HAL
open science

Quelques méthodes mathématiques pour la simulation moléculaire et multiéchelle

Gabriel Stoltz

To cite this version:

Gabriel Stoltz. Quelques méthodes mathématiques pour la simulation moléculaire et multiéchelle. Mathématiques [math]. Ecole des Ponts ParisTech, 2007. Français. NNT : . tel-00166728

HAL Id: tel-00166728

https://pastel.hal.science/tel-00166728

Submitted on 9 Aug 2007

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

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

Ecole Nationale des Ponts et Chaussées — Ecole doctorale ICMS

THÈSE

présentée pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'ECOLE NATIONALE DES PONTS ET CHAUSSEES

Spécialité : **Mathématiques**

Présentée par

Gabriel STOLTZ

QUELQUES METHODES MATHEMATIQUES POUR LA SIMULATION MOLECULAIRE ET MULTIECHELLE

Rapporteurs :	Andrew STUART	University of Warwick
	Eric DARVE	University of Stanford

Soutenue publiquement le **14 juin 2007** devant le jury composé de

Yvon MADAY	Université Paris VI	Président
Philippe CHARTIER	INRIA Rennes	Examinateur
Pierre DEL MORAL	Université de Nice	Examinateur
Claude LE BRIS	CERMICS	Examinateur
Gilles ZERAH	CEA/DAM	Examinateur
Eric CANCES	CERMICS	Directeur

Es gibt für Jeden keinen anderen Weg der Entfaltung und Erfüllung als den der möglichst vollkommenen Darstellung des eigenen Wesens. ≫ Sei Du Selbst ≪ ist das ideale Gesetz, zu mindest für den jungen Menschen, es gibt keinen andern Weg zur Warheit und zur Entwicklung.

Daß dieser Weg durch viele moralische and andre Hindernisse erschwert wird, daß die Welt uns lieber angepaßt und schwach sieht als eigensinnig, daraus entsteht für jeden mehr als durchschnittlich individualisierten Menschen der Lebenskampf. Da muß jeder für sich allein, nach seinen eigenen Kräften und Bedürfnissen, entscheiden, wieweit er sich der Konvention unterwerfen oder ihr trotzen will. Wo er die Konvention, die Forderungen von Familie, Staat, Gemeinschaft in den Wind schlägt, muß er es tun mit dem Wissen darum, daß es auf seine eigene Gefahr geschieht. Wiewiel Gefahr einer auf sich zu nehmen fähig ist, dafür gibt es keinen objektiven Maßstab. Man muß jedes Zuviel, jedes Überschreiten des eigenen Maßes büßen, man darf ungestraft weder im Eigensinn noch im Anpassen zu weit gehen.

Hermann Hesse, *Eigensinn macht Spaß*

A Axelle,
pour sa patience et son soutien
au cours de ces années.

Remerciements

Je tiens en premier lieu à remercier mon directeur de thèse, Eric Cancès, qui a su concilier avec brio un encadrement à la fois tactique (tes interventions techniques et idées de démonstration se sont souvent révélées fort à propos !), et surtout stratégique : tu as su me suggérer des directions de recherche intéressantes et fructueuses, et tu as toujours eu à coeur de me faire rencontrer tes relations et contacts scientifiques. J'en profite également pour remercier Claude Le Bris, qui m'a introduit au monde de la recherche en encadrant mon stage de DEA, et qui a été de très bon conseil tout au long de ces années au CERMICS.

Mes pensées se tournent ensuite vers tous mes collaborateurs : Jean-Bernard Maillet et Laurent Soulard, qui m'ont introduit au monde des ondes de choc et à leur simulation ; Frédéric Legoll, qui m'a appris la dynamique moléculaire ; Mathieu Lewin, qui a guidé mes premiers pas dans le monde quantique ; Anthony Scemama, avec qui nous avons eu des échanges fructueux sur VMC ; et surtout Mathias et Tony, pour notre travail commun sur le calcul des énergies libres : vous avez plus que contribué à mon initiation aux probabilités ! Je remercie enfin tous ceux qui m'ont invité à les visiter à l'étranger et qui m'ont à cette occasion consacré de leur temps scientifique, en particulier Andrew Stuart à Warwick et Arthur Voter à Los Alamos.

J'en profite pour remercier tous mes collègues, au CEA (surtout les "jeunes", que j'ai plus cotoyés, également François Jollet et Gilles Zérah), ou au CERMICS : pour vos conseils informatiques, scientifiques, pour les discussions que nous avons pu avoir (sur les mathématiques en général, la physique des solides, voire le foot ou la politique), ou simplement pour votre bonne humeur ou votre gentillesse (ça fait du bien les jours où la science ne va pas...!). Ceux qui me connaissent ne s'étonneront pas que j'ajoute un mot sur mes collègues coureurs à pied, qui ont eu la joie et le courage de m'accompagner lors de mes footings matinaux pendant les conférences et séjours à l'étranger – la palme revenant à François Castella, pour sa pugnacité et son enthousiasme ! J'ai aussi une pensée pour tous les élèves et stagiaires dont j'ai pu avoir la responsabilité pendant ces trois ans : j'espère qu'ils ne garderont pas un trop mauvais souvenir de mes enseignements ! Un grand merci enfin à Sylvie, Khadija et Martine pour votre soutien administratif au CERMICS.

Ces années de science n'auraient pas non plus été possibles sans un financement adéquat : c'est l'occasion rêvée d'exprimer ma gratitude envers mon employeur, le Ministère de l'Equipement, qui a bien voulu me laisser faire une thèse. Le CEA de Bruyères-le-Châtel a également été un contributeur notoire à la bonne marche de mon travail, en finançant une bonne partie de mes dépenses et voyages.

Un dernier mot pour tous ceux qui ne comprenaient pas vraiment les tourments qui ont pu être les miens, mais qui ont toutefois été d'un indéfectible soutien : mes parents, qui m'ont toujours poussé dans mes études, mon frère Gilles, qui a également le bon goût d'être mathématicien, ma famille, mes amis, et Jacques Darras, mon entraîneur d'athlétisme, parce qu'il savait y faire pour me changer les idées en me faisant suer sur autre chose que des problèmes de mathématiques.

Enfin, merci Axelle, tu sais me faire rire et me rendre heureux, et souffrir que je t'abandonne pour les besoins de la science...

Quelques méthodes mathématiques pour la simulation moléculaire et multiéchelle

Résumé : Ce travail présente quelques contributions à l'étude théorique et numérique des modèles utilisés en pratique pour la simulation moléculaire de la matière. En particulier, on présente et on analyse des méthodes numériques stochastiques dans le domaine de la physique statistique, permettant de calculer plus efficacement des moyennes d'ensemble. Une application particulièrement importante est le calcul de différences d'énergies libres, par dynamiques adaptatives ou hors d'équilibre. On étudie également quelques techniques, stochastiques ou déterministes, utilisées en chimie quantique et permettant de résoudre de manière approchée le problème de minimisation associé à la recherche de l'état fondamental d'un opérateur de Schrödinger en dimension grande. On propose enfin des modèles réduits permettant une description microscopique simplifiée des ondes de choc et de détonation par le biais d'une dynamique stochastique sur des degrés de liberté moyens, approchant la dynamique hamiltonienne déterministe du système complet.

Mots-clés : Equations aux dérivées partielles, équations différentielles stochastiques, systèmes dynamiques en physique statistique, méthodes de Monte-Carlo, ondes de choc.

Some Mathematical Methods for Molecular and Multiscale Simulation

Abstract: This work presents some contributions to the theoretical and numerical study of models used in practice in the field of molecular simulation. In particular, stochastic techniques to compute more efficiently ensemble averages in the field of computational statistical physics are presented and analyzed. An important application is the computation of free energy differences using nonequilibrium or adaptive dynamics. Some stochastic or deterministic techniques to solve approximately the Schrödinger ground state problem for high dimensional systems are also studied. Finally, some reduced models for shock and detonation waves, relying on an average stochastic dynamics reproducing in a mean sense the high dimensional deterministic hamiltonian dynamics, are proposed.

Keywords: Partial differential equations, stochastic differential equations, dynamical systems in statistical physics, Monte-Carlo methods, shock waves.

AMS Classification: 35P05, 35J60, 37A25, 37A60, 65C30, 65C40, 76L05, 82B30, 82B35.

Table des matières

1 Préambule . . . 1
1.1 Présentation des principaux résultats de la thèse . . . 1
1.1.1 Modèles de chimie quantique . . . 1
1.1.2 Dynamique moléculaire et calcul de différences d'énergie libre . . . 2
1.1.3 Modèles réduits pour les ondes de choc . . . 2
1.2 Liste des articles parus ou acceptés dans des revues à comité de lecture . . . 3
1.3 Autres travaux . . . 3

Partie I Introduction à la simulation moléculaire

2 Simulation moléculaire : une hiérarchie de modèles . . . 7
2.1 Description quantique de la matière . . . 10
2.1.1 Equation de Schrödinger et problème électronique . . . 11
2.1.2 Résolution directe du problème électronique . . . 12
2.1.3 Matrices densité d'ordre deux . . . 15
2.1.4 Méthodes de fonction d'onde . . . 16
2.1.5 Théorie de la fonctionnelle de la densité . . . 17
2.2 Description classique de la matière . . . 21
2.2.1 Représentation classique de la matière à l'échelle microscopique . . . 21
2.2.2 L'ensemble microcanonique . . . 23
2.2.3 L'ensemble canonique . . . 24
2.2.4 Autres ensembles thermodynamiques . . . 26
2.2.5 Propriétés dépendant du temps . . . 27
2.3 Simuler des systèmes plus grands pendant des temps plus longs . . . 28
2.3.1 Calcul de différences d'énergie libre . . . 28
2.3.2 Différentes approches pour augmenter le temps effectif de simulation . . . 37
2.3.3 Dynamiques réduites . . . 44

Part II Sampling Techniques in Molecular Dynamics

3 Phase-space sampling techniques . . . 53
3.1 Purely stochastic methods . . . 56
3.1.1 Rejection method . . . 56
3.1.2 Rejection control . . . 58
3.1.3 Metropolized independence sampler . . . 58
3.1.4 Importance sampling . . . 62

3.2 Stochastically perturbed Molecular Dynamics methods . . . 62
3.2.1 General framework for NVE Molecular Dynamics . . . 63
3.2.2 Hybrid Monte Carlo . . . 63
3.2.3 Biased Random-Walk . . . 75
3.2.4 Langevin dynamics . . . 78
3.3 Deterministic molecular dynamics sampling . . . 83
3.3.1 The Nosé-Hoover and Nosé-Hoover chains methods . . . 83
3.3.2 The Nosé-Poincaré and the Recursive Multiple Thermostat methods . . . 84
3.4 Numerical illustrations . . . 85
3.4.1 Description of the linear alkane molecule . . . 86
3.4.2 Discrepancy of sample points . . . 87
3.4.3 Choice of parameters . . . 89
3.4.4 Numerical results . . . 93
3.4.5 Improvement of the convergence rates . . . 94
3.4.6 Computation of correlation functions . . . 96
3.5 Stochastic boundary conditions . . . 96
3.5.1 Review of some classical stochastic boundary conditions . . . 97
3.5.2 An example of thermal boundary conditions . . . 99
3.6 Some background on continuous state-space Markov chains and processes . . . 105
3.6.1 Some background on continuous state-space Markov chains . . . 105
3.6.2 Some convergence results for Markov processes. . . . 114

4 Computation of free energy differences . . . 119
4.1 Nonequilibrium computation of free energy differences . . . 120
4.1.1 The Jarzynski equality (The alchemical case) . . . 120
4.1.2 The Jarzynski equality (The reaction coordinate case) . . . 122
4.1.3 Practical computation of free energy differences . . . 131
4.1.4 Numerical results . . . 134
4.2 Equilibration of the nonequilibrium computation of free energy differences . . . 138
4.2.1 The IPS and its statistical properties . . . 139
4.2.2 Consistency through a mean-field limit . . . 141
4.2.3 Numerical implementation . . . 143
4.2.4 Applications of the IPS method . . . 143
4.3 Path sampling techniques . . . 148
4.3.1 The path ensemble with stochastic dynamics . . . 150
4.3.2 Equilibrium sampling of the path ensemble . . . 152
4.3.3 (Non)equilibrium sampling of the path ensemble . . . 163
4.4 Adaptive computation of free energy differences . . . 169
4.4.1 A general framework for adaptive methods . . . 170
4.4.2 Rigorous convergence results for the Adaptive Biasing Force method . . . 179

Part III Shock Waves: a Multiscale Approach

5 A reduced model for shock waves . . . 191
5.1 A simplified one-dimensional model . . . 192
5.1.1 Shock waves in one-dimensional lattices . . . 192
5.1.2 An augmented one-dimensional model . . . 197
5.1.3 The stochastic limit . . . 205
5.1.4 Extension to the reactive case . . . 209
5.2 A reduced model based on Dissipative Particle Dynamics . . . 212

5.2.1 Previous mesoscopic models . . . 212
5.2.2 A reduced model in the inert case . . . 213
5.2.3 The reactive case . . . 218

Part IV Mathematical Study of some Quantum Models

6 Variational Monte-Carlo . . . 227
6.1 Description of the algorithms . . . 229
6.1.1 Random walks in the configuration space . . . 229
6.1.2 Random walks in the phase space . . . 231
6.2 Numerical experiments and applications . . . 234
6.2.1 Measuring the efficiency . . . 234
6.2.2 Numerical results . . . 236
6.2.3 Discussion of the results . . . 238

7 Second-order reduced density matrices . . . 241
7.1 The electronic structure problem in terms of second order reduced density matrices 242
7.1.1 The ensemble of N-repsentable second-order density matrices . . . 242
7.1.2 The energy minimization problem in terms of second order reduced-density matrices . . . 243
7.2 The N-representability problem . . . 244
7.2.1 Some necessary N-representability conditions for 2-RDMs . . . 244
7.2.2 An explicit (counter)example . . . 246
7.3 A dual formulation of the optimization problem . . . 247
7.3.1 Dual Formulation of the RDM Minimization Problem . . . 247
7.3.2 Algorithm for solving the dual problem . . . 248
7.3.3 Numerical results . . . 250

8 Local Exchange Potentials and Optimized Effective Potentials . . . 253
8.1 The Slater exchange potential . . . 255
8.2 The Optimized Effective Potential problem . . . 257
8.2.1 Usual formulation of the OEP problem . . . 257
8.2.2 A well-posed reformulation of the OEP problem . . . 258
8.3 The effective local potential minimization problem . . . 260
8.4 Mathematical proofs . . . 261
8.4.1 Some useful preliminary results . . . 261
8.4.2 Proofs for the Slater potential . . . 262
8.4.3 Proof of Proposition 8.4 . . . 267

Partie V Bibliographie

Bibliographie . . . 271

1

Préambule

1.1 Présentation des principaux résultats de la thèse

J'ai étudié pendant ma thèse plusieurs techniques de simulation moléculaire, d'un point de vue mathématique. On peut répartir ces études selon trois grands thèmes :

(A) l'analyse mathématique et numérique de certains modèles de chimie quantique (Partie IV) ;

(B) l'analyse mathématique et numérique de techniques d'échantillonnage en dynamique moléculaire, avec un accent particulier sur les techniques stochastiques et le calcul de différences d'énergie libre (Partie II) ;

(C) la recherche d'un modèle réduit pour les ondes de chocs décrites au niveau microscopique (Partie III).

1.1.1 Modèles de chimie quantique

Les méthodes que j'ai regardées en chimie quantique ne sont pas les plus couramment utilisées en pratique, mais sont toutefois très intéressantes d'un point de vue mathématique :

(a) avec Michel Caffarel, Eric Cancès, Tony Lelièvre, et Anthony Scemama, nous avons proposé une nouvelle méthode d'échantillonnage pour les techniques de Monte-Carlo variationnel (voir [P8] et le Chapitre 6), qui s'est révélée plus efficace et plus robuste que les approches précédentes, au moins pour les systèmes de référence considérés. Cette nouvelle méthode d'échantillonnage est obtenue en projetant une dynamique étendue de type Langevin sur l'espace des configurations électroniques, et est ainsi une extension de la traditionnelle marche aléatoire biaisée sur les configurations électroniques ;

(b) avec Eric Cancès et Mathieu Lewin nous avons proposé une formulation duale du problème de minimisation électronique formulé à l'aide de la matrice densité d'ordre deux (voir [P9] et le Chapitre 7), et avons testé numériquement la méthode numérique associée sur un ensemble de petites molécules ;

(c) avec Eric Cancès, nous avons également considéré le problème du potentiel effectif optimal (défini comme le potentiel local dans les équations de Kohn-Sham qui permet d'obtenir la meilleure énergie d'Hartree-Fock): plus précisément, nous avons étudié mathématiquement la proposition de Ernest Davidson, Artur Izmaylov, Gustavo Scuseria, et Viktor Staroverov, qui définissent un potentiel effectif local par le biais d'une procédure de minimisation annexe, et ce, afin de limiter les instabilités rencontrées dans les simulations numériques (voir [P5], [A2] et le Chapitre 8).

1.1.2 Dynamique moléculaire et calcul de différences d'énergie libre

J'ai étudié différentes techniques stochastiques permettant de calculer en pratique les quantités intéressantes définies en physique statistique :

(a) j'ai tout d'abord comparé différentes méthodes d'échantillonnage de la mesure canonique, d'un point théorique et numérique. Ce travail a été réalisé en collaboration avec Eric Cancès et Frédéric Legoll (voir [P3] et la Chapitre 3).

(b) je me suis ensuite intéressé au calcul de différences d'énergie libre :

  (i) en utilisant dans un premier temps des dynamiques hors d'équilibre et l'égalité de Jarzynski. Cette égalité peut être obtenue de manière rigoureuse lorsque le chemin de transition le long duquel on calcule les différences d'énergie libre est paramétré par un paramètre extérieur, et nous avons montré avec Tony Lelièvre et Mathias Rousset comment étendre ces résultats au cas de transitions indexées par une coordonnée de réaction (fonction de la configuration microscopique du système) grâce à des dynamiques stochastiques projetées (voir [P6] et la Section 4.1.2). Avec Mathias Rousset, nous avons également proposé une manière d'équilibrer la transition hors d'équilibre par le biais d'une procédure de sélection qui évite la dégénerescence des poids exponentiels dans l'inégalité de Jarzynski (voir [P10] et la Section 4.2);

  (ii) plus récemment, nous avons étudié les méthodes adaptatives de calcul de différences d'énergie libre. Toujours avec Tony Lelièvre et Mathias Rousset, nous avons proposé un formalisme général qui permet de présenter toutes les stratégies adaptatives de manière unifiée, avons montré l'existence d'un état stationnaire, et avons proposé une procédure de sélection qui permet de rendre plus efficace une implémentation parallèle des stratégies adaptatives (voir [P4] et la Section 4.4.1). Enfin, avec Tony Lelièvre, Felix Otto, et Mathias Rousset, nous sommes en train de montrer rigoureusement la convergence de certaines dynamiques adaptatives limites, en utilisant des méthodes entropiques (voir [A1] et la Section 4.4.2).

(c) j'ai également proposé quelques extensions des méthodes usuelles d'échantillonnage de chemins de réactions lorsque des dynamiques stochastiques sont utilisées (voir [P1] et la Section 4.3).

1.1.3 Modèles réduits pour les ondes de choc

Mon travail dans ce domaine a été effectué au CEA, en collaboration avec Jean-Bernard Maillet et Laurent Soulard. L'objectif était de trouver un modèle réduit mésoscopique permettant de décrire les principales caractéristiques des ondes de choc et de détonation simulées au niveau microscopique :

(a) j'ai tout d'abord proposé un modèle simplifié pour les ondes de choc unidimensionnelles, adapté au cas des solides cristallins (voir [P11] et la Section 5.1);

(b) j'ai ensuite proposé un modèle tridimensionnel pour les ondes de choc, fondé sur un modèle de type *Dissipative Particle Dynamics* (voir [P7] et la Section 5.2.2).

(c) Avec Jean-Bernard Maillet et Laurent Soulard, nous avons alors étendu ce modèle au cas d'ondes de choc réactives (voir [P2] et la Section 5.2.3).

Les modèles proposés dans [P7,P2] sont bien fondés du point de vue de la physique statistique, et les résultats numériques correspondants sont en bon accord avec les résultats de simulation tous-atomes, de manière qualitative [P2] et quantitative [P7].

1.2 Liste des articles parus ou acceptés dans des revues à comité de lecture

[P1] G. Stoltz, Path sampling with stochastic dynamics: some new algorithms, *J. Comput. Phys.* **225** (2007) 491-508

[P2] J.-B. Maillet, L. Soulard et G. Stoltz, A reduced model for shock and detonation waves. II. The reactive case, *Europhys. Lett.* **78**(6) (2007) 68001

[P3] E. Cancès, F. Legoll et G. Stoltz, Theoretical and numerical comparison of some sampling methods, *M2AN* **41**(2) (2007) 351-390

[P4] T. Lelièvre, M. Rousset and G. Stoltz, Computation of free energy profiles with parallel adaptive dynamics, *J. Chem. Phys.* **126** (2007) 134111.

[P5] A.F. Izmaylov, V.N. Staroverov, G. Scuseria, E.R. Davidson, G. Stoltz et E. Cancès, The effective local potential method: Implementation for molecules and relation to approximate optimized effective potential techniques, *J. Chem. Phys.* **126** (2007) 084107.

[P6] T. Lelièvre, M. Rousset et G. Stoltz, Computation of free energy differences through nonequilibrium stochastic dynamics: the reaction coordinate case, *J. Comp. Phys.* **222**(2) (2007) 624-643.

[P7] G. Stoltz, A reduced model for shock and detonation waves. I. The inert case, *Europhys. Lett.* **76**(5) (2006) 849-855.

[P8] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès et M. Caffarel, An efficient sampling algorithm for Variational Monte Carlo, *J. Chem. Phys.* **125** (2006) 114105.

[P9] E. Cancès, M. Lewin et G. Stoltz, The electronic ground state energy problem: a new reduced density matrix approach, *J. Chem. Phys.* **125** (2006) 064101.

[P10] M. Rousset et G. Stoltz, An interacting particle system approach for molecular dynamics, *J. Stat. Phys.* **123**(6) (2006) 1251-1272.

[P11] G. Stoltz, Shock waves in an augmented one-dimensional chain, *Nonlinearity* **18** (2005) 1967-1985.

1.3 Autres travaux

[A1] T. Lelièvre, F. Otto, M. Rousset et G. Stoltz, Long-time convergence of the Adaptive Biasing Force method, en préparation.

[A2] E. Cancès, E.R. Davidson, A.F. Izmaylov, G. Scuseria, V.N. Staroverov, et G. Stoltz, Local exchange potentials: a mathematical viewpoint, en préparation

[A3] T. Lelièvre, F. Legoll et G. Stoltz, Some remarks on sampling methods in Molecular Dynamics, Proceedings CANUM 2006, soumis à *ESAIM Proc* (2007)

[A4] J.N. Roux, S. Rodts et G. Stoltz, *Introduction à la physique statistique et quantique*, Cours de l'Ecole des Ponts et Chaussées (2007)

Partie I

Introduction à la simulation moléculaire

2

Simulation moléculaire : une hiérarchie de modèles

2.1 Description quantique de la matière **10**
2.1.1 Equation de Schrödinger et problème électronique 11
2.1.2 Résolution directe du problème électronique 12
2.1.3 Matrices densité d'ordre deux 15
2.1.4 Méthodes de fonction d'onde 16
2.1.5 Théorie de la fonctionnelle de la densité 17
2.2 Description classique de la matière **21**
2.2.1 Représentation classique de la matière à l'échelle microscopique 21
2.2.2 L'ensemble microcanonique 23
2.2.3 L'ensemble canonique 24
2.2.4 Autres ensembles thermodynamiques 26
2.2.5 Propriétés dépendant du temps 27
2.3 Simuler des systèmes plus grands pendant des temps plus longs . . . **28**
2.3.1 Calcul de différences d'énergie libre 28
2.3.2 Différentes approches pour augmenter le temps effectif de simulation . . 37
2.3.3 Dynamiques réduites 44

Physique quantique et physique statistique

La physique quantique et la physique statistique sont deux domaines importants de la physique contemporaine, et décrivent toutes deux la matière à l'échelle microscopique (voir respectivement les Sections 2.1 et 2.2). La physique quantique s'intéresse aux éléments constitutifs de la matière : protons, neutrons, électrons, dont l'évolution est régie par l'équation de Schrödinger. La physique statistique peut être utilisée pour décrire des systèmes quantiques ou classiques[1]. Cette théorie étudie le comportement des atomes, une entité résultant de la réunion d'un noyau (assemblage de protons et de neutrons) et de son nuage électronique. Des constantes physiques importantes sont rappelées dans la Table 2.1. On peut en déduire quelques ordres de grandeur de la description de la matière à l'échelle microscopique : les distances typiques s'expriment en Å (10^{-10} m), les énergies mises en jeu sont de l'ordre de $k_{\mathrm{B}}T \simeq 4 \times 10^{-21}$ J à température ambiante pour des systèmes classiques, alors qu'elles se mesurent en Hartrees (1 Ha = 27,2 eV = $43,6 \times 10^{-19}$ J) pour les systèmes quantiques ; enfin, l'unité de temps varie de 10^{-17} s à 10^{-15} s selon que l'on a affaire à un système quantique (la masse typique à prendre en compte est celle de l'électron) ou classique (la masse de référence est celle du proton).

[1] Par la suite, on utilisera souvent le terme de *classique* par opposition à *quantique* – et non pas comme synonyme de *courant* ou *usuel*...

Tableau 2.1. Quelques grandeurs ou constantes physiques importantes en physique quantique et en physique statistique.

Constante ou grandeur physique	Notation usuelle	Valeur
Nombre d'Avogadro	$\mathcal{N}_A$	$6,02 \times 10^{23}$
Constante de Boltzmann	k_{B}	$1,381 \times 10^{-23}$ J/K
Constante de Planck réduite	$\hbar$	$1,054 \times 10^{-34}$ Js
Charge élémentaire	e	$1,602 \times 10^{-19}$ C
Masse de l'électron	m_e	$9,11 \times 10^{-31}$ kg
Masse du proton	m_p	$1,67 \times 10^{-27}$ kg
Permittivité diélectrique du vide	ε_0	$8,854 \times 10^{-12}$ F/m
Electron-Volt	eV	$1,602 \times 10^{-19}$ J

Dans tous les cas, les ordres de grandeur utilisés dans la description microscopique de la matière sont loin des ordres de grandeur des quantités macroscopiques dont on a l'expérience quotidienne – de même que le nombre de particules étudiées, puisque les échantillons de matière macroscopiques contiennent de l'ordre de $\mathcal{N}_A \sim 10^{23}$ atomes ! Heureusement, la physique statistique permet de faire le lien entre les descriptions microscopique et macroscopique de la matière, en particulier

(i) dans le cadre de la *limite thermodynamique*, où le nombre de particules dans la description microscopique, ainsi que le volume de l'échantillon, tendent vers l'infini, alors que la densité est maintenue constante. Ce type de limite ne peut toutefois être justifié rigoureusement d'un point-de-vue mathématique que dans certains cas (voir par exemple les ouvrages de Ruelle [293] dans le cadre de la physique statistique classique et de Catto, Le Bris et Lions [55] pour l'étude de limites de modèles de physique quantique) ;

(ii) dans certains régimes physiques limites (basse densité, couplage faible, champ moyen,...), on peut décrire le système microscopique par le biais d'une équation cinétique sur la densité de probabilité d'une seule particule – telle que l'équation de Boltzmann (pour une justification mathématique de ces limites, on pourra se référer aux revues de Spohn sur ce sujet, en particulier à l'article [318] et à l'ouvrage [319]).

Physique quantique et physique statistique computationelles

Si le lien évoqué ci-dessus entre les descriptions microscopique et macroscopique est agréable d'un point-de-vue théorique, il est en revanche inutilisable pour des calculs pratiques des propriétés de la matière simulée à l'échelle moléculaire : cela demanderait en effet de simuler $\mathcal{N}_A$ atomes sur $\mathrm{O}(10^{15})$ pas d'intégration en temps. Il est bon de mettre ces nombres en regard des ordres de grandeur actuels (plutôt des records, en fait !) des problèmes pouvant être traités numériquement par la simulation moléculaire dans un cadre classique : la simulation complète du virus du tabac [111] a pu être menée pendant 50 ns pour un système de 1 million d'atomes ; le repliement de la tête d'une protéine (Villine) a été étudié *via* une trajectoire de 500 μs au total, pour un système de 20 000 atomes.[2]

La simulation moléculaire, malgré ses limitations spatiales et temporelles, a toutefois été utilisée de plus en plus couramment pendant les cinquante dernières années pour tester numériquement la validité de théories physiques – avant la vérification expérimentale proprement dite, qui reste l'ultime sanction. Les calculs numériques sont un complément au développement de théories physiques reposant sur des modèles simplifiés, d'où le terme d'*expérience numérique*. Dans ces expériences numériques, il est même de bon aloi d'enrichir autant que faire se peut le modèle physique sous-jacent. Cette utilisation de la simulation moléculaire a été initiée et soutenue par la physique des liquides simples, qu'aucune théorie physique ne décrivait correctement (voir en particulier le

[2] Voir la page internet du projet Folding@Home : `http://folding.stanford.edu/`

travail pionnier de METROPOLIS, ROSENBLUTH, ROSENBLUTH, TELLER et TELLER [238] en 1953, et la première simulation de dynamique moléculaire par ALDER et WAINWRIGHT en 1956 [3]). La chimie quantique computationelle a également débuté dans les années 50, avec, en chimie, les travaux de HALL [149] et ROOTHAN [288] en 1951, puis, en physique du solide, le modèle de KOHN et SHAM [195] en 1965.

La Microscopie numérique

La simulation moléculaire peut être utilisée comme un *microscope numérique*. En effet, il peut être difficile de comprendre la matière à l'échelle microscopique d'un point-de-vue expérimental, du fait de la grande précision requise, tant spatiallement que temporellement – ou même tout simplement, parce qu'on ne sait pas ce qu'on doit regarder ! Dans ce cas, des simulations numériques préliminaires sont un outil utile pour tester quelques idées sur les mécanismes en jeu (pour que le principe actif de ce médicament se lie bien à cette protéine, quelle est la séquence des changements de conformation des deux molécules qui doit avoir lieu ?), ou obtenir des données brutes que l'on peut traiter et analyser pour en tirer des informations sur les phénomènes observables par un dispositif expérimental. Ces considérations sont particulièrement pertinentes pour les nanosytèmes, très en vogue actuellement. Néanmoins, rappelons une dernière fois que les expérimentations numériques ne peuvent généralement pas remplacer complètement les validations expérimentales usuelles, et doivent donc plutôt être considérées comme un premier pas utile dans la construction de nouvelles théories ou la recherche de nouveaux résultats. On peut citer par exemple le criblage informatique des principes actifs de synthèse dans l'industrie pharmaceutique, qui permet de réduire de manière conséquente le nombre de molécules à synthétiser puis à tester par des protocoles expérimentaux longs et coûteux.

Calcul de propriétés moyennes des systèmes physiques

Un des principaux objectifs de la simulation moléculaire est le calcul de propriétés moyennes des systèmes physiques – *i.e.* des grandeurs macroscopiques qui pourraient également être mesurées expérimentalement, mais que l'on préfère calculer numériquement pour des raisons financières ou techniques. Un exemple prototypique d'une telle simulation est l'étude des propriétés de la structure interne de la Terre, en particulier son noyau, par des simulations *ab-initio* [316]. De manière générale, le calcul numérique est une alternative intéressante pour des régimes de haute pression, densité ou température.

La physique statistique permet de relier la simulation de sytèmes physiques à l'échelle moléculaire et les grandeurs macroscopiques par le biais de moyennes sur des *ensembles thermodynamiques* :

$$\langle A \rangle = \int_{\mathcal{M}^N \times \mathbb{R}^{3N}} A(q,p)\, d\mu(q,p). \tag{2.1}$$

Dans cette expression, la fonction $A \equiv A(q,p)$ est une observable, et la variable q donne les positions $q = (q_1, \ldots, q_N) \in \mathcal{M}^N$ des particules simulés, la variable p donnant les impulsions $p = (p_1, \ldots, p_N) \in \mathbb{R}^{3N}$ desdites particules. La mesure μ est une mesure de probabilité qui dépend de l'ensemble thermodynamique utilisé (voir Section 2.2).

Une grandeur que l'on calcule souvent pour des fluides est la pression P, par exemple dans le cas d'un matériau modélisé par un système de Lennard-Jones. L'observable associée, pour des particules de masses m_i, est

$$A(q,p) = \frac{1}{3|\mathcal{M}|} \sum_{i=1}^{N} \left(\frac{|p_i|^2}{m_i} - q_i \cdot \frac{\partial V}{\partial q_i}(q) \right),$$

où $|\mathcal{M}|$ est le volume accessible au système, l'énergie potentielle V du système étant dans ce cas donnée par (2.27)-(2.28).

En pratique, des moyennes telles que (2.1) sont calculées pour des systèmes de très petite taille par rapport aux dimensions macroscopiques typiques (on est donc loin du régime de la limite thermodynamique...). Cependant, l'expérience numérique montre que si les interactions entre les particules sont à courte portée, on peut obtenir tout de même de très bons résultats ! Par exemple, l'équation d'état de l'argon présentée en Figure 2.1 (courbe pression/densité à température fixée) a été obtenue pour un système de quelques milliers de particules seulement, soit 10^{20} fois moins que dans un échantillon macroscopique... La courbe ainsi calculée se compare toutefois très bien aux mesures expérimentales, et permet même de calculer des points dans un régime de forte densité difficilement accessible expérimentalement.

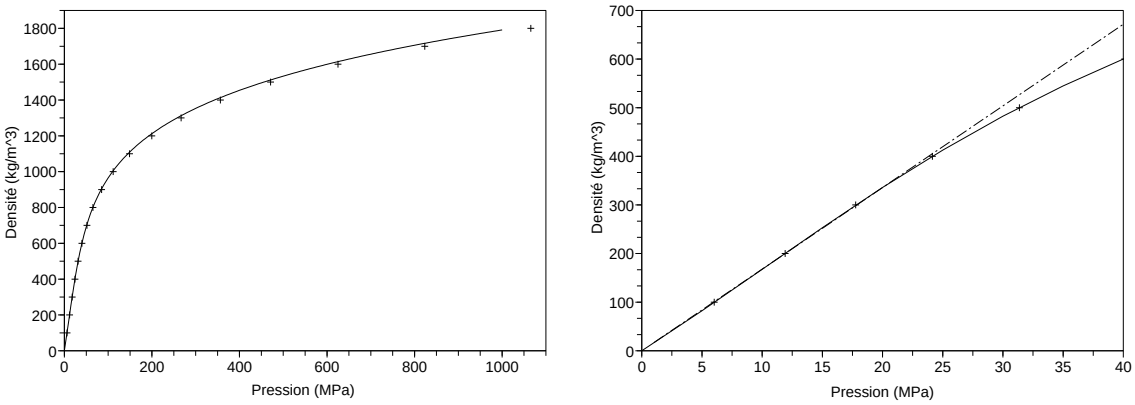


Fig. 2.1. Loi d'état de l'argon à $T = 300$ K : résultats numériques ('+') et courbe expérimentale de référence (ligne pleine). Le régime des gaz parfaits est indiqué en traits interrompus.

2.1 Description quantique de la matière

On considère, dans cette section, un système moléculaire composé de M noyaux, que l'on suppose figés à des positions $\bar{x}_i \in \mathbb{R}^3$ $(1 \leq i \leq M)$, et de N électrons, dont les variables de position et de spin sont notées respectivement $x_j \in \mathbb{R}^3$ et $\sigma_j \in \{|\uparrow\rangle, |\downarrow\rangle\}$ $(1 \leq j \leq N)$. L'état (électronique) du système est décrit à l'instant t par une fonction d'onde

$$\psi(t; (x_1, \sigma_1), \ldots, (x_N, \sigma_N)) \in \mathbb{C}.$$

Pour que la fonction d'onde ψ soit un état physiquement admissible, il faut que les conditions suivantes soit satisfaites :

(i) Normalisation: la fonction d'onde est normalisée pour la norme L^2, au sens où

$$\sum_{\sigma_1 \in \{|\uparrow\rangle, |\downarrow\rangle\}} \cdots \sum_{\sigma_N \in \{|\uparrow\rangle, |\downarrow\rangle\}} \int_{\mathbb{R}^{3N}} |\psi(t, (x_1, \sigma_1), \ldots, (x_N, \sigma_N))|^2 \, dx_1 \ldots dx_N = 1. \tag{2.2}$$

Cette propriété découle de l'interprétation de $|\psi(t, \cdot)|^2$ comme densité de probabilité ;

(ii) Propriété d'indiscernabilité des particules identiques : le principe de Pauli demande que la fonction d'onde soit antisymétrique pour l'échange de deux particules identiques. Plus précisément, pour une permutation p des indices $\{1, \ldots, N\}$, de signature $\varepsilon(p)$,

$$\psi(t, (x_{p(1)}, \sigma_{p(1)}), \ldots, (x_{p(N)}, \sigma_{p(N)})) = \varepsilon(p)\, \psi(t, (x_1, \sigma_1), \ldots, (x_N, \sigma_N)).$$

Les fonctions d'onde électroniques admissibles sont donc des élements de l'espace fonctionnel

$$\mathcal{H} = \bigwedge_{i=1}^{N} \mathrm{L}^2\left(\mathbb{R}^3 \times \{|\uparrow\rangle, |\downarrow\rangle\}, \mathbb{C}\right),$$

de norme 1 (pour le produit scalaire induit par la norme (2.2)).

Pour préciser plus avant l'espace fonctionnel des fonction d'onde, on introduit l'opérateur Hamiltonien du système :

$$H = -\sum_{i=1}^{N} \frac{\hbar^2}{2m}\Delta_{x_i} - \sum_{i=1}^{N}\sum_{k=1}^{M} \frac{Z_k e^2}{4\pi\varepsilon_0|x_i - \bar{x}_k|} + \sum_{1\le i<j\le N} \frac{e^2}{4\pi\varepsilon_0|x_i - x_j|},$$

où $Z_k e$ est la charge du k-ième noyau, et m la masse d'un électron. On travaille dans la suite avec les unités atomiques, qui sont telles que

$$m = 1, \qquad e = 1, \qquad \hbar = 1, \qquad \frac{1}{4\pi\varepsilon_0} = 1.$$

Dans ce système d'unité, l'unité de masse est $9,11 \times 10^{-31}$ kg, l'unité de longueur est le rayon de Bohr $a_0 = 5,29 \times 10^{-11}$ m, l'unité de temps est $2,42 \times 10^{-17}$ s, et l'unité d'énergie est le Hartree $\mathrm{Ha} = 4,36 \times 10^{-18}$ J = 27,2 eV = 627 kcal/mol. Ce changement d'unité permet de manipuler des quantités accessibles intuitives : pour de petits systèmes à l'équilibre (N et $Z = \sum_{k=1}^{M} Z_k$ assez petits), la distance typique entre un électron et le noyau auquel il est rattaché est en moyenne de l'ordre du rayon de Bohr, et les énergies des états fondamentaux (à l'équilibre) sont de quelques Ha. L'opérateur Hamiltonien est, en unités atomiques,

$$H = -\sum_{i=1}^{N} \frac{1}{2}\Delta_{x_i} - \sum_{i=1}^{N}\sum_{k=1}^{M} \frac{Z_k}{|x_i - \bar{x}_k|} + \sum_{1\le i<j\le N} \frac{1}{|x_i - x_j|}. \tag{2.3}$$

On note, pour le reste de cette section,

$$V_{\mathrm{nuc}}(x) = -\sum_{k=1}^{M} \frac{Z_k}{|x - \bar{x}_k|}.$$

L'opérateur Hamiltonien est auto-adjoint sur $\mathcal{H}$ (pour une introduction à la théorie spectrale des Hamiltoniens quantiques, on pourra consulter les ouvrages de Reed et Simon [277] ou de Dautray et Lions [99]).

2.1.1 Equation de Schrödinger et problème électronique

On s'intéresse dans la suite aux propriétés des états fondamentaux des systèmes décrits dans le formalisme quantique, *i.e.* aux propriétés du premier vecteur propre et de la première valeur propre de l'opérateur Hamiltonien. On introduit donc le problème de minimisation suivant :

$$E = \inf\{\langle\psi, H\psi\rangle \mid \psi \in \mathcal{H},\ \|\psi\|_{\mathrm{L}^2} = 1\}. \tag{2.4}$$

Un minimiseur de (2.4) est un vecteur propre de l'opérateur Hamiltonien, associé à la valeur propre E :

$$H\psi = E\psi.$$

L'existence de tels minimiseurs pour des potentiels de type Coulombien lorsque $\sum_{k=1}^{M} Z_k \ge N$ est assurée par des résultats de théorie spectrale [99, 277, 377]. Comme l'opérateur H est réel, on peut

se restreindre en fait à des fonctions d'ondes à valeurs réelles. Au vu du laplacien intervenant dans l'expression (2.3) de l'opérateur Hamiltonien, on peut également se restreindre à une minimisation sur l'espace fonctionnel

$$\mathcal{H}^1 = \bigwedge_{i=1}^{N} \mathrm{H}^1\left(\mathbb{R}^3 \times \{|\uparrow\rangle, |\downarrow\rangle\}, \mathbb{R}\right).$$

Remarque 2.1. *Pour éviter d'alourdir les notations, on n'a pas noté explicitement la dépendance de l'énergie de l'état fondamental (2.4) par rapport aux positions des noyaux atomiques. L'énergie de l'état fondamental est paramétrisée par ces positions $\bar{x}_1, \ldots, \bar{x}_M$, et on peut définir*

$$U(\bar{x}_1, \ldots, \bar{x}_M) = \inf\left\{\langle \psi, H_{\bar{x}_1,\ldots,\bar{x}_M}\psi\rangle \mid \psi \in \mathcal{H}^1,\ \|\psi\|_{\mathrm{L}^2} = 1\right\}, \tag{2.5}$$

La fonction U définie ci-dessus peut être utilisée pour étudier la dynamique et les propriétés statistiques de systèmes moléculaires décrits dans un formalisme quantique (voir Section 2.2). On parle dans ce cas de dynamique ab-initio. Cette manière de procéder repose donc sur l'approximation que les évolutions des degrés de liberté électroniques et nucléaires peuvent être découplées, plus précisément que les degrés de liberté électroniques peuvent effectivement être décrits par une fonction d'onde où seules les positions des noyaux sont des paramètres extérieurs (en particulier, on n'a pas besoin de prendre en compte les vitesses des noyaux). On trouvera plus de précisions mathématiques sur cette approximation (dite de Born-Oppenheimer) dans l'ouvrage de Teufel *[342].*

Pour alléger les notations et simplifier l'exposé, on omet par la suite la variable de spin dans la minimisation (2.4). Ceci ne change rien aux difficultés mathématiques rencontrées.

2.1.2 Résolution directe du problème électronique

On présente en premier lieu des méthodes cherchant à résoudre directement le problème de minimisation (2.4) (éventuellement approximativement, par le biais d'une borne supérieure par exemple). C'est une tâche non-triviale parce que (2.4) est un problème de minimisation en dimension grande (posé dans $\mathrm{L}^2(\mathbb{R}^{3N})$), et donc les méthodes usuelles d'optimisation (techniques de gradient en particulier) sont souvent vouées à l'échec.

Méthode de Monte-Carlo variationnelle

La méthode de Monte-Carlo variationnelle (*Variational Monte-Carlo*, VMC) repose sur la borne supérieure suivante de l'énergie fondamentale (2.4) : pour toute fonction $\psi \in \mathcal{H}$,

$$E \leq \frac{\langle \psi, H\psi\rangle}{\langle \psi, \psi\rangle} = \frac{\displaystyle\int_{\mathbb{R}^{3N}} E_L^{\psi}(x)\,|\psi(x)|^2\,dx}{\displaystyle\int_{\mathbb{R}^{3N}} |\psi(x)|^2\,dx}, \tag{2.6}$$

où $E_L^{\psi}(x) = [H\psi](x)/\psi(x)$. La fonction E_L^{ψ} est appelée l'énergie locale de la fonction ψ. Remarquons que si ψ était un vecteur propre de H associé à la valeur propre E, on aurait $E_L^{\psi}(x) = E$ pour tout x, et dans ce cas la variance de la fonction E_L^{ψ} (pour la mesure de densité $|\psi(x)|^2$) serait nulle.

La plupart du temps, les calculs VMC sont faits avec des fonctions d'onde test ψ qui sont de bonnes approximations de la fonction d'onde fondamentale ψ_0. Souvent, ψ est une somme de déterminants construits avec des orbitales atomiques de Slater, multipliée par un facteur de Jastrow prenant en compte les corrélations électroniques (voir la formule (6.8) pour plus de précisions, et l'analyse mathématique de Fournais, Hoffmann-Ostenhof, Hoffmann-Ostenhof et Ostergaard Sorensen [110] motivant l'introduction de du facteur de corrélation de Jastrow).

Lorsque l'on considère une famille de fonctions d'ondes, dépendant de certains paramètres, et que l'on optimise lesdits paramètres (de façon à minimiser l'énergie ou la variance de E_L^ψ), de bonnes bornes supérieures de l'énergie fondamentale peuvent être obtenues (voir en particulier le travail de Umrigar et Fillippi [351]).

En pratique, la borne supérieure (2.6) peut être vue comme la moyenne de la quantité E_L par rapport à la mesure de probabilité $Z_\psi^{-1}|\psi(x)|^2\,dx$ (avec la constante de normalisation $Z_\psi = \int_{\mathbb{R}^{3N}} |\psi|^2$). Comme l'intégrale (2.6) est posée sur un espace de grande dimension, il est naturel de recourir à des méthodes stochastiques pour l'évaluer. De telles méthodes sont présentées au Chapitre 3, et peuvent toutes être adaptées au cadre de VMC. En particulier, nous avons montré dans [P8], avec E. Cancès, M. Caffarel, A. Scemama et T. Lelièvre, qu'il était intéressant de remplacer la dynamique de gradient usuellement utilisée dans la communauté VMC par une dynamique de type Langevin (avec quelques adaptations techniques, voir le Chapitre 6 pour une présentation complète de cette nouvelle stratégie numérique, et les résultats numériques correspondants).

Méthode de Monte-Carlo diffusif

La principe de la méthode de Monte-Carlo diffusive (*Diffusion Monte-Carlo*, DMC) repose sur la remarque que le premier état propre d'un opérateur elliptique peut être retrouvé comme la limite en temps long d'un processus de diffusion. En effet, lorsque l'opérateur Hamiltonien est auto-adjoint et qu'il existe un trou spectral $\gamma > 0$ entre la première valeur propre (supposée isolée et de multiplicité 1) du spectre discret et la seconde, la solution de l'équation aux dérivées partielles (EDP)

$$\frac{\partial \phi}{\partial t} = -H\phi, \qquad \phi(0,x) = \psi_I(x), \tag{2.7}$$

est telle que

$$\|\mathrm{e}^{E_0 t}\phi(t) - \langle \psi_I, \psi_0\rangle\, \psi_0\| \leq C\mathrm{e}^{-\gamma t},$$

où ψ_0 est la fonction d'onde fondamentale, et E_0 l'énergie fondamentale associée. On montre également que l'énergie calculée au temps t converge exponentiellement rapidement vers l'énergie fondamentale ; plus précisément,

$$0 \leq \frac{\langle \psi_I, H\phi(t)\rangle}{\langle \psi_I, \phi(t)\rangle} - E_0 \leq \frac{\langle H\psi_I, \psi_I\rangle - E_0}{\langle \psi_0, \psi_I\rangle}\,\mathrm{e}^{-\gamma t}.$$

En pratique, il est une fois de plus délicat de résoudre directement (2.7) (du fait de la grande dimension du problème). On recours plutôt à une méthode stochastique : pour ce faire, on interprète (2.7) comme l'équation de Fokker-Planck d'une équation différentielle stochastique (EDS). L'énergie de l'état fondamental est alors estimée en simulant l'EDS associée et en utilisant une formule de Feynman-Kac. Cependant, cette technique n'est pas suffisante en soi, car les estimations que l'on obtient ainsi souffrent d'une trop grande variance. On a alors recours à des techniques de réduction de variance telle que l'échantillonnage d'importance (*importance sampling*) : cela consiste à choisir une fonction test ψ_I telle que $E_L(x) = [H\psi_I](x)/\psi_I(x)$ soit aussi constant que possible (c'est une situation tout à fait analogue à celle rencontrée pour les calculs VMC), faire le changement de fonction inconnnue $\tilde{\phi} = \psi_I\phi$, et résoudre l'équation de diffusion associée à $\tilde{\phi}$ par des méthodes stochastiques.

L'introduction d'une fonction d'importance ψ_I a cependant l'inconvénient que l'équation vérifiée par $\tilde{\phi}$ n'est pas complètement équivalente à l'équation (2.7). Les noeuds $\psi_I^{-1}(0)$ de la fonction d'importance imposent en effet des contraintes supplémentaires, et on ne peut obtenir ainsi qu'une borne supérieure sur l'énergie fondamentale. Cette erreur est connue sous le nom d'approximation des noeuds fixés (*fixed node approximation*) dans la littérature. Une analyse mathématique de la

méthode DMC et de l'approximation des noeuds fixés est présentée par Cancès, Jourdain et Lelièvre dans [50].

Méthodes déterministes

Bien que le problème de minimisation (2.4) posé en grande dimension ne soit pas traitable par des méthodes de gradient usuelles, on peut toutefois tenter d'appliquer de telles méthodes pour obtenir des références précises sur de petits systèmes (qui permettront ensuite de tester la précision de méthodes plus approximatives). Cependant, la méthode de gradient simple fondée sur la minimisation de

$$E(\psi) = \frac{\langle \psi, H\psi \rangle}{\langle \psi, \psi \rangle}$$

conduit à des itérations de la forme

$$\psi_{n+1} = \psi_n + c_n(H - E(\psi_n))\psi_n.$$

Or, cette procédure itérative n'est pas bien posée en général car l'opérateur H est non-borné. Pour remédier à ce problème, Nakatsuji propose d'introduire un opérateur régularisant auto-adjoint S, et de résoudre l'équation de Schrödinger modifiée (*scaled Schrödinger equation*) [254, 255]

$$SH\psi = E_S\, S\psi,$$

où l'énergie fondamentale E_S est obtenue comme

$$E_S = \inf \left\{ \frac{\langle \psi, S^{1/2} H S^{1/2} \psi \rangle}{\langle \psi, S\psi \rangle} \;\middle|\; \psi \in \mathcal{H} \right\}. \tag{2.8}$$

L'opérateur de régularisation est tel que $S^{1/2}HS^{1/2}$ est un opérateur borné, et $S\psi = 0$ implique $\psi = 0$ (par exemple, dans $\mathcal{H}$). On obtient ainsi $E_S = E$, et l'équivalence des problèmes de minimisation (2.8) et (2.4). L'intérêt de la formulation (2.8) est que la méthode de gradient

$$\psi_{n+1} = \psi_n + c_n S^{1/2}(H - E_S(\psi_n))S^{1/2}\psi_n \tag{2.9}$$

est cette fois bien posée. Un autre intérêt de la méthode itérative (2.9) est qu'elle permet d'améliorer de manière systématique des fonctions tests utilisées pour des calculs VMC ou DMC [255].

Une approche plus courante pour obtenir des résultats numériques de référence pour des petits systèmes est la méthode de l'interaction totale des configurations (*full configuration interaction*). Dans ce cas, on considère une base de Galerkin de fonctions $(\phi_1, \ldots, \phi_{N_b})$ de $\mathcal{H}^1$ $(N_b \geq N)$, $\mathcal{I}$ l'ensemble des N-uplets d'éléments distincts de $\{1, \ldots, N_b\}$, et on écrit

$$\psi = \sum_{i \in \mathcal{I}} c_I\, \psi_I,$$

où, pour $I = (i_1, \ldots, i_N)$, ψ_I est le déterminant de Slater $\psi_I = (N!)^{-1/2}\,\mathrm{Det}(\phi_{i_1}, \ldots, \phi_{i_N})$. Le problème de minimisation approché associé

$$E_{\mathrm{FCI}} = \inf \left\{ \langle \psi, H\psi \rangle \;\middle|\; \psi = \sum_{i \in \mathcal{I}} c_I\, \psi_I,\ \|\psi\|_{\mathrm{L}^2} = 1 \right\}$$

donne une borne supérieure de l'énergie fondamentale. Remarquons toutefois que le nombre de déterminants à considérer augmente factoriellement avec N_b, ce qui limite considérablement l'applicabilité de cette méthode.

2.1.3 Matrices densité d'ordre deux

Dès les années 50, des chercheurs comme MAYER [232], LÖWDIN [220] ou COULSON [72] se sont rendus compte que la fonction d'onde n'a pas besoin d'être connue dans toute sa généralité si on cherche simplement à calculer l'énergie d'un système décrit par un Hamiltonien (2.3) ne faisant intervenir que des interactions de paire. En effet,

$$\langle \psi, H\psi \rangle = \mathrm{Tr}(h\gamma) + \frac{1}{2}\int_{\mathbb{R}^3\times\mathbb{R}^3} \frac{\Gamma(x,y\,;\,x,y)}{|x-y|}\,dx\,dy = \mathrm{Tr}(K\Gamma) \tag{2.10}$$

où l'opérateur

$$h_x = -\frac{1}{2}\Delta_x + V(x)$$

est auto-adjoint sur $\mathrm{L}^2(\mathbb{R}^3)$, et l'opérateur à 2 corps

$$K = \frac{1}{2(N-1)}(h_{x_1} + h_{x_2}) + \frac{1}{2|x_1 - x_2|}$$

est auto-adjoint sur $\mathrm{L}^2(\mathbb{R}^3 \times \mathbb{R}^3)$. Les fonctions γ et Γ sont respectivement les matrices densités d'ordre 1 et 2, la matrice densité d'ordre p associée à une fonction d'onde ψ étant définie de manière générale par

$$\begin{aligned}\Gamma^{(p)}&(x_1,\dots,x_p;y_1,\dots,y_p)\\ &= \frac{N!}{(N-p)!}\int_{\mathbb{R}^{3(N-p)}} \overline{\psi}(x_1,\dots,x_p,x_{p+1},\dots,x_N)\psi(y_1,\dots,y_p,x_{p+1},\dots,x_N)\,dx_{p+1}\dots dx_N.\end{aligned} \tag{2.11}$$

On a donc en particulier la relation entre matrices densité d'ordre 1 et 2 :

$$\gamma(x,y) = \frac{1}{N-1}\int_{\mathbb{R}^3} \Gamma(x,x_2;y,x_2)\,dx_2.$$

La formulation (2.10), du problème de minimisation (2.4) montre donc qu'on peut se restreindre à une minimisation sur des fonctions $\Gamma \equiv \Gamma^{(2)}$ de 4 variables. Cependant, on ne connaît pas de conditions nécessaires et suffisantes simples pour assurer qu'une matrice densité d'ordre 2 est obtenue à partir d'une fonction d'onde ψ par la contraction (2.11) avec $p = 2$. Ce problème est connu sous le nom de *problème de la N-représentabilité* des matrices densités d'ordre 2 pour les états purs. Une extension de ce problème consiste à caractériser les matrices densités d'ordre 2 qui sont des combinaisons convexes des opérateurs densité à 2-corps admissibles :

$$\Gamma(x,y) = \sum_{i=1}^{+\infty} n_i\Gamma_i(x,y), \qquad 0 \le n_i \le 1, \qquad \sum_{i=1}^{+\infty} n_i = N,$$

l'opérateur densité à 2-corps Γ_i étant obtenu à partir de fonctions d'onde $\psi_i \in \mathcal{H}^1$ par (2.11) dans le cas $p = 2$. L'ensemble des combinaisons convexes des opérateurs densité à 2-corps est noté $\mathcal{C}_N$ (*ensemble second order density matrices*). Les premiers travaux concernant la N-représentabilité sont ceux de COLEMAN [69], et le récent ouvrage de COLEMAN et YUKALOV [71] décrit la situation actuelle de ce champ de recherche (voir également la Section 7.2). A ce jour, seules des conditions nécessaires de N-représentabilité sont connues ; ces conditions nécessaires sont exprimées comme des (in)égalités linéaires. On obtient donc en pratique des bornes inférieures de l'énergie fondamentale car on minimise sur un espace variationnel trop grand.

D'un point-de-vue numérique, les premiers résultats encourageants ont été obtenus en 1975 par GARROD, MIHAILLOVIC et ROSINA [120], et récemment, de très bons résultats numériques ont été obtenus avec des techniques de programmation semi-définie, telles que les méthodes de point intérieur (NAKATA *et al.* [253]) et des formulations utilisant un Lagrangien augmenté (voir les

articles de Mazziotti [234–236]). Avec E. Cancès et M. Lewin, nous avons proposé dans [P9] une approche duale. En effet, introduisant le Lagrangien augmenté

$$\mathcal{L}(\Gamma, B, \mu) = \mathrm{Tr}(K\Gamma) - \mathrm{Tr}(B\Gamma) - \mu\{\mathrm{Tr}(\Gamma) - N(N-1)\},$$

on montre que

$$E = \inf_{\Gamma} \sup_{B \in (\mathcal{C}_N)^*,\ \mu \in \mathbb{R}} \mathcal{L}(\Gamma, B, \mu)$$

où $\mathcal{C}_N$ est le cône des matrices densité d'ordre 2 admissibles et $(\mathcal{C}_N)^*$ son cône polaire, la minimisation sur Γ étant restreinte aux fonctions symétriques. De manière duale, on a alors

$$E = \inf_{\Gamma} \sup_{B \in (\mathcal{C}_N)^*,\ \mu \in \mathbb{R}} \mathcal{L}(\Gamma, B, \mu) = N(N-1) \sup\{\mu \mid K - \mu \in (\mathcal{C}_N)^*\},$$

la minimisation sur Γ étant également restreinte aux fonctions symétriques (voir Section 7.3). Ainsi, on a ramené le problème de minimisation (2.10) à un problème unidimensionnel. L'implémentation pratique de cette idée utilise un algorithme de Newton pour l'optimisation sur μ, combiné à une boucle interne pour trouver la projection de $K - \mu^n$ sur $(\mathcal{C}_N)^*$ à l'itération n (voir [P9] et l'Algorithme 7.1 en Section 7.3).

2.1.4 Méthodes de fonction d'onde

Les méthodes de fonctions d'onde variationnelles partent d'un *ansatz* sur la forme fonctionnelle de la fonction d'onde ψ, et considèrent alors un problème de minimisation analogue à (2.4), restreint aux fonctions d'onde de la forme donnée par l'*ansatz*. Une des approximations les plus courantes est l'approximation de Hartree-Fock (HF), qui consiste à se restreindre à des fonctions d'ondes qui peuvent s'écrire comme un déterminant de Slater (et sont donc en particulier antisymétriques), c'est-à-dire

$$\psi(x_1, \ldots, x_N) = \frac{1}{\sqrt{N!}} \operatorname{Det}(\phi_i(x_j)), \tag{2.12}$$

où le N-uplet $\Phi = \{\phi_i\}_{i=1,\ldots,N}$ est tel que

$$\phi_i \in \mathrm{H}^1(\mathbb{R}^3), \qquad \int_{\mathbb{R}^3} \phi_i(x)\phi_j(x)\,dx = \delta_{ij}.$$

L'énergie associée à la fonction d'onde (2.12) est

$$\begin{aligned}\langle \psi, H\psi \rangle = E^{\mathrm{HF}}(\Phi) = \frac{1}{2}\sum_{i=1}^{N} \int_{\mathbb{R}^3} |\nabla \phi_i(x)|^2\,dx - \int_{\mathbb{R}^3} V_{\mathrm{nuc}}(x)\rho_\Phi(x)\,dx \\ + \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho_\Phi(x)\rho_\Phi(y)}{|x-y|}\,dx\,dy - \frac{1}{2}\sum_{i=1}^{N}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\,dx\,dy,\end{aligned} \tag{2.13}$$

où on a introduit la matrice densité d'ordre 1 et la densité associées à Φ :

$$\gamma_\Phi(x,y) = \sum_{i=1}^{N} \phi_i(x)\phi_i(y), \qquad \rho_\Phi(x) = \gamma_\Phi(x,x).$$

Le problème de minimisation que l'on obtient finalement est

$$E_{\mathrm{HF}} = \inf\left\{ E^{\mathrm{HF}}(\Phi) \;\middle|\; \Phi = \{\phi_i\}_{i=1,\ldots,N},\ \phi_i \in \mathrm{H}^1(\mathbb{R}^3),\ \int_{\mathbb{R}^3} \phi_i\phi_j = \delta_{ij} \right\}. \tag{2.14}$$

Du fait de l'*ansatz* particulier (2.12), l'espace variationnel est trop petit, et l'énergie HF n'est ainsi qu'une borne supérieure de l'énergie fondamentale (2.4). L'existence d'un minimiseur pour le problème (2.14) lorsque $Z = \sum_{k=1}^{M} Z_k > N-1$ a été montrée par Lieb et Simon [211]. Cependant, on ne sait rien concernant l'unicité du minimiseur (à une transformation orthogonale sur le N-uplet Φ près).

D'un point-de-vue physique, la différence entre l'énergie de l'état fondamental et l'énergie de Hartree-Fock est appelée l'*énergie de corrélation*. En effet, la forme (2.12) de la fonction conduit à faire une hypothèse implicite d'indépendance entre les électrons, compatible avec le principe de Pauli. Lorsque la variable de spin est prise en compte, seuls deux électrons ayant le même spin peuvent être corrélés avec un *ansatz* de type Hartree-Fock, alors que pour la fonction d'onde décrivant l'état fondamental, des électrons de spins différents sont corrélés du fait de l'interaction coulombienne (qui empêche les électrons d'être trop proches les uns des autres).

Tout minimiseur de (2.14) vérifie les équations de Hartree-Fock, qui sont les équations d'Euler-Lagrange associées à (2.14) (après une transformation orthogonale convenable, voir par exemple Cancès, Defranceschi, Kutzelnigg, Le Bris et Maday [53]) :

$$\mathcal{F}_\Phi \phi_i = -\frac{1}{2}\Delta\phi_i + V_{\mathrm{nuc}}\phi_i + \left(\rho_\Phi \star \frac{1}{|x|}\right)\phi_i + K_\Phi \phi_i = \epsilon_i \phi_i. \tag{2.15}$$

Dans cette expression, l'opérateur d'échange K est défini par

$$K_\Phi \varphi(x) = -\int_{\mathbb{R}} \frac{\gamma_\Phi(x,y)}{|x-y|}\varphi(y)\,dy. \tag{2.16}$$

Sous l'hypothèse $Z \geq N$, Lions a montré dans [214] qu'il existe une infinité de solutions du problème aux valeurs propres non-linéaire (2.15). On ne sait pas quelles conditions il faut imposer aux solutions de (2.15) pour qu'elles soient des minimiseurs de (2.14). En revanche, si Φ est un minimiseur de (2.14), alors on sait que les valeurs propres ϵ_i sont les N plus petites valeurs propres de $\mathcal{F}_\Phi$ [214], et que $\epsilon_{N+1} > \epsilon_N$ (voir Bach, Lieb, Loss et Solovej [17]).

D'un point-de-vue numérique, on cherche un point fixe de (2.15), généralement par le biais d'algorithmes auto-consistants ; en effet, même si (2.15) n'est pas équivalent à (2.14), (2.15) est néanmoins plus facile à résoudre numériquement. Une première introduction aux méthodes numériques correspondantes et à l'analyse mathématique de leur convergence est l'ouvrage de Cancès, Le Bris et Maday [52] (voir également [53] pour un traitement plus approfondi).

De nombreuses méthodes ont été proposées et developpées pour améliorer l'approximation de Hartree-Fock. Une classification de ces méthodes, dites post Hartree-Fock, est présentée dans [53], où sont distinguées les approches variationnelles et les approches non-variationnelles. Un exemple d'approche variationnelle est la méthode MCSCF (*multiconfiguration self-consistent field method*), pour laquelle on écrit la fonction d'onde comme une somme (finie) de déterminants de Slater (rappelons en effet que toute fonction d'onde admissible peut être écrite comme une somme infinie de déterminants). L'analyse mathématique de cette méthode a récemment été menée par Friesecke [114] et Lewin [208].

2.1.5 Théorie de la fonctionnelle de la densité

L'idée de Hohenberg et Kohn

Le théorème de Hohenberg et Kohn [161] exprime le fait que la connaissance de la densité électronique de l'état fondamental détermine complètement le potentiel V_{nuc} (à une constante près), et donc la fonction d'onde fondamentale ψ. Ainsi, la minimisation (2.4) sur toutes les fonctions d'onde admissibles peut être remplacée par une minimisation sur toutes les densités admissibles (voir la formule (2.19) ci-dessous). Heuristiquement en effet, on s'attend à ce que la dérivée de la densité électronique présente des singularités aux positions des noyaux atomiques, dont l'intensité

est liée à la charge du noyau en question (les conditions de *cusp* de KATO [190]) ; on peut ainsi retrouver tous les paramètres du potentiel coulombien.

On définit l'énergie électronique d'un système, pour un potentiel extérieur $V \in \mathrm{L}^{3/2}(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3)$ (en fait, $V \equiv V_{\mathrm{nuc}}$ avec les notations employées jusqu'à présent), par

$$E(V) = \inf_{\psi \in \mathcal{H}^1} \left\{ \left\langle \psi, \left(H_0 + \sum_{i=1}^N V(x_i) \right) \psi \right\rangle \right\} = \inf_{\psi \in \mathcal{H}^1} \left\{ \langle \psi, H_0 \psi \rangle + \int_{\mathbb{R}^3} \rho_\psi V \right\}, \tag{2.17}$$

où le Hamiltonien

$$H_0 = \sum_{i=1}^N -\frac{1}{2} \Delta_{x_i} + \sum_{1 \le i < j \le N} \frac{1}{|x_i - x_j|}$$

ne dépend pas de V, et où la densité électronique ρ_ψ associée à ψ est

$$\rho_\psi(x) = N \int_{\mathbb{R}^{3(N-1)}} |\psi(x, x_2, \ldots, x_N)|^2 \, dx_2 \ldots dx_N.$$

Les injections de Sobolev montrent que $\rho_\psi \in \mathrm{L}^1(\mathbb{R}^3) \cap \mathrm{L}^3(\mathbb{R}^3)$. La fonctionnelle définie pour $\rho \in \mathrm{L}^1(\mathbb{R}^3) \cap \mathrm{L}^3(\mathbb{R}^3)$ par

$$F_{\mathrm{L}}(\rho) = \sup_{V \in \mathrm{L}^{3/2}(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3)} \left\{ E(V) - \int_{\mathbb{R}^3} \rho V \right\}, \tag{2.18}$$

a été introduite par LIEB [210]. Notons que F_{L} est convexe, et que l'on peut retrouver l'énergie fondamentale par la formule

$$\boxed{E(V) = \inf_{\rho \in \mathrm{L}^1(\mathbb{R}^3) \cap \mathrm{L}^3(\mathbb{R}^3)} \left\{ F_{\mathrm{L}}(\rho) + \int_{\mathbb{R}^3} \rho V \right\}.} \tag{2.19}$$

Ceci résulte du fait que F_{L} est la transformée de Legendre de E (en effet, $\mathrm{L}^{3/2}(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3)$ est l'espace dual de $\mathrm{L}^1(\mathbb{R}^3) \cap \mathrm{L}^3(\mathbb{R}^3)$ et la fonctionelle E définie par (2.17) est concave [210]). Le fait que la minimisation dans (2.19) porte uniquement sur la densité électronique justifie le nom de théorie de la fonctionnelle de la densité (*density functional theory*, DFT).

Une définition alternative de la fonctionelle de Lieb repose sur l'utilisation de combinaisons convexes d'opérateurs densité à N corps, de la forme

$$\Gamma^{(N)}(x, y) = \sum_{i=1}^{+\infty} n_i \Gamma_i^{(N)}(x, y), \qquad 0 \le n_i \le 1, \qquad \sum_{i=1}^{+\infty} n_i = N,$$

l'opérateur densité à p-corps $\Gamma_i^{(p)}$ étant obtenu à partir de fonctions d'onde $\psi_i \in \mathcal{H}^1$ par (2.11). L'ensemble des combinaisons convexes des opérateurs densité à N-corps est noté $\mathcal{D}^N$ (*ensemble N-particle density operators*). Dans ce formalisme,

$$F_{\mathrm{L}}(\rho) = \inf \left\{ \mathrm{Tr}(H_0 \Gamma^{(N)}), \quad \Gamma^{(N)} \in \mathcal{D}^N, \ \Gamma^{(1)}(x, x) = \rho(x) \right\}.$$

Le fait que cette définition coïncide avec la définition précédente est prouvé dans [210].

Pour obtenir des modèles utilisables en pratique, il faut recourir à des approximations raisonnables de la fonction (inconnue) F_{L}. Pour ce faire,

(i) les méthodes sans orbitales proposent une forme fonctionelle explicite pour $F_{\mathrm{L}} \equiv F_{\mathrm{L}}(\rho)$. Par exemple, le modèle de Thomas-Fermi approche F_{L} par

$$F_{\mathrm{TF}}(\rho) = \frac{10}{3}(3\pi^2)^{2/3} \int_{\mathbb{R}^3} \rho^{5/3} + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}\, dx\, dy;$$

(ii) le modèle de Kohn-Sham prend comme référence un gaz de N électrons non-interagissants, et ρ est alors la somme des densités électroniques de chaque électron.[3]

Implémentation pratique de la théorie de la fonctionelle de la densité par le modèle de Kohn-Sham

Dans la majorité des calculs pratiques, la DFT est implémentée selon le modèle de Kohn et Sham (KS) [195]. Partant d'un gaz d'électrons non-interagissants, on commence par approximer l'opérateur Hamiltonien H_0 par sa partie cinétique $T = -\frac{1}{2}\sum_{i=1}^N \Delta_{x_i}$. L'énergie associée à T est donnée par la fonctionelle d'énergie cinétique de Janak

$$\begin{aligned} T_{\mathrm{J}}(\rho) &= \inf\left\{ \mathrm{Tr}(H_0 \Gamma^{(N)}), \quad \Gamma^{(N)} \in \mathcal{D}^N,\ \Gamma^{(1)}(x,x) = \rho(x) \right\}, \\ &= \inf\left\{ \frac{1}{2}\sum_{i=1}^{+\infty} n_i \int_{\mathbb{R}^3} |\nabla \phi_i|^2, \quad \phi_i \in \mathrm{H}^1(\mathbb{R}^3),\ \int_{\mathbb{R}^3} \phi_i \phi_j = \delta_{ij},\ 0 \le n_i \le 1, \right. \\ &\qquad\qquad \left. \sum_{i=1}^{+\infty} n_i = N,\ \sum_{i=1}^{+\infty} n_i |\phi_i|^2 = \rho \right\}. \end{aligned}$$

Cette approche est celle du modèle de Kohn-Sham étendu, et autorise des nombres d'occupation n_i fractionnaires. La fonctionelle T_{J} ci-dessus est définie pour des densités ρ ensemble N-représentables, c'est-à-dire provenant de la contraction d'un opérateur densité de $\mathcal{D}^N$. Coleman [69] a montré que l'ensemble des densités ensemble N-représentables d'énergie cinétique finie est

$$\mathcal{I}_N = \left\{ \rho \ge 0,\ \sqrt{\rho} \in \mathrm{H}^1(\mathbb{R}^3),\ \int_{\mathbb{R}^3} \rho = N \right\}.$$

L'énergie électrostatique est approximée par l'énergie de Coulomb

$$J(\rho) = \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}\, dx\, dy.$$

Enfin, les erreurs provenant des approximations ci-dessus des énergies électrostatique et cinétique sont compensées par l'énergie dite d'*échange-corrélation* :

$$E_{\mathrm{xc}}(\rho) = F_{\mathrm{L}}(\rho) - T_{\mathrm{KS}}(\rho) - J(\rho). \tag{2.20}$$

Le modèle de Kohn-Sham étendu est donc le problème de minimisation suivant :

$$\begin{aligned} E^{\mathrm{KS}}(V) = \inf\Bigg\{ &\frac{1}{2}\sum_{i=1}^{+\infty} n_i \int_{\mathbb{R}^3} |\nabla \phi_i|^2 + \int_{\mathbb{R}^3} \rho V + \frac{1}{2} \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \frac{\rho(x)\rho(y)}{|x-y|}\, dx\, dy + E_{\mathrm{xc}}(\rho), \\ &\phi_i \in \mathrm{H}^1(\mathbb{R}^3),\ \int_{\mathbb{R}^3} \phi_i \phi_j = \delta_{ij},\ 0 \le n_i \le 1,\ \sum_{i=1}^{+\infty} n_i = N,\ \sum_{i=1}^{+\infty} n_i |\phi_i|^2 = \rho \Bigg\}. \end{aligned} \tag{2.21}$$

Si E_{xc} est différentiable dans $\mathcal{I}_N$ autour de $\rho \in \mathcal{I}_N$, et notant $v_{\mathrm{xc}}(\rho)$ sa dérivée fonctionnelle, les équations d'Euler-Lagrange associées à (2.21) sont les équations de Kohn-Sham (étendues) :

$$-\frac{1}{2}\Delta \phi_i(x) + V(x)\phi_i(x) + \left(\int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}\, dy \right) \phi_i(x) + v_{\mathrm{xc}}(\rho)\phi_i(x) = \epsilon_i \phi_i(x), \tag{2.22}$$

[3] Ce qui explique *a posteriori* pourquoi les modèles de type Thomas-Fermi sont appelés modèles sans orbitales...

avec les contraintes $\int_{\mathbb{R}^3} \phi_i \phi_j = \delta_{ij}$, et $n_i = 1$ si $\epsilon_i < \varepsilon_F$, $0 \leq n_i \leq 1$ si $\epsilon_i = \varepsilon_F$, $n_i = 0$ si $\epsilon_i > \varepsilon_F$. Le multiplicateur de Lagrange ε_F de la contrainte $\sum_{i=1}^{+\infty} n_i = N$ est appelé le niveau de Fermi. Les équations de Kohn-Sham usuelles

$$-\frac{1}{2}\Delta\phi_i(x) + V(x)\phi_i(x) + \left(\int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}\,dy\right)\phi_i(x) + v_{\mathrm{xc}}(\rho)\phi_i(x) = \epsilon_i\phi_i(x), \tag{2.23}$$

avec $n_i = 1$ si $1 \leq i \leq N$, et $n_i = 0$ sinon, correspondent au cas où seuls des nombres d'occupation entiers sont admis. L'existence d'un minimiseur du problème de minimisation associé à (2.23) (et donc d'une solution normalisée de (2.23)) a été prouvée par Le Bris [42] pour certaines approximations usuelles de v_{xc}.

Rappelant à ce point que le potentiel V utilisé ici est le potentiel extérieur (par exemple, le potentiel V_{nuc} engendré par les noyaux atomiques), on s'aperçoit que les équations de Kohn-Sham sont formellement similaires aux équations de Hartree-Fock (2.15). La différence entre ces équations est que l'opérateur d'échange non-local des équations de Hartree-Fock est remplacé dans les équations de Kohn-Sham par un potentiel d'échange-corrélation local. Cette similarité a été utilisée dans les premiers temps de la chimie quantique computationnelle pour simplifier les équations de Hartree-Fock, en replaçant le potentiel local par sa "meilleure" approximation. La qualité de cette approximation doit être comprise en un sens variationnel, et est connue sous le nom de potentiel effectif optimisé (*Optimized Effective Potential* (OEP), voir ci-après et le Chapitre 8 pour plus de précisions).

Fonctionnelles d'échange-corrélation

L'approximation la plus simple de $E_{\mathrm{xc}}(\rho)$ est l'approximation de la densité locale (*local density approximation*, LDA), fondée sur le modèle du gaz d'électron homogène. On a dans ce cas

$$E_{\mathrm{xc}}^{\mathrm{LDA}}(\rho) = \int_{\mathbb{R}^3} \rho(x)\varepsilon_{\mathrm{xc}}^{\mathrm{LDA}}(\rho(x))\,dx,$$

où $\varepsilon_{\mathrm{xc}}^{\mathrm{LDA}} = \varepsilon_{\mathrm{x}}^{\mathrm{LDA}} + \varepsilon_{\mathrm{c}}^{\mathrm{LDA}}$. La partie correspondant au terme d'échange électronique peut être calculée analytiquement : $\varepsilon_{\mathrm{x}}^{\mathrm{LDA}}(\rho) = -C_D\rho^{4/3}$, où $C_D = \frac{3}{4}(\frac{3}{\pi})^{1/3}$ est la constante de Dirac. En revanche, la partie correspondant aux corrélations électroniques doit être approximée, par exemple par des techniques de Monte-Carlo quantique. Une amélioration de ce modèle consiste à prendre en compte des densités électroniques $\rho_{|\uparrow\rangle}$ et $\rho_{|\downarrow\rangle}$ dépendant de la variable de spin, ainsi que des corrections modélisant les inhomogénéités de la densité par des termes de gradients (d'où le nom d'approximation du gradient généralisée, *Generalized Gradient Approximation*). La recherche de fonctionnelles d'échange-corrélation de bonne qualité et applicables à une grande variété de systèmes est un champ de recherche toujours actif en physique et chimie (voir par exemple la revue de Scuseria et Staroverov [304]).

Recherche variationnelle de potentiels d'échange-corrélation pertinents

Sharp et Horton [308] ont proposé une manière systématique d'obtenir des potentiels locaux approchant au mieux l'opérateur d'échange non local de Hartree-Fock K_Φ donné par (2.16). Ils ont suggéré de minimiser l'énergie du déterminant de Slater construit à partir des fonctions propres correspondant aux N premières valeurs propres d'un opérateur de Schrödinger à un électron $-\frac{1}{2}\Delta + W$, W étant un "potentiel local".[4] Cette stratégie a ensuite été approfondie par Talman et

[4] La notion de "potentiel local" n'a pas de définition précise dans la littérature en chimie ou en physique. Cela dit, on peut se restreindre, pour cette introduction, à des opérateurs multiplicatifs, pour un potentiel $W \in \mathrm{L}^{3/2}(\mathbb{R}^3) + \mathrm{L}^\infty_\varepsilon(\mathbb{R}^3)$. Rappelons que dans ce cas, le spectre essentiel de l'opérateur

Shadwick [338]. Le problème de minimisation correspondant est la recherche du "potentiel effectif optimal" (*Optimized Effective Potential*, OEP) qui peut être énoncé de manière vague comme

$$\inf_W \left\{ E^{\mathrm{HF}}(\phi_1^W, \dots, \phi_N^W) \;\middle|\; \int_{\mathbb{R}^3} \phi_i^W \phi_j^W = \delta_{ij},\ (\phi_1^W, \dots, \phi_N^W) \text{ sont les premiers vecteurs propres correspondant aux } N \text{ premières valeurs propres de } H_W = -\frac{1}{2}\Delta + W \right\}. \tag{2.24}$$

Cependant, formulé de cette manière, ce problème de minimisation ne semble pas être bien posé car il n'y a pas de borne évidente pour les suites minimisantes $(W_n)_{n\geq 0}$ (voir l'étude mathématique de Ben-Haj-Yedder, Cancès et Le Bris [25]). Une manière de contourner cette difficulté est de remplacer le problème de minimisation (2.24) par des conditions équivalentes (au moins formellement) qui ne font pas intervenir explicitement un potentiel local W. Dans ce cas, on peut obtenir quelques résultats mathématiques concernant le caractère bien posé du problème (voir [25], ainsi que la Section 8.2).

A côté de ces questions mathématiques, se posent également des problèmes numériques dans le calcul du potentiel effectif optimal lorsque l'on discrétise le problème de minimisation dans une base d'orbitales (voir par exemple [321]). Il est donc tentant de remplacer le problème de minimisation (2.24) par un problème de minimisation plus simple, qui assure cependant toujours que le potentiel local d'échange à considérer approche en un certain sens l'opérateur d'échange non local (2.16). Avec E. Cancès, E. Davidson, A. Izmaylov et G. Scuseria [P5,A2], nous avons montré qu'il est possible de définir de manière unique (à une constante additive près toutefois) un potentiel effectif local (*Effective Local Potential*, ELP) qui est tel que

$$v_{\mathrm{ELP}} = \operatorname*{arginf}_{v \in \mathrm{L}^3(\mathbb{R}^3)+\mathrm{L}^\infty(\mathbb{R}^3)} \left\{ \frac{1}{2} \left\| [v - K_\Phi, \gamma_\Phi] \right\|_{\mathrm{HS}}^2 \right\}, \tag{2.25}$$

où $\|\cdot\|_{\mathrm{HS}}$ est la norme de Hilbert-Schmidt pour des opérateurs sur $\mathrm{L}^2(\mathbb{R}^3)$, et $[A, B] = AB - BA$. L'étude mathématique du caractère bien posé du problème (2.25) est menée à la Section 8.3.

On peut montrer que le potentiel effectif local a une forme analytique simple, ce qui est très intéressant en pratique. Notons également que ce potentiel n'est pas nouveau, et a déjà été proposé par d'autres auteurs [138, 297], par des arguments non-variationnels toutefois. Nous n'avons cependant pas pu montrer l'existence d'une solution des équations de Kohn-Sham que pour un potentiel d'échange local plus simple que v_{ELP} (voir la Section 8.1). Ce potentiel a été proposé par Slater [312], mais approche également le potentiel d'échange non local au sens d'un problème variationnel en norme Hilbert-Schmidt analogue à (2.25).

2.2 Description classique de la matière

2.2.1 Représentation classique de la matière à l'échelle microscopique

On considère dans cette section des systèmes microscopiques formés de N particules (typiquement, des atomes), décrits par les positions $q = (q_1, \cdots, q_N) \in \mathbb{R}^{3N}$ et les impulsions $p = (p_1, \cdots, p_N) \in \mathbb{R}^{3N}$ de ces particules. Pour les systèmes physiques et biologiques étudiés à l'heure actuelle, N est typiquement de l'ordre de 10^3 à 10^9. L'interaction entre les particules est prise en compte par le biais d'un potentiel $V \equiv V(q)$, et l'énergie totale du système est donnée par le Hamiltonien

$-\frac{1}{2}\Delta + W$ est encore $[0, +\infty[$ [52, 277]. L'espace fonctionnel $\mathrm{L}^{3/2}(\mathbb{R}^3) + \mathrm{L}^\infty_\varepsilon(\mathbb{R}^3)$ est l'ensemble de toutes les fonctions ϕ qui sont telles que, pour tout $\varepsilon > 0$, on peut écrire $\phi = \phi_{3/2} + \phi_\infty$ avec $\phi_{3/2} \in \mathrm{L}^{3/2}(\mathbb{R}^3)$ et $\|\phi_\infty\|_{\mathrm{L}^\infty(\mathbb{R})} \leq \varepsilon$.

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q), \tag{2.26}$$

où $M = \mathrm{Diag}(m_1, \ldots, m_N)$ est la matrice de masse.

Energie potentielle

Le potentiel d'interaction V est, en principe, obtenu par (2.5). C'est même ainsi qu'est calculé le potentiel d'interaction pour des simulations moléculaires *ab-initio* : dans ce cas, à chaque itération (à chaque modification de la géométrie des noyaux), on recalcule le potentiel d'interaction par (2.5).

Cette approche est toutefois très coûteuse en termes de temps de calcul, et ne peut être appliquée qu'à de petits systèmes. Pour simuler de plus grands systèmes, on utilise généralement des formules empiriques. Ces formules empiriques sont typiquement obtenues en supposant une forme fonctionnelle *a priori* pour le potentiel d'interaction, et en optimisant alors les paramètres entrant dans la formule analytique du potentiel pour obtenir le meilleur accord entre les résultats de simulation numérique et les données de références. Ces données de référence (loi d'état, compressibilité, vitesse du son,...) peuvent être soit expérimentales, soit obtenues numériquement avec un modèle plus précis, notamment des simulations *ab-initio* sur de petits systèmes à l'équilibre. Dans ce dernier cas, l'approximation de Born-Oppenheimer utilisée implicitement pour écrire le potentiel d'interaction entre particules comme (2.5) peut ne plus être valide. Ceci est le cas par exemple lorsque des réactions chimiques ont lieu, et que des liaisons sont créées ou coupées (mais certaines approches cherchent toutefois à modéliser ces événements dans le cadre des potentiels classiques, voir ci-dessous).

Un exemple simple de potentiel d'interaction est celui d'un système de N particules interagissant additivement via un potentiel d'interaction de paire. Dans ce cas,

$$V(q_1, \ldots, q_N) = \sum_{1 \le i < j \le N} \mathcal{V}(|q_i - q_j|). \tag{2.27}$$

Par exemple, le fluide d'argon est bien décrit par un potentiel de Lennard-Jones

$$\mathcal{V}(r) = \epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right), \tag{2.28}$$

avec $\epsilon/k_{\mathrm{B}} = 120$ K, et $\sigma = 3.405$ Å. Des termes d'interactions entre plus que deux atomes peuvent être considérés, particulièrement dans les modèles de biologie moléculaire. Ces termes permettent de prendre en compte les interactions locales entre les nuages électroniques des atomes (interactions à 3-corps par les angles de liaisons, interactions à 4-corps par les angles dihèdres, voir Section 3.4.1 pour des expressions explicites de ces potentiels d'interaction) et non-locales entre atomes non liés (forces de van der Waals, interactions coulombiennes) – voir par exemple l'ouvrage de Schlick [299] pour plus de précisions sur les modèles utilisés en biologie moléculaire.

Considérer des potentiels d'interaction à 2-, 3- ou 4-corps n'est toutefois pas toujours suffisant. De nombreuses études, en particulier dans le domaine de la matière condensée, ont encore pour objectif de proposer de meilleurs potentiels d'interaction empiriques, et d'optimiser les paramètres correspondants sur une base de données de référence. Des exemples récents sont les potentiels de type (M)EAM ((*Modified) Embedded-Atom Model*) [22], qui utilisent une densité électronique de référence autour de l'atome, et les potentiels à ordre de liaison, de type REBO [341] ou ReaxFF [353], qui contiennent des termes dépendant de l'environnement de la particule (fonctions de la coordination locale). Ces deux derniers types de potentiels permettent même de traiter des dissociations chimiques dans certains régimes physiques.

Conditions de bord

On peut imposer plusieurs types de conditions de bord au système :

(i) de nombreuses simulations sont faites à l'heure actuelle avec des conditions de bord périodiques, qui limitent les effets de surface dans la simulation et permettent d'approcher les conditions règnant au sein d'un échantillon macroscopique homogène de matériau. Dans ce cas, chaque particule interagit avec toutes les particules du système, mais aussi avec leurs images périodiques ;

(ii) certaines simulations sont faites avec des conditions au bord libres, comme dans le cas de systèmes isolés (molécules simulées *in vacuo*). Il peut être utile de quotienter la dynamique par les transformations galiléennes (translations, rotations), dans la mesure où, en l'absence de champ de force extérieur, l'énergie potentielle du système est invariante par ces transformations ;

(iii) il est parfois intéressant de simuler des systèmes confinés. Dans ce cas, les positions accessibles aux particules sont restreintes à une région prédéfine de l'espace ambiant, et on doit se donner des règles pour traiter les réflections sur le bord (réflection spéculaire des impulsions par exemple) ;

(iv) finalement, on peut considérer un terme de forçage (stochastique ou déterministe) à la frontière (voir Section 3.5.1).

On note dans la suite $\mathcal{M}$ l'espace des positions, et $T^*\mathcal{M}$ son espace cotangent, qui est l'ensemble des configurations microscopiques (q,p) accessibles au système (on parle également d'*espace des phases* en physique). Typiquement, $\mathcal{M} = \mathbb{T}^{3N}$ (un tore de dimension $3N$) lorsque l'on simule N particules avec des conditions de bord périodiques. Dans ce cas, $T^*\mathcal{M} = \mathbb{T}^{3N} \times \mathbb{R}^{3N}$.

Ensembles thermodynamiques

L'état (macroscopique) d'un système est décrit, dans le contexte de la physique statistique, par une mesure de probabilité μ sur l'espace des phases $T^*\mathcal{M}$. Les propriétés macroscopiques du système sont alors calculées comme des moyennes par rapport à cette mesure, selon (2.1) :

$$\langle A \rangle = \int_{T^*\mathcal{M}} A(q,p)\, d\mu(q,p).$$

On s'intéresse par la suite aux deux ensembles thermodynamiques les plus couramment utilisés dans la pratique, à savoir les ensembles microcanonique et canonique, qui décrivent respectivement des systèmes isolés et des systèmes à température fixée (en contact avec un thermostat ou un réservoir d'énergie).

2.2.2 L'ensemble microcanonique

L'ensemble thermodynamique le plus simple conceptuellement est l'ensemble microcanonique, qui décrit les systèmes isolés, donc d'énergie constante. La mesure correspondante est la mesure de probabilité uniforme sur les configurations accessibles, à savoir

$$\mu_{\mathrm{mc}}(dq,dp) = \delta_{H(q,p)-E} = \frac{d\sigma_E}{|\nabla H|}, \tag{2.29}$$

où $d\sigma_E$ est la mesure de surface induite par la mesure de Lebesgue sur la sous-variété $\mathcal{M}(E) = \{(q,p) \in T^*\mathcal{M} \mid H(q,p) = E\}$. Les intégrales thermodynamiques de la forme (2.1) sont calculées en pratique en utilisant une hypothèse d'ergodicité de la dynamique sous-jacente :

$$\boxed{\langle A \rangle = \lim_{T \to +\infty} \frac{1}{T} \int_0^T A(\Phi_t(q,p))\, dt,} \tag{2.30}$$

où Φ_t est le flot de la dynamique hamiltonienne associée au Hamiltonien (2.26), à savoir :

$$\begin{cases} \dot{q}_i(t) = \dfrac{\partial H}{\partial p_i}(q(t),p(t)) = \dfrac{p_i(t)}{m_i}, \\ \dot{p}_i(t) = -\dfrac{\partial H}{\partial q_i}(q(t),p(t)) = -\nabla_{q_i} V(q(t)). \end{cases} \tag{2.31}$$

L'ergodicité peut être démontrée rigoureusement pour des systèmes complètement intégrables, ainsi que pour des perturbations de systèmes complétement intégrables (voir par exemple l'ouvrage de référence d'Arnol'd [11]).

D'un point de vue numérique, la propriété d'ergodicité demande des algorithmes très stables, permettant une intégration en temps très long. La dynamique (2.31) est une équation différentielle ordinaire qui est souvent intégrée numériquement par l'algorithme de Verlet[5] [360]

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla V(q^n), \\ q^{n+1} = q^n + \Delta t\ M^{-1}p^{n+1/2}, \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}), \end{cases} \tag{2.32}$$

où $\Delta t > 0$ est la pas de temps choisi. Le flot numérique associé au schéma de Verlet partage deux propriétés qualitatives avec le flot exact Φ_t de à (2.31) : il est *réversible en temps* et *symplectique*. Ces deux propriétés sont très importantes pour l'intégration numérique en temps long de la dynamique hamiltonienne : un résultat bien établi, et rappelé dans l'ouvrage de référence de Hairer, Lubich et Wanner [146] sur l'intégration numérique géométrique (voir en particulier les Chapitres VIII et IX), est que l'énergie du système est conservée à $\mathrm{O}(\Delta t^2)$ sur des temps $\mathrm{O}(\mathrm{e}^{-c/\Delta t})$ lorsque l'on utilise un schéma de Verlet. L'analyse numérique de méthodes d'échantillonnage de l'ensemble microcanonique fondées sur ces propriétés (dans le cas très particulier de systèmes complètement intégrables) a été faite par Cancès, Castella, Chartier, Le Bris, Legoll, Faou et Turinici [48, 49, 203].

2.2.3 L'ensemble canonique

Les systèmes à température fixée (en particulier, les systèmes en contact avec un thermostat) sont décrits par la mesure de probabilité canonique sur $T^*\mathcal{M}$:

$$d\mu(q,p) = Z^{-1}\exp(-\beta H(q,p))\,dq\,dp, \tag{2.33}$$

où $\beta = 1/k_{\mathrm{B}}T$ (T est la température du système et k_{B} la constante de Boltzmann). La constante Z dans (2.33) est une constante de normalisation, donnée par

$$Z = \int_{T^*\mathcal{M}} \exp(-\beta H(q,p))\ dq\,dp.$$

On l'appelle également *fonction de partition* en physique statistique. Comme le Hamiltonien H est séparable, la mesure canonique peut se mettre sous la forme tensorisée

$$d\mu(q,p) = d\pi(q)\,d\kappa(p),$$

avec

$$d\kappa(p) = Z_p^{-1}\exp\left(-\frac{\beta}{2}p^T M^{-1}p\right)\,dp, \tag{2.34}$$

[5] Voir aussi [145] pour plus de précisions historiques : l'algorithme de Verlet était déjà connu par Störmer au début du 20ème siècle, et même déjà par Newton!

et

$$d\pi(q) = Z_q^{-1} \mathrm{e}^{-\beta V(q)}\, dq. \tag{2.35}$$

Les constantes $Z_q = \int_{\mathcal{M}} \mathrm{e}^{-\beta V(q)}\, dq$ et $Z_p = (2\pi/\beta)^{3N/2} \prod_{i=1}^N m_i^{3/2}$ sont à nouveau des constantes de normalisation. Notons au passage que l'on fait l'hypothèse implicite que les mesures μ et π sont des mesures de probabilité, ce qui est le cas lorsque $\mathrm{e}^{-\beta V} \in L^1(\mathcal{M})$. L'échantillonnage de la mesure $d\kappa$ ne posant pas de problème, la vraie question est d'échantillonner la mesure $d\pi$.

Comparaison théorique et numérique de quelques méthodes d'échantillonnage usuelles

Des méthodes numériques permettant d'engendrer des configurations $(q^n, p^n)_{n\geq 0}$ telles que

$$\lim_{N\to+\infty} \frac{1}{N} \sum_{n=0}^{N-1} A(q^n, p^n) = \int_{T^*\mathcal{M}} A(q,p)\, d\mu(q,p) \tag{2.36}$$

sont présentées dans l'article de revue [P3] co-écrit avec E. Cancès et F. Legoll (voir également le Chapitre 3). En particulier, nous proposons une classification de méthodes usuelles d'échantillonnage de la mesure canonique en trois catégories, et précisons leurs propriétés théoriques d'ergodicité. On distingue ainsi

(i) les méthodes purement stochastiques, telles que la méthode du rejet ou l'utilisation de fonctions d'importance, dont la convergence résulte des théorèmes usuels de probabilité (Loi des Grands Nombres (LGN), Theorème de la Limite Centrale (TCL)) ;

(ii) des méthodes fondées sur la dynamique déterministe hamiltonienne, modifiée par des perturbations stochastiques pour assurer que des niveaux d'énergie différents sont explorés. On peut utiliser des chaînes de Markov, par exemple un schéma de Metropolis-Hastings [153, 238] utilisant la dynamique hamiltonienne comme fonction de proposition (schéma de Hybrid Monte-Carlo [88]), ou des équations différentielles stochastiques ayant la dynamique hamiltonienne comme cas limite (dynamique de Langevin). On s'arrange pour que la mesure canonique soit préservée par ces schémas, et on obtient, sous des hypothèses assez générales, l'équivalent de la LGN et du TCL pour des chaînes ou processus de Markov, ce qui assure l'ergodicité théorique de ces méthodes (voir en particulier l'excellent ouvrage de Meyn et Tweedie [240] en ce qui concerne les résultats sur la convergence des chaînes de Markov, ainsi que la Section 3.6 pour un résumé des résultats les plus pertinents dans le contexte de la dynamique moléculaire) ;

(iii) des méthodes complètement déterministes fondées sur le paradigme de la dynamique de Nosé-Hoover [259, 260], et qui considèrent des variables étendues (q, p, x), dont la dynamique est telle que la marginale de la mesure invariante par rapport à la variable supplémentaire x soit bien la mesure canonique. Malgré ce résultat de consistance, aucune preuve théorique d'ergodicité n'existe ; en revanche, il existe des résultats de non-ergodicité dans certains cas (voir la preuve de Legoll, Luskin et Moeckel [204] fondée sur une perturbation de systèmes complètement intégrables).

Nous avons également comparé l'ergodicité numérique de ces méthodes pour une molécule d'alcane simple, tant pour le calcul de propriétés statiques (intégrales thermodynamiques de la forme (2.1)) que de propriétés dépendant du temps comme des fonctions d'autocorrélations (voir Section 2.2.5). Les résultats montrent, comme on pouvait d'ailleurs qualitativement s'y attendre, que les méthodes purement stochastiques marchent mal dès que la dimension du système est un peu trop grande, et que les méthodes complètement déterministes peuvent être délicates d'emploi (choix des paramètres, nécessité de petits pas de temps pour assurer la conservation de certains invariants de la dynamique), alors que les méthodes mêlant dynamique moléculaire et techniques stochastiques s'avèrent plus robustes et plus efficaces.

Métastabilité et obstruction à l'ergodicité numérique

Même si l'ergodicité est assurée théoriquement et numériquement pour des systèmes simples, ce n'est souvent pas le cas en pratique pour des systèmes physiquement intéressants, du fait de la présence d'échelles de temps très différentes dans le système. Les échelles de temps rapides sont typiquement associées à des composantes raides de l'énergie potentielle, et demandent, pour être résolues, des pas d'intégration temporelle très petits. Par exemple, les longueurs de liaison dans une molécule ont une période de vibration de l'ordre de la femtoseconde (10^{-15} s), alors que d'autres quantités (telles que la conformation d'une protéine) évoluent sur des échelles de temps beaucoup plus longues. Les échelles de temps grandes proviennent souvent de la présence d'états métastables, qui sont des minima locaux de la surface d'énergie. Les événements intéressants tel que le repliement des protéines ne se produisent que lorsque le système a exploré plusieurs bassins d'énergie, ce qui peut demander des temps de l'ordre de la microseconde (10^{-6} s) ou plus [299].

Une manière de traiter les métastabilités du système, lorsqu'elles sont identifiées, est de découpler les variables ayant un comportement métastable et les autres degrés de liberté du système, puis d'utiliser des techniques de calcul de différences d'énergie libre (voir Section 2.3.1). Il existe bien sûr de nombreuses autres façons de faire, comme les dynamiques accélérées de la Section 2.3.2, ainsi que des méthodes permettant d'identifier les états métastables (fondées sur les propriétés spectrales des noyaux de transition d'une chaîne de Markov et initiées par les travaux de Schütte [301]). Il n'existe toutefois pas à ce jour de méthode générale et robuste permettant de traiter de manière systématique des potentiels complexes, comme ceux régissant les systèmes biologiques.

2.2.4 Autres ensembles thermodynamiques

Outre les ensembles canonique et microcanonique, il existe plusieurs autres ensembles thermodynamiques courants, par exemple l'ensemble NPT, où le nombre de particules, la pression et la température sont maintenus constants, ou l'ensemble grand-canonique, où le volume, la température et le nombre *moyen* de particules sont constants. Ce dernier ensemble est aussi appelé ensemble μVT, μ étant le potentiel chimique.[6] La mesure de probabilité grand-canonique est [270]

$$d\nu(N, q^N, p^N) = Z^{-1} \frac{1}{h^{3N} N! V^N} \, \mathrm{e}^{\beta(\mu N - H_N(q^N, p^N))} \, dq^N \, dp^N, \tag{2.37}$$

où d est la dimension de l'espace, V le volume de la cellule de simulation, h la constante de Planck, et H_N le Hamiltonien (2.26) pour N particules en interaction. La constante de normalisation Z est (notant $T^*\mathcal{M}_N$ l'espace cotangent de la variété des positions accessibles $\mathcal{M}_N$)

$$Z = \sum_{N=0}^{\infty} \frac{1}{\Lambda^N N! V^N} \, \mathrm{e}^{\beta\mu N} \int_{T^*\mathcal{M}_N} \mathrm{e}^{-\beta H_N(q^N, p^N)} \, dq^N \, dp^N,$$

où $\Lambda = h(2\pi m\beta^{-1})^{-1/2}$ est la longueur d'onde thermique de De Broglie. Les premières techniques développées pour l'échantillonnage de la mesure (2.37) ont été des techniques purement stochastiques [258] (voir [113, Chapitre 5] pour plus de références). Les techniques d'échantillonnage de la mesure canonique ont ensuite été progressivement transposées au contexte grand-canonique, en particulier les approches de type Hybrid Monte-Carlo [218] ou Nosé-Hoover [56, 57, 216] (voir [296, Chapitre 8] pour plus de précisions et de références sur ces méthodes).

Il se peut également que l'on considère un système régi par la dynamique hamiltonienne, à des termes de forçage au bord près (voir Section 3.5). Par exemple, on peut considérer une création ou destruction de particules, ou une thermalisation sur le bord du domaine. Dans ces cas, les quantités

[6] Dans cette section et dans cette section seulement, la notation μ renvoie au potentiel chimique, et non à une mesure de probabilité sur l'espace des phases.

préservées par la dynamique (exactement ou en moyenne) ne sont pas toujours clairement établies, et l'ensemble thermodynamique à utiliser n'est pas toujours bien défini *a priori*.

2.2.5 Propriétés dépendant du temps

Les propriétés dépendant du temps sont de la forme

$$\langle B \rangle(t) = \int_{T^*\mathcal{M}} B(\Phi_t(q,p),(q,p))\, d\mu, \tag{2.38}$$

où Φ_t est le flot de la dynamique utilisée pour engendrer les trajectoires $(q(t),p(t))_{t\geq 0} = (\Phi_t(q,p))_{t\geq 0}$. Celles-ci peuvent être calculées selon la dynamique hamiltonienne associée à (2.26). Ce choix est consistant avec des conditions initiales microcaniques, et même canoniques, car la mesure canonique (2.33) est invariante par la dynamique hamiltonienne (2.31).

Les coefficients de transport sont des exemples de propriétés dynamiques. L'auto-diffusion d'une particule marquée dans un fluide de N particules identiques de masse m est ainsi donnée par la relation d'Einstein [276]

$$D = \lim_{t\to+\infty} \frac{1}{6Nt} \left\langle \sum_{i=1}^{N} |q_i(t) - q_i(0)|^2 \right\rangle,$$

où $q_i(t)$ est la position de la i-ème particule au temps t, et $\langle \cdot \rangle$ est une moyenne sur les conditions initiales. Une expression alternative est la formule de Green-Kubo utilisant la fonction d'autocorrélation des impulsions [276] :

$$D = \frac{1}{3Nm^2} \int_0^{+\infty} \left\langle \sum_{i=1}^{N} p_i(t)\cdot p_i(0) \right\rangle dt,$$

où $p_i(t)$ est l'impulsion de la i-ème particule au temps t. D'autres exemples courant de coefficients de transport sont la viscosité de cisaillement d'un fluide, ou sa diffusivité thermique.

Un calcul précis d'intégrales thermodynamiques dépendant du temps demande dans un premier temps un bon échantillon de conditions initiales, distribuées selon la mesure canonique typiquement. Ces configurations initiales ne doivent pas être trop nombreuses, car il faut pouvoir intégrer numériquement une ou plusieurs trajectoires pour chacune. Le coût de calcul d'une trajectoire sur un intervalle de temps $[0,T]$ est en $(\Delta t)^{-1}$, ce qui donne un coût total en $\mathrm{O}(M(\Delta t)^{-1})$ pour M configurations initiales. Il y a donc un compromis à faire entre la qualité de l'échantillon initial tiré selon $d\mu$ (erreur de l'ordre de $M^{-1/2}$) et la précision de la trajectoire numérique approchant la dynamique (2.31) (liée au paramètre Δt).

En pratique, certaines études calculent des propriétés dépendant du temps (à température constante) comme des moyennes le long d'une trajectoire sous une hypothèse implicite d'ergodicité. Il n'est pas clair que cette procédure soit correcte, car la dynamique en question est soit hamiltonienne, auquel cas on échantillonne mal les conditions initiales, soit consistante avec la mesure canonique (dynamique de Langevin ou Nosé-Hoover), et il n'est alors pas clair que le résultat final soit indépendant des paramètres numériques choisis (masse de Nosé, paramètre de friction dans la dynamique de Langevin). Par exemple, le coefficient d'auto-diffusion d'une particule dans un fluide dépend *a priori* de la friction choisie pour la dynamique de Langevin.

Dans tous les cas cependant, il est rare que les systèmes conservent leur énergie sur des temps très longs, du fait d'échanges avec leur environnement. Utiliser une dynamique hamiltonienne partant de conditions canoniques semble justifié pour le calcul de propriétés dépendant du temps sur des temps courts seulement. Pour les temps longs, on peut penser à considérer des systèmes dont la partie centrale évolue selon une dynamique hamiltonienne (on calcule les propriétés dans cette région seulement), mais dont les parties proches de la frontière du domaine subissent des

perturbations stochastiques, modélisant les perturbations extérieures (voir Section 3.5). Comme la procédure de thermalisation n'influence pas directement la dynamique, on s'attend à une dépendance moins forte en fonction des paramètres choisis.

2.3 Simuler des systèmes plus grands pendant des temps plus longs

Les techniques de simulation moléculaire présentées dans les sections précédentes ne permettent que de simuler des systèmes très petits comparés aux systèmes réels, et ce, pour des temps très courts seulement. En particulier, l'existence d'états métastables est un frein considérable à une simulation numérique efficace. Or, le comportement macroscopique de certains systèmes est déterminé sur le long terme par des événements rares, d'origine microscopique. Par exemple, les propriétés mécaniques de matériaux soumis à des dommages radiatifs (tels que les enceintes de cuves des centrales nucléaires) doivent idéalement être prédites sur des périodes de plusieurs dizaines d'années, alors que les évenements microscopiques conduisant à ces modifications de propriétés mécaniques (migration de dislocations ou de lacunes, par exemple) se produisent sur des temps de l'ordre de la picoseconde. Il est donc vraiment nécessaire de considérer des méthodes permettant de simuler de plus gros systèmes sur des temps plus longs. On se concentre dans cette section sur trois types de stratégies :

(i) les techniques de calcul de différences d'énergie libre, qui permettent de forcer la transition d'un état métastable à un autre lorsque l'on sait paramétrer correctement cette transition (Section 2.3.1) ;

(ii) il existe également des techniques permettant d'augmenter le temps de simulation total, par le biais de pas de temps d'intégration plus grands pour une dynamique moyenne, une dynamique accélérée, ou une approche de type Monte-Carlo cinétique (Section 2.3.2) ;

(iii) les dynamiques réduites ou effectives, sont obtenues à partir de la dynamique tous-atomes par une procédure de moyennisation ou de projection, et sont moins gourmandes en temps de calcul (Section 2.3.3).

On ne parle pas ici de techniques permettant d'augmenter la taille (spatiale) des systèmes étudiés, telles que des méthodes de décomposition de domaines, ou de couplage de modèles, où une petite région du système est décrite par un modèle précis, alors que la majeure partie est décrite par un modèle plus grossier. Un exemple de cette dernière approche est la méthode QCM (*quasicontinuum method*) de Tadmor, Ortiz et Phillips [334], où on couple une description atomique et une discrétisation par éléments finis. Cette méthode a été étudiée d'un point-de-vue mathématique pour un système modèle unidimensionnel par Blanc, Le Bris et Legoll dans [32].

2.3.1 Calcul de différences d'énergie libre

Lorsque la variable à l'origine du comportement métastable du système est connue (ou supposée connue), on peut recourir à des techniques de calcul de différences d'énergie libre pour forcer les transitions entre les états métastables du système. Bien sûr, l'efficacité de telles techniques dépend cruciallement du choix de la coordonnée de réaction, qui représente les degrés de liberté essentiels du système. Le choix de cette coordonnée de réaction est un problème très important en pratique, et que nous n'évoquerons pas plus avant ici, nous contentant de supposer une fois pour toutes qu'une coordonnée de réaction convenable est disponible. On se limite ainsi au problème de calcul de différences d'énergie libre entre des états associés à des valeurs différentes de la coordonnée de réaction.

Remarque 2.2 (Motivation mathématique du choix de la coordonnée de réaction). *Peu d'études mathématiques ont traité du choix optimal de la coordonnée de réaction. On peut toutefois citer le travail de* Vanden-Eijnden *et* Tal *[357] qui propose une définition variationnelle de la*

coordonnée de réaction et de la surface de séparation entre deux zones de métastabilité, à la base de la méthode de la corde [370] (voir également la discussion à ce sujet dans [91]).

L'énergie libre absolue d'un système est définie par

$$F = -\frac{1}{\beta} \ln Z,$$

où $Z = \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H(q,p)}\, dq\, dp$ est la fonction de partition. L'énergie libre absolue ne peut être calculée que pour certains systèmes comme le gaz parfait ou les solides à basse température (en utilisant le spectre de phonons) [113, 281]. Heureusement, dans de nombreuses applications, c'est la *différence d'énergie libre* qui est la quantité pertinente. Le profil de la différence d'énergie libre est en effet une information précieuse sur la stabilité relative des espèces en présence, et est utilisé pour préciser la cinétique de transition entre l'état initial et l'état final. Les transitions que l'on considère peuvent être classées en deux catégories :

(i) les transitions de type alchimiques sont indexées par un paramètre extérieur λ, qui peut représenter l'intensité d'un champ magnétique pour un système de spin, la température, un paramètre des potentiels d'interaction,[7] etc. Pour une valeur de λ fixée, le système est ainsi décrit par un Hamiltonien H_λ (ou une énergie potentielle V_λ). La différence d'énergie libre correspondante est alors

$$\Delta F = -\beta^{-1} \ln \left(\frac{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_1(q,p)}\, dq\, dp}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_0(q,p)}\, dq\, dp} \right) ;$$

(ii) dans le cas où la transition est indexée par une coordonnée de réaction de dimension m, notée $\xi : \mathbb{R}^{3N} \to \mathbb{R}^m$ (les différents états de la transition étant dans les sous-variétés correspondant à des lignes de niveau de ξ), la différence d'énergie libre est

$$\Delta F = -\beta^{-1} \ln \left(\frac{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z_1}\, dq\, dp}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z_0}\, dq\, dp} \right).$$

Rappelons que $\delta_{\xi(q)-z}$ est la mesure portée par $\Sigma(z) = \{q,\ \xi(q) = z\}$ et définie par

$$\delta_{\xi(q)-z} = |\nabla \xi|^{-1} d\sigma_{\Sigma(z)},$$

où $d\sigma_{\Sigma(z)}$ est la mesure de Lebesgue induite sur $\Sigma(z)$.

Il est bien plus facile de calculer des différences d'énergie libre que des énergies libres absolues. Les techniques les plus courantes pour ce faire sont de quatre types (voir Figure 2.2 pour une comparaison schématique) :

(i) l'intégration thermodynamique, remontant à Kirkwood [194], consiste à écrire l'évolution comme une succession quasi-statique d'états d'équilibre (ce qui correspond à une transition infiniment lente) ;

(ii) la méthode perturbative de Zwanzig (*free energy perturbation method*) [380] ne peut être employée que pour les transitions alchimiques. Elle consiste à ré-écrire les différences d'énergie libre comme des moyennes canoniques, ce qui permet d'utiliser toutes les techniques

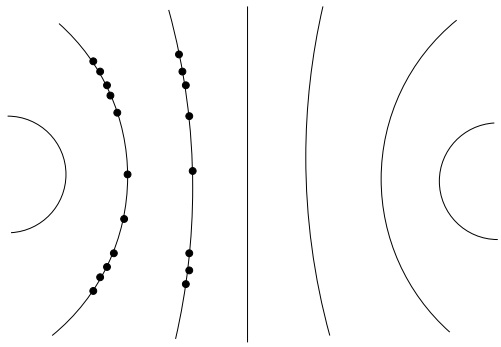
[7] C'est d'ailleurs ce cas de figure qui est à l'origine du terme *alchimique* : en effet, il est facile, sur son ordinateur, de changer du plomb en or, en faisant passer la charge du noyau atomique de $Z = 82$ à $Z = 79$!

d'échantillonnage usuelles, et ouvre la voie à de nombreux raffinements, dont l'utilisation de fonctions d'importance (tel que l'*Umbrella sampling* de Torrie et Valleau [345]) ;

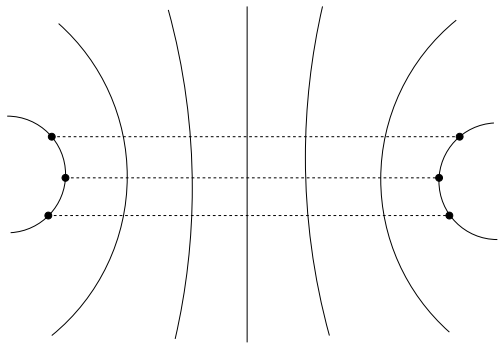
(iii) une technique plus récente utilise une dynamique de transition à vitesse finie, et donc hors d'équilibre. Dans ce cas, il est nécessaire de donner un poids exponentiel aux différentes trajectoires de transition possibles afin de retrouver la bonne différence d'énergie libre. C'est la fameuse égalité que Jarzynski a introduite dans [187] ;

(iv) enfin, des *dynamiques adaptatives* peuvent également être employées. Dans ce cas, la vitesse et le chemin de transition entre les états initiaux et finaux n'est pas fixée *a priori*, mais un terme de biais force la transition en pénalisant les régions qui ont été suffisamment échantillonnées. Ce terme de biais peut être une force biaisante, comme pour l'*Adaptive Biasing Force de* Darve et Pohorille [75], ou un potentiel biaisant, comme pour les méthodes de Wang et Landau [368] ou de Iannuzzi, Laio et Parrinello [179] (*nonequilibrium metadynamics*).

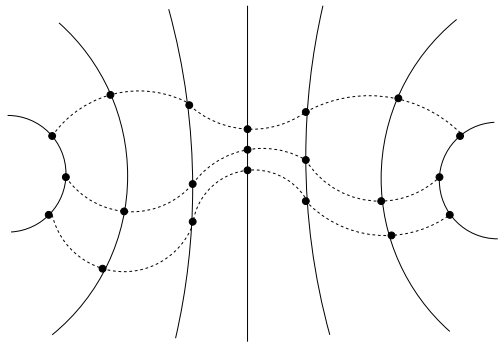
Nous détaillons à présent ces différentes approches dans le cas de transitions alchimiques (par souci de simplicité), et indiquons comment elles se généralisent au cas de transitions indexées par une coordonnée de réaction.



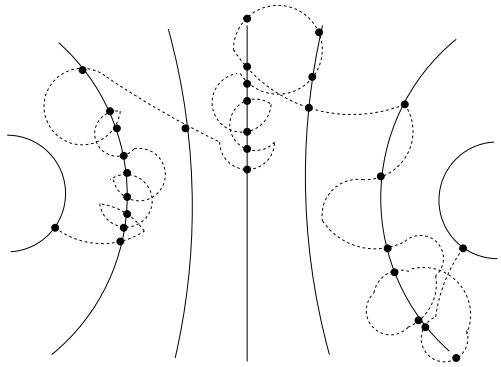
(a) Intégration thermodynamique : on utilise une dynamique projetée et on échantillonne "feuille" par "feuille".



(b) Méthode perturbative : on considère une transition immédiate entre les états initiaux et les états finaux.



(c) Dynamique hors d'équilibre : la vitesse de transition est la même pour chaque trajectoire et est imposée *a priori*.



(d) Dynamique adaptative : on force le système à quitter les zones où l'échantillonnage est suffisant.

Fig. 2.2. Comparaison schématique des différentes méthodes de calcul de différences d'énergie libre.

Intégration thermodynamique

Pour une transition alchimique,

$$F(\lambda) = -\frac{1}{\beta} \ln \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp.$$

L'intégration thermodynamique repose sur le fait que $F(\lambda) - F(0) = \int_0^\lambda F'(s)\, ds$, sachant que

$$F'(\lambda) = \frac{\displaystyle\int_{T^*\mathcal{M}} \frac{\partial H_\lambda}{\partial \lambda}(q,p)\, \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp}$$

est la moyenne canonique de $\frac{\partial H_\lambda}{\partial \lambda}$ pour la mesure canonique $d\mu_\lambda = Z_\lambda^{-1}\, \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp$. Ainsi, en pratique, on calcule $F'(\lambda_i)$ pour un ensemble de valeurs $\lambda_i \in [0,1]$ et on intègre numériquement la dérivée pour obtenir le profil voulu.

L'extension au cas de transitions indexées par une coordonnée de réaction peut être faite par exemple par une dynamique stochastique projetée (voir le travail de Ciccotti, Lelièvre et Vanden-Eijnden [66], rappelé en Section 4.1.2). Dans ce cas, on peut montrer rigoureusement que la dérivée de l'énergie libre peut être obtenue par une moyenne en temps des multiplicateurs de Lagrange associés à la contrainte $\xi(q)$ fixé. Notons également qu'il est possible de recourir à un schéma de type Hybrid Monte-Carlo convenablement modifié (voir Schütte et Hartmann [151]).

Méthode perturbative

La méthode perturbative consiste à ré-écrire ΔF comme

$$\Delta F = -\beta^{-1} \ln \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta(H_1 - H_0)} d\mu_0.$$

Notons immédiatement que cette technique ne peut pas être utilisée en tant que telle pour des transitions indexées par une coordonnée de réaction car dans ce cas, les mesures initiales et finales ont des supports disjoints.[8]

Une approximation de ΔF est obtenue en tirant des configurations (q^n, p^n) distribuées selon $d\mu_0$ et en faisant la moyenne des quantités $\mathrm{e}^{-\beta(H_1 - H_0)(q^n,p^n)}$ correspondantes. Cependant, les distributions initiales et finales $d\mu_0$ et $d\mu_1$ ont souvent un faible recouvrement, et il est préférable, par souci de fiabilité des estimations, de considérer une succession d'états intermédaires. Décomposant la variation d'énergie libre en n pas intermédiaires $\lambda_i = i/n$, il vient

$$\Delta F_i = -\beta^{-1} \ln \frac{Z_{\lambda_{i+1}}}{Z_{\lambda_i}} = -\beta^{-1} \ln \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta(H_{\lambda_{i+1}} - H_{\lambda_i})} d\mu_{\lambda_i},$$

et $\Delta F = \Delta F_0 + \cdots + \Delta F_{n-1}$. On s'attend à ce que les distribution $d\mu_i$ et $d\mu_{i+1}$ aient un meilleur recouvrement (si n est assez grand) et donc que l'estimation de ΔF_i soit plus fiable.

La différence d'énergie libre élémentaire ΔF_i peut être calculée plus efficacement en utilisant une fonction d'importance, technique qui dans ce contexte porte le nom d'*Umbrella sampling* [345]. On a en effet

[8] Mais cette technique est tout de même utilisée en pratique pour le calcul de différences d'énergie libre dans le cas d'une coordonnée de réaction : pour ce faire, on considère la transition associée à $V_\lambda(q) = V(q) + K(\xi(q) - \lambda)^2$. On montre que la différence d'énergie libre associée à cette transition (pour K fixé) converge vers la différence recherchée lorsque $K \to +\infty$.

$$\Delta F = -\beta^{-1} \ln \frac{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta(H_1 - W)} d\pi_W}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta(H_0 - W)} d\pi_W},$$

où $d\pi_W(q,p) = Z^{-1}\mathrm{e}^{-\beta W(q,p)}\, dq\, dp$. La mesure $d\pi_W$ doit être choisie de manière à ce qu'elle ait un recouvrement appréciable avec $d\mu_0$ et $d\mu_1$. C'est d'ailleurs cette propriété de double recouvrement qui a motivé le nom de *Umbrella sampling* (littéralement, "échantillonnage parapluie"). Des choix possibles sont par exemple

$$d\pi_W(q) = Z_{1/2}^{-1}\, \mathrm{e}^{-\beta H_{1/2}(q,p)}\, dq\, dp,$$

ou l'utilisation de $\tilde{H}_{1/2}$ tel que

$$d\pi_W(q) = \tilde{Z}_{1/2}^{-1}\, \mathrm{e}^{-\beta \tilde{H}_{1/2}(q,p)}\, dq\, dp = \frac{1}{2}(d\mu_0 + d\mu_1).$$

L'égalite de Jarzynski

L'égalité de Jarzynski est facile à obtenir si on considère un système gouverné par une dynamique hamiltonienne, partant de conditions initiales canoniquement distribuées. On considère une transition à taux fini, se produisant en un temps $0 < T < +\infty$, partant de $\lambda(0) = 0$ et allant à $\lambda(T) = 1$. Plus précisément, partant de conditions initiales distribuées selon $d\mu_0$, on considère l'équation différentielle ordinaire non-autonome ($0 \le t \le T$)

$$\begin{cases} \dot{q}_i(t) = \dfrac{\partial H_{\lambda(t)}}{\partial p_i}(q(t), p(t)), \\ \dot{p}_i(t) = -\dfrac{\partial H_{\lambda(t)}}{\partial q_i}(q(t), p(t)). \end{cases} \tag{2.39}$$

Notant Φ^λ le flot associé, le travail fourni au système partant d'une condition initiale (q,p) est

$$W(q,p) = \int_0^T \frac{\partial H_{\lambda(t)}}{\partial \lambda}(\Phi_t^\lambda(q,p))\, \lambda'(t)\, dt = H_1(\Phi_T^\lambda(q,p)) - H_0(q,p).$$

En effet, avec $\Phi_t^\lambda(q,p) = (Q(t), P(t))$,

$$\partial_t \left(H_{\lambda(t)}(\Phi_t^\lambda(q,p))\right) = \frac{\partial H_{\lambda(t)}}{\partial \lambda}(\Phi_t^\lambda(q,p))\, \lambda'(t) + \frac{\partial H_{\lambda(t)}}{\partial q} \cdot \partial_t Q(t) + \frac{\partial H_{\lambda(t)}}{\partial p} \cdot \partial_t P(t),$$

et les deux derniers termes du membre de droite se simplifient pour la dynamique (2.39). Ainsi,

$$\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta W(q,p)}\, d\mu_0(q,p) = Z_0^{-1} \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_1(\Phi_T^\lambda(q,p))}\, dq\, dp = Z_0^{-1} \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_1(q,p)}\, dq\, dp$$

car Φ_T^λ définit un changement de variables de Jacobien 1. L'égalité ci-dessus peut être ré-écrite comme

$$\boxed{\mathbb{E}(\mathrm{e}^{-\beta W}) = \frac{Z_1}{Z_0} = \mathrm{e}^{-\beta \Delta F},} \tag{2.40}$$

où l'espérance est prise par rapport à des conditions initiales distribuées selon $d\mu_0$. L'extension à des dynamiques stochastiques est présentée dans le cas alchimique en Section 4.1.1, et suit la preuve de Hummer et Szabo [177] reposant sur une formule de Feynman-Kac.

Extension au cas de transitions indexées par une coordonnée de réaction

En se fondant sur les dynamiques stochastiques projetées, et toujours par une démonstration utilisant une formule de Feynman-Kac, nous avons proposé avec T. Lelièvre et M. Rousset [P6] une extension de la dynamique hors d'équilibre de Jarzynski lorsque la transition est indexée par une coordonnée de réaction, ainsi que l'égalité (2.40) associée (voir également la Section 4.1.2). La dérivation de cette égalité demande en particulier une définition correcte du travail exercé sur le système, et on montre que le travail peut être calculé comme une moyenne trajectorielle des multiplicateurs de Lagrange utilisés pour projeter la dynamique sur les différentes sous-variétés traversées, à un terme de correction près prenant en compte le caractère hors d'équilibre de la procédure (on exerce une force non nulle sur le système pour forcer les transitions, et il faut enlever le travail exercé par cette force pour obtenir la bonne définition des travaux).

Dégénerescence de la distribution des travaux

La formule (2.40) montre que l'on peut obtenir la différence d'énergie libre comme une moyenne non-linéaire sur plusieurs réalisations indépendantes du processus de transition. On peut donc très facilement paralléliser le calcul de (2.40), et l'indépendance des trajectoires calculées (si les conditions initiales sont bien indépendantes et identiquement distribuées selon la mesure canonique) permet d'obtenir des estimations d'erreur par des théorèmes de type TCL. Cependant, aussi élégante que soit l'égalité de Jarzynski, on se heurte en pratique à des problèmes de variance provenant de la dégénerescence de la distribution des poids $\mathrm{e}^{-\beta W}$. Ainsi, quelques rares réalisations de la dynamique de transition dominent complètement la moyenne (2.40).

Ces considérations heuristiques peuvent être précisées dans certaines situations où des calculs analytiques peuvent être menés complètement. Soit la famille de Hamiltoniens

$$H_\lambda(q,p) = \frac{1}{2}\omega^2(q-\lambda)^2 + \frac{1}{2}p^2,$$

et le chemin de transition linéaire $\lambda(t) = t/T$. La solution générale de la dynamique hamiltonienne est

$$q(t) = q(0)\cos(\omega t) + \frac{p(0)}{\omega}\sin(\omega t) + \int_0^t \omega \sin(\omega s)\lambda(t-s)\,ds.$$

Par souci de simplicité, on prend un temps de transition T tel que $\omega T = \pi/2 \bmod \pi$ (mais l'analyse suivante reste valable dès que $\omega T \neq 0 \bmod \pi$). Alors,

$$q(T) = q(0) + \lambda(T) - \frac{1}{\omega T}, \quad p(T) = -\omega q(0) + \frac{1}{T}.$$

Ainsi, pour des conditions initiales canoniquement distribuées (*i.e.* $q(0) \sim \omega^{-1}\beta^{-1/2}\mathcal{N}(0,1)$), les positions et impulsions finales sont distribuées selon

$$q(T) \sim 1 - \frac{1}{\omega T} + \omega^{-1}\beta^{-1/2}\mathcal{N}(0,1), \quad p(T) \sim \frac{1}{T} + \beta^{-1/2}\mathcal{N}(0,1).$$

Ces formules montrent que la distribution des positions n'est pas la distribution d'équilibre au temps final, et même plus précisément, qu'en moyenne les positions sont en retard (position moyenne $1-(\omega T)^{-1}$ au lieu de 1), ce retard augmentant avec la rapidité de la transition. On peut également calculer les travaux exercés sur chaque système lors de la transition :

$$W(q(0),p(0)) = H_1(q(T),p(T)) - H_0(q(0),p(0)) = \frac{1}{T^2} + \frac{2\omega}{T}q(0) \sim \frac{1}{T^2} + \frac{2}{T\sqrt{\beta}}\mathcal{N}(0,1). \quad (2.41)$$

Ceci montre que $\mathbb{E}(W) = T^{-2} > \Delta F = 0$ (l'espérance étant prise par rapport à des conditions initiales $(q(0),p(0))$ canoniquement distribuées), alors que $\mathbb{E}(\mathrm{e}^{-\beta W}) = 1 = \mathrm{e}^{-\beta\Delta F}$ comme prévu.

Lorsque le temps de transition est court, l'expression (2.41) montre clairement que la queue inférieure de la distribution des travaux est particulièrement importante pour obtenir des estimations correctes de la différence d'énergie libre, et qu'une faible fraction de la distribution des travaux domine l'estimation proposée. Plus précisément, notant $P(W)$ la distribution des travaux,

$$\mathbb{E}(\mathrm{e}^{-\beta W}) = \int_{\mathbb{R}} \mathrm{e}^{-\beta W} P(W)\, dW = C \int_{\mathbb{R}} \exp\left[-\frac{\beta T^2}{8}\left(W + \frac{4}{T^2}\right)^2\right] dW.$$

Lorsque T est petit, les valeurs des travaux contribuant le plus à l'intégrale ci-dessus sont distribués autour de la valeur $-4/T^2$, avec une déviation standard $\mathrm{O}(T)$. Ces valeurs sont fort improbables au vu de (2.41). Notons enfin que la queue de la distribution des travaux est liée à la queue inférieure de la distribution des positions initiales. Ainsi, il faut que les conditions initiales soient échantillonnées avec une très grande précision si on veut réaliser une transition très rapide (ce qui demande donc de très nombreuses conditions initiales). En pratique, il vaut donc mieux se placer dans des régimes de transition moins brusques ($\mathrm{O}(T) \sim 1$ au moins).

Pour éviter la dégénerescence des poids $\mathrm{e}^{-\beta W}$ et les fâcheuses conséquences qui s'en suivent, nous avons proposé, avec M. Rousset dans [P10], d'utiliser un méchanisme de sélection sur différentes répliques du système simulées en parallèle (voir également la Section 4.3.3 pour plus de précisions). Pour ce faire, on utilise un système de particules en interaction, stratégie inspirée par les techniques de rééchantillonnage (voir la littérature sur les algorithmes de Monte-Carlo séquentiels, en particulier l'ouvrage de Doucet, Freitas et Gordon [84] et l'article de revue de Doucet, Del Moral et Jasra [85]). Dans ce cas, il n'est plus nécessaire d'attacher un poids à chaque particule, l'équilibre étant maintenu à tout instant par le biais de règles de sélection probabilistes (mort/naissance) : les systèmes ayant un travail inférieur à la moyenne sont favorisés, les autres sont pénalisés. On peut montrer la consistance de cette approche dans la limite d'une infinité de répliques simulées en parallèle en utilisant les travaux de Rousset [289, 290].

Une autre approche pour calculer de manière plus fiable l'espérance (2.40) est de considérer cette espérance comme une espérance sur tous les chemins de transition possibles. On peut alors utiliser des techniques d'échantillonnage de chemins combinées à l'utilisation d'une fonction d'importance [331, 374] pour biaiser l'échantillonnage en faveur des chemins correspondant aux valeurs improbablement basses des travaux W (voir la Section 4.3 pour plus de précisions sur l'échantillonnage des chemins et son application au calcul de différences d'énergie libre).

Dynamiques adaptatives

L'objectif des dynamiques adaptatives est d'échantillonner les mesures $d\mu_\lambda$ juste le temps qu'il faut, tout en passant les barrières d'énergie libre du système. Pour décrire précisément les dynamiques adaptatives en terme d'un procédé de point-fixe, il est utile d'utiliser la formulation que nous avons proposée avec T. Lelièvre et M. Rousset dans [P4] (voir également la Section 4.4). Nous présentons cette dynamique dans le cas alchimique, mais les formulations originelles de cette méthode ont toutes été faites dans le cas d'une coordonnée de réaction.

Considérons donc la variable étendue $X = (q, \lambda)$, la coordonnée de réaction associée étant $\xi(X) = \lambda \in \mathbb{T}$. On suppose que la transition est paramétrée par le biais d'un potentiel $V(q, \lambda)$, la mesure canonique associée (en positions) étant $d\pi_\lambda(q) = Z_\lambda \mathrm{e}^{-\beta V(q,\lambda)}\, dq$. Lorsque X_t évolue selon une dynamique de Langevin suramortie (*overdamped Langevin*), à savoir

$$\begin{cases} dq_t = -\nabla_q V(q_t, \lambda_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t^q, \\ d\lambda_t = -\partial_\lambda V(q_t, \lambda_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t^\lambda, \end{cases} \tag{2.42}$$

où W_t^q, W_t^λ sont des mouvements browniens standard indépendants, la mesure $d\Pi(q,\lambda) = Z^{-1}\mathrm{e}^{-\beta V(q,\lambda)}\,dq\,d\lambda$ est invariante.[9] En principe, on peut donc utiliser la dynamique (2.42) pour échantillonner des configurations distribuées selon $d\Pi(X)$, et calculer les différences d'énergie libre selon

$$F(\lambda_2) - F(\lambda_1) = -\beta^{-1}\ln\frac{\overline{\psi}_{\mathrm{eq}}(\lambda_2)}{\overline{\psi}_{\mathrm{eq}}(\lambda_1)},$$

où les marginales $\overline{\psi}_{\mathrm{eq}}$ de la distribution d'équilibre sont

$$\overline{\psi}_{\mathrm{eq}}(\lambda) = \int_{\mathcal{M}} \mathrm{e}^{-\beta V(q,\lambda)}\,dq.$$

Cependant, la dynamique (2.42) ne peut être utilisée en tant que telle si le système présente de la métastabilité le long du profil $F(\lambda) - F(0)$, car dans ce cas le paramètre λ risque de rester coincé un long moment dans certaines régions de l'intervalle $[0,1]$. Les barrières d'énergie libre associées correspondent à des petites valeurs de $\overline{\psi}_{\mathrm{eq}}(\lambda)$ (comparées à $\overline{\psi}_{\mathrm{eq}}(0)$, par exemple).

L'idée des dynamiques adaptatives est d'ajouter un terme de biais dans la dynamique du paramètre extérieur (variable λ_t) qui force l'exploration de tout l'intervalle $[0,1]$. On peut même demander mieux : il est possible en effet de construire ce terme de biais de manière à ce que les variables λ soient asymptotiquement uniformément réparties sur $[0,1]$, et que le terme de biais donne la différence d'énergie libre. Pour préciser ces considérations heuristiques, il est utile de recourir à un ensemble de réalisations d'une dynamique stochastique sur X_t de la forme

$$\begin{cases} dq_t = -\nabla_q V(q_t,\lambda_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t^q, \\ d\lambda_t = -\partial_\lambda\left[V(q_t,\lambda_t) - F_{\mathrm{bias}}(t,\lambda_t)\right]\,dt + \sqrt{2\beta^{-1}}\,dW_t^\lambda, \end{cases} \tag{2.43}$$

où on a introduit un terme de biais $F_{\mathrm{biais}}(t,\lambda)$. Les configurations X_t sont distribuées au temps t selon une densité $\psi_t(q,\lambda)\,dq\,d\lambda$ (en pratique, cela revient à simuler un nombre infini de répliques en parallèle, et à prendre la limite de la mesure empirique associée). La distribution des variables λ_t est donnée par les marginales

$$\overline{\psi}_t(\lambda) = \int_{\mathcal{M}} \psi_t(q,\lambda)\,dq.$$

Si le terme de biais $F_{\mathrm{biais}}(t,\lambda)$ converge en effet vers $F(\lambda)$, alors la variable X_t évoluant selon (2.43) est distribuée selon $d\Pi_\infty(q,\lambda) = Z_\infty^{-1}\mathrm{e}^{-\beta(V(q,\lambda)-F(\lambda))}$, et donc λ est distribuée selon les marginales

$$\overline{\psi}_\infty(\lambda) = \int_{\mathcal{M}} \exp(-\beta[V(q,\lambda) - F(\lambda)])\,dq = 1.$$

Ceci signifie qu'on a éliminé le caractère métastable du profil d'énergie libre, et que toutes les régions sont explorées uniformément.

Potentiel biaisant adaptatif

En pratique, la clé du succès des dynamiques adaptatives est de proposer une mise à jour adéquate du terme biaisant, pris sous forme potentielle pour commencer. Une première idée est de forcer la convergence de la marginale $\overline{\psi}_t(\lambda)$ vers sa valeur limite $\overline{\psi}_\infty(\lambda) = 1$, et d'utiliser les propriétés de la dynamique sur q_t dans (2.43) pour obtenir la bonne distribution des configurations pour une valeur fixée de λ. Si on suppose que les configurations du système sont distribuées instantanément selon $\psi_t(q,\lambda) = Z_t^{-1}\mathrm{e}^{-\beta(V(q,\lambda)-F_{\mathrm{biais}}(t,\lambda))}$ (ce qui est en effet le cas si la dynamique de la variable q est plus rapide que la dynamique de la variable λ) la mise à jour

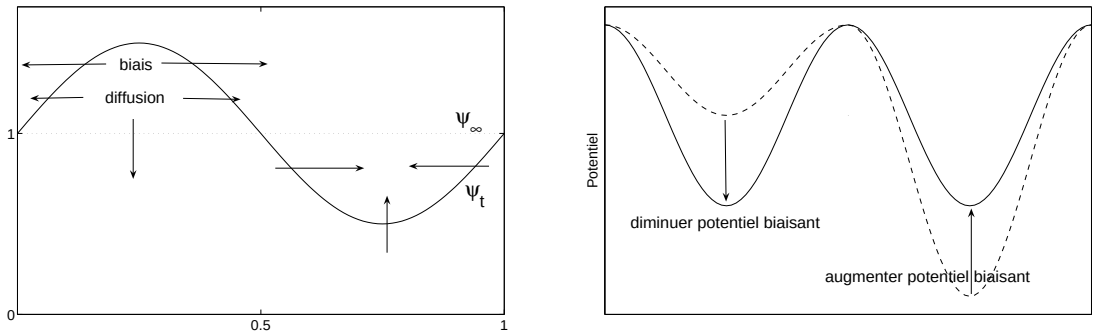
[9] Il faudrait bien sûr préciser les conditions de bord sur la variable λ. Dans le cas de certaines coordonnées de réaction, on peut utiliser des conditions périodiques. Ici, on préférerait probablement des conditions de réflection au bord. Une discussion plus approfondie des conditions de bord est menée à la Section 4.4.

$$\partial_t F_{\text{biais}}(t,\lambda) = -\frac{\beta^{-1}}{\tau} \ln \overline{\psi}_t(\lambda) = \frac{1}{\tau}(F(\lambda) - F_{\text{biais}}(t,\lambda)) + c_t$$

avec $\tau > 0$ est telle que $F_{\text{biais}}(t,\lambda) \to F(\lambda)$ lorsque $t \to +\infty$ (à une constante additive près). En général, $\psi_t(q,\lambda) \neq Z_t^{-1} \mathrm{e}^{-\beta(V(q,\lambda) - F_{\text{biais}}(t,\lambda))}$, mais le potentiel biaisant est toujours mis à jour selon

$$\partial_t F_{\text{biais}}(t,\lambda) = -\frac{\beta^{-1}}{\tau} \ln \overline{\psi}_t(\lambda). \tag{2.44}$$

Dans ce cas, on peut montrer que, s'il existe un point fixe de la dynamique (2.43) avec la mise à jour (2.44), alors on a $F_{\text{biais}}(t,\lambda) \to F(\lambda)$ (à une constante près). La mise à jour (2.44) est assez naturelle si on demande que $\overline{\psi}_t(\lambda)$ soit constant : lorsque $\overline{\psi}_t(\lambda) > 1$ (région sur-explorée), alors on décroît le biais qui pousse les répliques vers cette région ; alors qu'on augmente le terme de biais dans les régions sous-explorées, qui correspondent aux valeurs de λ telles que $\overline{\psi}_t(\lambda) < 1$ (voir la Figure 2.3).



(a) Distributions marginales

(b) Energies libres.

Fig. 2.3. (a) Distributions marginales de la variable λ, au temps t (ligne pleine) et limite en temps long (pointillés). (b) Energie libre cible (ligne pleine) et biais proposé au temps t (ligne interrompue). Dans ce cas, on diminue le biais là où il y a trop d'échantillonnage en λ ($\overline{\psi}_t(\lambda) > \overline{\psi}_\infty(\lambda)$), et on l'augmente là où il n'y a pas assez d'échantillonnage en λ.

Force adaptative biaisante et preuve de convergence

De la même manière, le terme biaisant peut être introduit *via* une force (au lieu d'un potentiel), auquel cas on propose directement une mise à jour de $\partial_\lambda F_{\text{biais}}(t,\lambda)$. Si on suppose à nouveau que les configurations du système sont distribuées au temps t selon $\psi_t(q,\lambda) = Z_t^{-1} \mathrm{e}^{-\beta(V(q,\lambda) - F_{\text{biais}}(t,\lambda))}$, la mise à jour

$$\partial_t \partial_\lambda F_{\text{biais}}(t,\lambda) = -\frac{\beta^{-1}}{\tau} \left(\frac{\displaystyle\int_{\mathcal{M}} \partial_\lambda V(q,\lambda) \psi_t(q,\lambda)\, dq}{\displaystyle\int_{\mathcal{M}} \psi_t(q,\lambda)\, dq} - \partial_\lambda F_{\text{biais}}(t,\lambda) \right) = \frac{1}{\tau}(\partial_\lambda F(\lambda) - \partial_\lambda F_{\text{biais}}(t,\lambda))$$

pour $\tau > 0$ est telle que la force biaisante $\partial_\lambda F_{\text{bias}}(t,\lambda)$ converge vers $\partial_\lambda F(\lambda)$. Dans le cas général, la force biaisante est encore mise à jour selon

$$\partial_t \partial_\lambda F_{\text{biais}}(t,\lambda) = -\frac{\beta^{-1}}{\tau}\left(\frac{\displaystyle\int_{\mathcal{M}} \partial_\lambda V(q,\lambda)\psi_t(q,\lambda)\,dq}{\displaystyle\int_{\mathcal{M}} \psi_t(q,\lambda)\,dq} - \partial_\lambda F_{\text{biais}}(t,\lambda)\right). \tag{2.45}$$

On peut montrer, comme dans le cas d'un potentiel biaisant adaptatif, que si la dynamique admet un état stationnaire, alors celui-ci est bien tel que $F_{\text{biais}}(t,\lambda) \to F(\lambda)$ (à une constante près) en temps long.

Avec T. Lelièvre, F. Otto et M. Rousset [A1] (voir également la Section 4.4.2 pour une preuve dans un cas simplifié et une brève introduction aux techniques mathématiques nécessaires à la preuve), nous avons pu faire une preuve de convergence dans le cas limite $\tau \to 0$ pour la dynamique (2.45). La démonstration repose sur l'introduction d'une fonction d'entropie de la mesure ψ_t, et sa décomposition en une contribution macroscopique (liée aux marginales $\overline{\psi}_t$) et une contribution microscopique (qui ne dépend que des mesures conditionnées $\psi_t/\overline{\psi}_t$). Remarquant que

$$\partial_t \overline{\psi}_t = \partial_{\lambda\lambda}\,\overline{\psi}_t,$$

on montre facilement la convergence des marginales $\overline{\psi}_t$ et la décroissance de l'entropie macroscopique. La décroissance de l'entropie microscopique est assurée lorsque les mesures conditionnées $\psi_\infty(\cdot,\lambda)/\overline{\psi}_\infty(\lambda)$ satisfont une inégalité de Sobolev logarithmique avec une constante uniforme en λ. Au final, la vitesse de convergence est le minimum entre la vitesse de convergence macroscopique (exploration diffusive) et la vitesse de convegence microscopique (constante de l'inégalité de Sobolev logarithmique). L'extension au cas d'une coordonnée de réaction générale suit les mêmes lignes, mais demande toutefois de modifier un peu la dynamique (2.45).

Accélération de la convergence

Le formalisme ci-dessus, fondé sur des ensembles de réalisations d'une dynamique idoine, suggère naturellement une implémentation parallèle de la dynamique par l'utilisation de répliques concourrant à construire ensemble un même potentiel biaisant. Cette implémentation naturelle peut toutefois être améliorée par le biais d'un processus de sélection sur les répliques (voir [P4]). En effet, dans l'analyse heuristique des dynamiques adaptatives, on a insisté sur l'importance d'une répartition uniforme des coordonnées de réaction λ. Il semble donc intéressant, lorsque l'on simule plusieurs répliques du système en parallèle, d'ajouter un processus de mort/naissance au processus de diffusion (2.43), toujours dans l'esprit des méthodes particulaires déjà employées pour éviter la dégénerescence des poids exponentiels dans l'égalité de Jarzynski. L'idée est ici de dupliquer les répliques dans les régions sous-explorées (répliques innovantes), et d'éliminer les répliques dans les régions sur-explorées (répliques redondantes). Contrairement au processus de diffusion, on permet ainsi des mouvements "non-locaux" dans l'espace des phases, complémentaires du processus de diffusion, et qui permettent d'accélérer l'étalement de la distribution des valeurs de λ. On trouvera dans [P4] une analyse sur un cas simple de l'effet de ce terme de sélection (voir aussi la Section 4.4.1).

2.3.2 Différentes approches pour augmenter le temps effectif de simulation

On présente dans cette section quelques stratégies pour simuler des temps plus longs. Notons qu'il est généralement plus difficile d'augmenter le temps de simulation que la taille des systèmes étudiés, car des stratégies parallèles sont souvent limitées par la nature séquentielle du temps.

La stratégie pararéelle

Une exception notoire à la limitation intrinsèque évoquée ci-dessus est la stratégie pararéelle, proposée par Lions, Maday et Turinici dans [213], puis appliquée au domaine de la dynamique

moléculaire dans [18]. Cette stratégie consiste en une première itération, séquentielle mais peu coûteuse : la proposition d'une trajectoire grossière par le biais d'un intégrateur approché (grand pas de temps ou potentiel d'interaction simplifié) ; dans un deuxième temps, on raffine en parallèle les différents segments de la trajectoire. On répète ensuite cette procédure jusqu'à convergence.

Pas de temps d'intégration plus grands

Il arrive couramment en dynamique moléculaire que l'énergie d'un système soit la somme d'une contribution évoluant rapidement, et d'une contribution évoluant sur des temps bien plus longs :

$$V(q) = V_{\text{lent}}(q) + V_{\text{rapide}}(q).$$

Le terme rapide peut provenir de composantes raides de l'énergie potentielle (ou de degrés de liberté associés à de petites masses), et est souvent moins coûteux à évaluer que le terme lent. En effet, les termes rapides sont souvent associés à des interactions locales (à courte portée), dont le coût de calcul croît linéairement avec la taille du système ; alors que les termes lents correspondent typiquement à des interactions à longue portée, ce qui fait que le coût de calcul croît quadratiquement avec la taille du système.

Dans cette situation, le pas de temps d'intégration est déterminé par la partie rapide du potentiel. Il y a plusieurs façons de remédier à ce problème :

(i) lorsque le terme rapide provient de composantes raides dans le potentiel, et que ces composantes raides servent à pénaliser une contrainte (longueur de liaison presque fixée par exemple), il peut se révéler intéressant de recourir à une dynamique contrainte, telle que RATTLE [8] ou SHAKE [295] ;

(ii) on peut utiliser des intégrateurs à pas de temps multiples. Les forces rapides sont alors évaluées avec un pas de temps Δt proche du pas de temps utilisé pour un algorithme de Verlet standard, alors que l'intégration temporelle des forces lentes se fait avec un pas de temps Δt_{lent} bien plus grand. La méthode *Impulse* [141, 347] est un tel algorithme, qui correspond à la décomposition de Strang du Hamiltonien originel en deux composantes, $H = H_{\text{lent}} + H_{\text{rapide}}$, avec

$$H_{\text{lent}}(q,p) = V_{\text{lent}}(q), \qquad H_{\text{rapide}}(p,q) = V_{\text{rapide}}(q) + \frac{1}{2}p^T M^{-1} p.$$

Cependant, les résonances numériques qui apparaissent demandent que le temps Δt_{lent} entre les évalutions de la force lente soit plus petit que la moitié de la période du mouvement [31, 119]. Ainsi, Δt_{lent} est encore limité par les modes de haute fréquence. On pourra se reporter à l'ouvrage de Hairer, Lubich et Wanner [146, Chap. XIII] pour un traitement mathématique complet dans le cas où la composante rapide est harmonique.

Approches de type Monte-Carlo cinétique

Dans les approches de type Monte-Carlo cinétique (*Kinetic Monte-Carlo*, KMC), on considère une liste d'états métastables du système, ainsi que tous les événements pouvant intervenir (transitions entre états métastables). Le système en question peut être une modélisation tous-atomes, éventuellement *ab-initio*, ou une version réduite du système physique (par exemple, pour des événements se produisant dans un cristal parfait à basse température, on peut ne considérer que les défauts dudit cristal, tels que les atomes en position interstitielle, les lacunes, ou les aggrégats). Cette dernière approche est une approche Monte-Carlo cinétique sur objets (*Object KMC*).

Pour simplifier, on présente seulement l'algorithme KMC d'équilibre, pour lequel la liste des événements possibles et leurs probabilités d'occurrence sont fixées (des techniques permettant de mettre à jour la liste des événements et leurs probabilités au cours de la simulation ont également

été developpées, et sont fondées sur les travaux de Henkelman et Jonsson [158]). On suppose que les événements se produisent à des temps aléatoires exponentiellement distribués. Indexant par i les événements possibles, et notant r_i les taux de réactions associés (les temps d'occurrence associés étant alors des variables aléatoires exponentiellement distribuées, selon la densité $r_i \mathrm{e}^{-r_i t}$), l'algorithme KMC, proposé par Bortz, Kalos et Lebowitz [37] dans le contexte de la science des matériaux (et proposé indépendamment un peu plus tard par Gillespie [129,130] pour traiter les réactions chimiques) est

Algorithme de Monte-Carlo cinétique

Algorithme 2.1. Pour une liste de M événements possibles $i = 1, \ldots, M$, de taux de réaction r_i, et partant d'une configuration initiale du système à $t^0 = 0$,

(1) tirer un événement k, selon la loi de probabilité discrète $(w_i)_{i=1,\ldots,M}$ avec

$$w_i = \frac{r_i}{\sum_{j=1}^{M} r_j} \ ;$$

(2) effectuer le mouvement correspondant à l'événement k;

(3) incrémenter le temps d'un temps aléatoire exponentiellement distribué selon une loi exponentielle de paramètre $\sum_{j=1}^{M} r_j$:

$$t^{n+1} = t^n + \tau^n, \qquad \tau^n \sim \mathcal{E}\left(\sum_{j=1}^{M} r_j\right) \ ;$$

(4) retourner en (1).

Cet algorithme n'est pas efficace sous cette forme lorsque les taux de réactions des différents événements possibles sont très différents les uns des autres : en effet, dans ce cas, les événements les moins rares sont effectués très souvent (et avec une très forte probabilité), et les incréments de temps sont petits (de l'ordre du plus petit temps caractéristique des événements). Dans cette situation, Gillespie et Petzold [131, 132] ont montré comment procéder à une réduction de la dynamique rapide, en remplaçant les réactions élémentaires par la réaction de plusieurs entités (méthode τ-leap), une équation de Langevin chimique, ou même une équation déterministe.

Une autre manière de gagner en temps de simulation est de remarquer que des événements quasi simultanés mais spatiallement éloignés peuvent être considérés comme indépendants. Les techniques de décomposition de domaine pour les algorithmes KMC [309] se fondent sur cette remarque ; le principal défi que rencontrent ces méthodes est la nécessaire synchronisation du temps dans les sous-domaines, et le traitement des événements se produisant aux bords des sous-domaines. Des approches adaptatives en espace peuvent également être considérées [63].

Calcul des taux de réaction

L'étape qui requiert le plus de temps de calcul dans une simulation KMC est le calcul des taux de réaction (voire l'identification, parfois) des événements pouvant arriver. Ces événements sont des transitions d'un état métastable à un autre, et ainsi, si la température n'est pas trop haute (et si les effets entropiques ne sont pas trop importants), ces états métastables sont des minima locaux de la surface d'énergie potentielle. Les états de transitions sont alors des points-selles de la surface d'énergie potentielle, situés sur le chemin d'énergie minimale reliant l'état initial et l'état final. La fait que le point-selle soit situé sur le chemin d'énergie minimale est justifié à basse température par la théorie des grandes déviations de Freidlin et Wentzell [112].

Nous décrivons ici des techniques employées dans de nombreux calculs pratiques, reposant sur la méthode du flux réactif (*Reactive Flux method*) de Bennett et Chandler [26,60], et la théorie de l'état de transition (*Transition State Theory*, TST) introduite dès les années 30 par Eyring et Wigner [102,371]. La première tâche à effectuer pour toutes ces méthodes est de localiser les états de transition, sous l'hypothèse que ce sont des point-selles d'ordre 1 (la matrice hessienne a une seule valeur propre négative).[10] On paramètre ensuite la transition de l'état métastable initial à l'état métastable final par une coordonnée de réaction $\xi(q)$ (aussi appelée *variable collective* ou *paramètre d'ordre* dans ce contexte), telle que l'état initial corresponde à $\xi^{-1}(0)$, l'état final à $\xi^{-1}(1)$, et que l'état de transition soit sur $\Sigma = \xi^{-1}(\frac{1}{2})$. On note $n(q)$ la normale à la surface Σ en $q \in \Sigma$, $A = \xi^{-1}[0, \frac{1}{2})$ la région des réactifs, et $\xi^{-1}(\frac{1}{2}, 1]$ la région des produits.

Le flux réactif sortant (qui mesure ce qui s'échappe de A pour aller en A^c) est défini, pour un temps t, comme le flux sortant passant par l'interface séparant les deux régions (voir Figure 2.4, Gauche):

$$\begin{aligned} k_+(t) &= \frac{\left\langle \delta_{\xi(q(0))=1/2}\, \partial_t(\xi(q))|_{t=0}\, \mathbf{1}_{A^c}(q(t)) \right\rangle}{\langle \mathbf{1}_{A^c}(q) \rangle} \\ &= \frac{\displaystyle\int_\Sigma n(q(0)) \cdot \frac{p(0)}{m}\, \mathbf{1}_{A^c}(q(t))\, \mathrm{e}^{-\beta H(q,p)}\, d\sigma_\Sigma(q,p)}{\displaystyle\int_{T^*\mathcal{M}} \mathbf{1}_{A^c}(q)\, \mathrm{e}^{-\beta H(q,p)}\, dq\, dp}. \end{aligned}$$

Le flux réactif entrant $k_-(t)$ est défini de manière similaire. En pratique, $k_+(t)$ atteint une valeur plateau pour des temps $t \ll (k_+(0) + k_-(0))^{-1}$, et cette limite est le taux de flux réactif. Le choix de la surface de séparation Σ est très important en pratique, car un mauvais choix de surface de séparation peut conduire à de nombreuses traversées successives de l'interface (sortie du domaine suivie d'une nouvelle entrée), voire à un faible taux de transition (voire la Remarque 2.2 pour une optimisation du choix de l'interface).

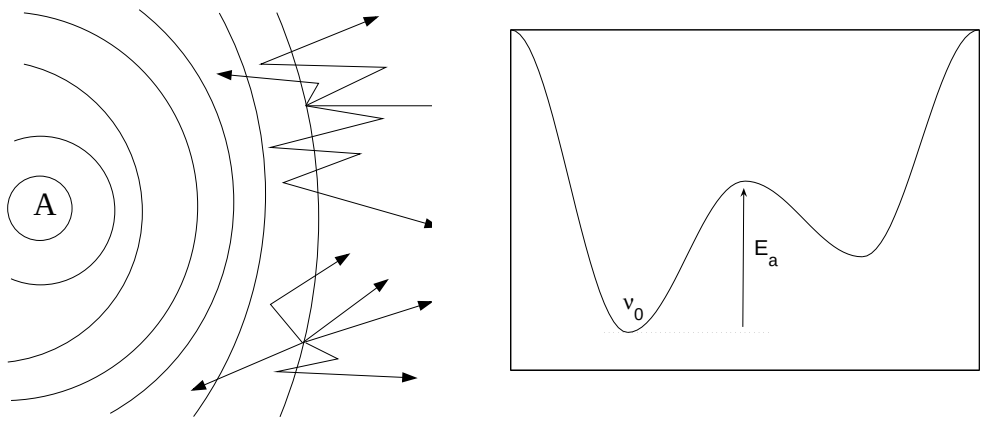


Fig. 2.4. Gauche : représentation schématique du flux sortant de la région A et passant par l'interface. Des tentatives de sortie sont fructueuses (trajectoires finissant hors de A), d'autres ne le sont pas (trajectoires entrant à nouveau dans le domaine et finissant à gauche de la surface de séparation). Droite : approximation de la TST harmonique.

Le taux de réaction classiquement utilisé par la TST est en fait $k(0)$, c'est-à-dire la valeur prédite par la méthode du flux réactif lorsque l'on néglige les retours dans le domaine après traversée de l'interface. Ce taux est donc une borne supérieure du vrai taux de transition, et la

[10] Par exemple, en utilisant une méthode qui suit les vecteurs propres correspondants aux plus petites valeurs propres de la matrice hessienne, partant d'un minima local où cette matrice est définie positive.

TST est d'autant plus une mauvaise approximation que la transition se fait de manière diffusive (par opposition à une image ballistique de la transition). L'approximation de TST harmonique consiste à supposer de plus que le taux de transition peut être écrit comme

$$k^{\mathrm{HTST}} = \nu_0 \mathrm{e}^{-\beta E_a},$$

où l'énergie d'activation E_a est la différence entre l'énergie du point-selle et celle du minima local d'où on part, et ν_0 est la homogène à une fréquence (voir Figure 2.4, Droite). La TST harmonique peut être justifiée de manière rigoureuse dans le cas d'un potentiel harmonique lorsque $\beta E_a \gg 1$ (voir par exemple [178, Section III.A]). Dans le cas général, le taux de réaction de la TST harmonique est donné par l'expression de Vineyard [362]

$$k^{\mathrm{HTST}} = \frac{\prod_{i=1}^{3N} \nu_i^{\min}}{\prod_{i=1}^{3N-1} \nu_i^{\mathrm{sad}}} \mathrm{e}^{-\beta E_a},$$

où $(\nu_i^{\min})_{i=1,\ldots,3N}$ sont les fréquences de la matrice hessienne au minima local de départ, et $(\nu_i^{\min})_{i=1,\ldots,3N-1}$ les fréquences positives de la matrices hessienne au point-selle.

Echantillonnage de chemins de réaction

Echantillonnage des chemins de transition

L'échantillonnage des chemins de transition (*Transition Path sampling*, TPS) est une technique développée par Bolhuis, Chandler, Dellago et Geissler [80,81] et qui permet d'échantillonner des chemins de réaction de longueur fixée (c'est-à-dire qu'on ne considère que les réactions ayant lieu pendant un temps T). Un chemin réactif est défini dans ce contexte comme une trajectoire (déterministe ou stochastique) partant d'un sous-ensemble A de l'espace des phases, et arrivant au temps T dans un sous-ensemble B. L'échantillonnage de tels chemins n'est pas possible par des simulations de dynamique moléculaire directe lorsque de grandes barrières d'énergie libre séparent les deux sous-ensembles. C'est encore plus difficile dans le cas de surfaces d'énergie dites rugueuse, où de nombreux minima locaux séparent les deux régions, et aucune coordonnée de réaction satisfaisante n'est disponible (bien qu'il existe éventuellement des coordonnées de réaction incomplètes ou partielles).

TPS est une méthode permettant d'échantillonner des chemins de transition entre A et B, partant d'un chemin de transition initial. Concrètement, l'algorithme correspondant est un algorithme de Metropolis-Hastings, avec une fonction de proposition convenable, qui permet, partant d'un chemin réactif à l'itération n, de proposer un chemin réactif modifié à l'itération $n+1$. Lorsque la dynamique sous-jacente est déterministe, une fonction de proposition efficace consiste à choisir aléatoirement un temps le long d'un chemin réactif, et à modifier un peu l'impulsion des particules à ce temps-là, puis intégrer la dynamique en temps positif et négatif (voir la revue de Dellago, Bolhuis et Geissler [81] pour plus de précisions). Lorsque la dynamique sous-jacente est stochastique, par exemple dans le cas d'une dynamique de Langevin, ce même algorithme est souvent utilisé, auquel cas la proposition d'un nouveau chemin de transition n'utilise que l'information à temps donné le long du chemin pour proposer un nouveau chemin : en particulier, on n'utilise pas du tout la réalisation du bruit brownien qui a conduit à une transition fructueuse (ce qui peut poser problème pour des phénomènes diffusifs). Inversement, Crooks et Chandler [74] ont proposé de conserver entièrement la réalisation du bruit brownien ayant conduit à une transition, sauf sur un petit intervalle de temps. Dans ce cas, deux chemins successifs sont très similaires, et on peut penser que la grande corrélation entre deux itérations ralentit la convergence numérique de cet algorithme. J'ai proposé dans [P1] une approche généralisant ces deux techniques : dans ce cas, on représente un chemin par ses conditions initiales et la réalisation du mouvement brownien correspondante, et on propose un nouveau chemin en sélectionnant un temps au hasard le long de

la trajectoire, et en intégrant en temps positif et négatif en partant de la configuration courante avec une réalisation du mouvement brownien corrélée à la réalisation utilisée à l'itération n (le taux de corrélation étant un paramètre à optimiser). Des tests numériques montrent que cette approche est en effet intéressante (voir [P1] et la Section 4.3)

Calcul des constantes de réaction

On peut calculer des constantes de réactions grâce aux chemins échantillonnés selon les procédures brièvement évoquées ci-dessus. De telles constantes sont utiles pour des approches de type Monte-Carlo cinétique. Partant au temps $t = 0$ avec une distribution de configurations du système en A, la fonction de corrélation peut être approximée comme

$$C(t) = \frac{\langle \mathbf{1}_A(q_0)\mathbf{1}_B(q_t)\rangle}{\langle \mathbf{1}_A(q_0)\rangle} \simeq k_{AB}\, t,$$

pour des temps $\tau_{\mathrm{mol}} \ll t \ll \tau_{\mathrm{rxn}}$, $\langle\cdot\rangle$ étant une moyenne sur tous les chemins possibles (voir [81] et la Section 4.3 pour plus de précisions sur la mesure utilisée dans l'espace des chemins). Les temps τ_{mol} et τ_{rxn} sont respectivement les temps moléculaires de décorrélation et le temps de réaction typique :

$$\tau_{\mathrm{rxn}} = \frac{1}{k_{AB} + k_{BA}}.$$

La procédure exacte pour extraire la constante de réaction d'un seul échantillon de chemins est expliquée dans [81, Section 4.4].

Une des difficultés de cette approche est qu'il faut partir d'un chemin initialement réactif. Des stratégies pour ce faire sont proposées dans [81]. Il est également possible de forcer progressivement un chemin à quitter la région A pour aller vers la région finale B, par exemple par une dynamique de transition hors d'équilibre "à la Jarzynski" comme proposé par Geissler et Dellago dans [122]. On peut également dans ce cas effectuer cette transition avec plusieurs chemins simulés en parallèle, et utiliser un processus de sélection entre les chemins pour assurer que l'échantillon de chemins réactifs final est non-dégénéré (voir [P1] et la Section 4.3.3).

Formulation de l'échantillonnage par une équation aux dérivées partielles stochastique

Dans l'approche ci-dessus, un chemin est représenté par une trajectoire numérique, et est donc une suite de configurations séparées par un temps Δt. En particulier, la mesure sur l'espace des chemins dépend du pas de temps choisi, et les résultats ne sont pas formulés de manière intrinsèque. D'un point de vue mathématique, il est intéressant de formuler le problème de l'échantillonnage des chemins de transition d'un point-de-vue continu. Dans la formulation de Hairer, Stuart, Voss et Wiberg [147,330], le problème d'échantillonnage des chemins liant un état initial x_0 et un état final x_1 est formulé comme une équation aux dérivées partielles stochastique (EDPS). Ensuite seulement, on discrétise cette EDPS pour calculer concrètement des chemins de transition. De cette manière, on peut proposer des algorithmes numériques plus efficaces (voir Beskos, Roberts, Stuart et Voss [29]).

Dynamiques accélérées

Plusieurs techniques ont été développées pour accélérer les simulations de dynamique moléculaire. On présente ici trois stratégies, proposées par Voter, de la plus rigoureuse à la moins rigoureuse (au sens où les méthodes reposant sur le moins d'hypothèses sont présentées en premier).

Hyperdynamique

L'hyperdynamique (*Hyperdynamics*), introduite par Voter dans [365], est réminiscente des techniques d'échantillonnage d'importance utilisées pour le calcul de différences d'énergie libre

(voir Section 2.3.1). L'idée est de considérer un potentiel biaisant $\Delta V \geq 0$ (éventuellement, uniquement sur certains degrés de liberté), n'agissant que dans le voisinage des minima locaux d'énergie potentielle. Ainsi, la dynamique près des point-selles n'est pas modifiée. De cette manière, le système passe moins de temps près des minima d'énergie locaux, et plus de temps près des régions de transition. Le temps physique effectivement simulé est alors

$$t_{\text{hyper}} = \int_0^t \mathrm{e}^{\beta \Delta V(q_s)}\, ds \geq t,$$

d'où un facteur de gain $t_{\text{hyper}}/t \geq 1$. Le principal défi de cette méthode, comme pour toutes les méthodes de fonction d'importance, est de construire un potentiel biaisant efficace pour des systèmes de grande dimension. Des propositions en ce sens sont faites dans [364] (utilisant un potentiel biaisant fondé sur la matrice hessienne locale) et dans [243] (sous l'hypothèse que les transitions peuvent être détectées comme des changements significatifs dans certaines longueurs de liaison).

Dynamique sur les répliques parallèles

La méthode des répliques parallèles (*Parallel Replica dynamics*) a été proposée par Voter dans [366], et permet de paralléliser l'évolution en temps pour un système dont l'évolution globale est pilotée par des événements rares. On a dans ce cas un coût de calcul décroissant (presque) linéairement avec le nombre de processeurs.

On suppose que les temps de sorties du système de tous les états métastables sont distribués selon une loi exponentielle. Le principe de la méthode repose sur la remarque mathématique suivante : la somme de M variables aléatoires indépendantes et identiquement distribuées selon une loi exponentielle de paramètre τ, est encore une variable aléatoire distribuée selon une loi exponentielle, mais de paramètre $M\tau$. Ainsi, au lieu de simuler un seul système et d'attendre une transition (ce qui donne un temps de transition T^1), il est équivalent de simuler M répliques indépendantes du système (en parallèle), d'arrêter toutes les simulations dès qu'une transition a lieu, et d'incrémenter le temps physique total de la simulation de la somme de tous les temps de simulation T_i de la i-ème réplique (ce qui donne un temps total effectif $T^M = T_1 + \cdots + T_M$). On peut même utiliser des processeurs de vitesses différentes.

En pratique, on détecte les transitions en effectuant régulièrement une minimisation locale de l'énergie, et en vérifiant que la géométrie du minimum local est inchangée. Lorsqu'une transition est détectée, on prolonge la simulation jusqu'au minimum suivant, puis on arrête toutes les simulations et on réplique le système ayant subi la transition. On effectue ensuite une phase de décorrélation entre les systèmes (échantillonnage local autour du nouveau minimum), avant de relancer la procédure parallèle d'attente de transitions. On trouvera des applications de cette méthode dans [367].

Dynamique accélérée en température

La dynamique accélérée en température (*Temperature accelerated dynamics*) a été proposée par Sorensen et Voter [317]. Cette technique peut être utilisées avec des systèmes pour lesquels la TST harmonique est une bonne approximation. Une application typique est le cas d'un matériau soumis à des dommages d'irradiation dont on cherche à simuler le comportement en temps long à température ambiante. Pour ce faire, on simule le système à une température T_+ assez grande, alors qu'on est intéressé par la dynamique à une température $T_- < T_+$. Partant d'un système dans un état métastable, les tentatives de sorties hors de cet état métastable sont interceptées, et le taux de réaction correspondant, donné par la TST harmonique, est calculé. Plus précisément, pour la i-ème tentative de sortie au temps t_+^i, le taux de réaction est

$$k_+^i = \nu_0^i \mathrm{e}^{-E_a^i/k_\mathrm{B}T_+}.$$

Le temps de transition t_-^i correspondant à la température T_-^i est alors

$$t_-^i = t_+^i \exp\left(\frac{E_a^i}{k_\mathrm{B}}\left(\frac{1}{T_-} - \frac{1}{T_+}\right)\right).$$

Une fois ce temps calculé, on renvoie le système dans l'état métastable, et la simulation continue. Si on suppose que tous les préfacteurs ν_0^i sont bornés inférieurement, on peut alors donner *a priori* un temps de simulation maximal tel que toutes les transitions possibles à température basse aient eu lieu, avec une probabilité donnée (par exemple, 95%). Finalement, à la fin de la simulation à la température haute, et après calcul de tous les temps de réaction associés à la température physique T_-, on sélectionne l'événement ayant le plus petit temps d'occurrence t_-^i, on avance le temps de simulation de t_-^i, et on effectue la transition associée à l'événement i. La principale limitation pratique est que la température T_+ ne peut être choisie trop grande sans quoi l'approximation de la TST harmonique n'est plus valide.

2.3.3 Dynamiques réduites

Il existe des dynamiques posées dans des espaces de grande dimension qui admettent une expression réduite dans certains régimes limites. Ceci peut être montré rigoureusement pour des systèmes modèles (voir ci-dessous le cas d'un couplage avec un bain déterministe composé d'oscillateurs harmoniques), mais n'est pas possible en général. Dans ce cas, une analyse formelle suggère souvent une forme raisonnable pour la dynamique réduite, et on cherche ensuite à bien choisir les paramètres des modèles correspondants pour reproduire au mieux les résultats de simulation obtenus avec les modèles complets.

Dynamiques réduites dans le cas d'un couplage avec un bain déterministe

Pour une particule couplée à de nombreux oscillateurs harmoniques, Zwanzig [379] a formellement montré que la dynamique limite sur la particule couplée est une équation de Langevin généralisée (à mémoire). Cette preuve formelle a été justifiée d'un point de vue mathématique par Kupferman, Stuart, Terry et Tupper pour un couplage harmonique avec les particules du bain [199], qui précisent notamment en quel sens entendre la convergence formelle de [379].

Dans [P11], j'ai proposé un tel modèle de couplage avec des degrés de liberté harmoniques, qui a servi de base à un modèle simplifié pour les ondes de choc. Ce modèle simplifié est unidimensionnel, mais capture quelques effets des dimensions supérieures, en particulier une sorte de relaxation de l'énergie au passage du choc, ce qui lui permet de corriger le comportement non-physique des modèles microsopiques d'ondes de choc purement unidimensionnels (voir également la Section 5.1 pour plus de précisions sur la dynamique limite correspondante dans le cas où le nombre de degrés de liberté du bain tend vers l'infini).

Dynamique de particules dissipatives

La dynamique de particules dissipatives (*Dissipative Particle Dynamics*, DPD) a été introduite par Hoogerbrugge et Koelman [170], puis étendue par Español et Warren [98] de manière à préserver la mesure canonique. L'objectif initial des modèles de type DPD est la simulation de fluides complexes, la philosophie de ces modèles reposant heuristiquement sur le remplacement de petits échantillons de fluide composés de molécules se déplaçant de manière coordonnée (*blob*) par des particules mésoscopiques ponctuelles avec interactions de paire. Ces interactions sont des forces conservatives, ainsi que des termes de dissipation visqueuse entre particules voisines, compensés par des forces aléatoires. Les modèles DPD sont conceptuellement souples, et n'ont pas en particulier d'échelles temporelle et spatiale clairement établies *a priori* (on ne dit pas à quel point il sont mésoscopiques...).

Les modèles de type DPD peuvent être dérivés de modèles tous-atomes microscopiques dans le cas d'une chaîne d'atomes harmonique unidimensionnelle [94] (voir également l'équation limite obtenue pour le modèle unidimensionnel de [P11], qui est de type DPD généralisé, c'est-à-dire à mémoire). Dans un contexte plus général, Flekkoy, Coveney et De Fabritiis [106] ont proposé une motivation fondée sur l'utilisation de cellules de Voronoï. Dans tous les cas, la dynamique déterministe tous-atomes est remplacée par une dynamique stochastique, dont la partie déterministe provient du comportement moyen du système, alors que la partie stochastique modélise les fluctuations autour du comportement moyen, dues aux degrés de liberté non explicitement traités.

La mesure d'équilibre de la dynamique (voir (2.46) ci-dessous) montre que la partie conservative de la dynamique rend compte des propriétés thermodynamiques moyennes du système, alors que la dissipation et les fluctuations aléatoires caractérisent la viscosité du système [97]. La dynamique DPD est plus précisément

$$\begin{cases} dq_i = \dfrac{p_i}{m_i}\,dt, \\ dp_i = \displaystyle\sum_{j\neq i} -\nabla_{q_i}\mathcal{V}(r_{ij})\,dt - \gamma\chi^2(r_{ij})(v_{ij}\cdot e_{ij})e_{ij}\,dt + \sqrt{\frac{2\gamma}{\beta}}\chi(r_{ij})dW_{ij}\,e_{ij}, \end{cases}$$

avec

$$\gamma > 0, \quad r_{ij} = |q_i - q_j|, \quad e_{ij} = \frac{q_i - q_j}{r_{ij}}, \quad v_{ij} = \frac{p_i}{m_i} - \frac{p_j}{m_j},$$

où χ désigne une fonction de poids (à support dans une boule de rayon r_c, r_c étant un rayon de coupure), et où les processus de Wiener unidimensionnels W_{ij} standards sont tels que $W_{ij} = -W_{ji}$. Comme toutes les interactions sont des interactions additives de paire (y compris les termes de fluctuation/dissipation), la dynamique DPD est telle que l'impulsion totale du système et le moment angulaire total sont conservés.

Notant $H(q,p) = \frac{1}{2}p^T Mp + V(q)$ le Hamiltonien du système, et $V(q) = \sum_{1\leq i<j\leq N}\mathcal{V}(r_{ij})$, on montre facilement que la mesure

$$d\mu(q,p) = \frac{1}{Z}\exp\left(-\beta H(q,p)\right)\,dq\,dp \tag{2.46}$$

(Z étant une constante de normalisation) est une mesure de probabilité invariante de la dynamique (5.35), car cette mesure est une solution stationnaire de l'équation de Fokker-Planck associée à (5.35) (voir [98]). Cependant, il est délicat de montrer l'ergodicité de la dynamique DPD. Citons toutefois le résultat de Shardlow et Yan qui montre l'ergodicité de la dynamique DPD lorsqu'elle posée dans un tore unidimensionnel (sous certaines conditions sur le potentiel d'interaction et les fonctions de poids, et à condition que la densité soit suffisamment grande).

Notons enfin que les modèles de type DPD sont proches à la fois des modèles microscopiques tous-atomes usuellement utilisés en dynamique moléculaire, et des discrétisations particulaires de l'équation de Navier-Stokes, telles que la *Smoothed Particle Hydrodynamics* de Lucy et Monaghan [217, 246]. Un premier pas vers un formalisme général permettant d'envisager un couplage de modèles est proposé par Español et Revenga dans [96].

Potentiels d'interaction entre particules

Le choix d'un potentiel d'interaction décrivant les interactions entre les mésoparticules est une question importante en pratique. Par ordre de technicité et de complexité croissante, on peut citer plusieurs approches :

(i) l'utilisation d'une force moyenne, qui peut être une moyenne de type thermodynamique (on obient alors la force moyenne et le potentiel de force moyenne) [97, 142], ou un autre type de moyenne (moyenne sur les temps courts) [109]. La force moyenne exercée par une particule q_2 sur une particule q_1 est définie par la moyenne sur les configurations des

particules restantes (de manière générale, pour la force entre deux molécules, on considère une moyenne des forces entre les centres de masses, en moyennant sur les configurations des autres particules) [142] :

$$-\nabla_{q_1}\mathcal{V}(q_1,q_2) = \frac{\displaystyle\int -\nabla_{q_1}V(q)\,\mathrm{e}^{-\beta V(q)}\,dq_3\ldots dq_N}{\displaystyle\int \mathrm{e}^{-\beta V(q)}\,dq_3\ldots dq_N} = -\frac{1}{\beta}\nabla_{q_1}\left[\ln g(|q_1-q_2|)\right],$$

où $g(r)$ est la fonction de distribution de paires ;

(ii) la recherche d'un potentiel de paire optimal, selon des critères à définir, et éventuellement selon une forme fonctionnelle donnée ;

(iii) des potentiels effectifs plus complexes, par exemple anisotropes (pour prendre en compte des effets stériques par exemple).

Beaucoup d'approches relèvent de la seconde catégorie, et tombent sous le thème général de l'optimisation de potentiel. Les quantités que l'on cherche à reproduire au mieux sont souvent des quantités d'équilibre (propriétés structurelles et/ou moyennes thermodynamiques), en particulier la fonction de paire radiale $g(r)$ (voir les méthodes d'inversion Monte-Carlo [219, 279] reposant sur une bijection entre un potentiel de paire et le $g(r)$ associé [155]), les lois d'état (pression en fonction de la densité, etc ; voir par exemple [245] pour un protocole modèle) ou certains coefficients thermodynamiques dérivés (compressibilité, etc). Parfois on cherche également à reproduire des coefficients de transports (auto-diffusion dans les fluides en particulier). Un point important à noter toutefois est que ces différents potentiels effectifs dépendent en général des conditions thermodynamiques dans lesquelles ils sont obtenus (ceci est immédiat pour le potentiel de force moyenne, mais reste vrai pour tous les autres également).

Application aux ondes de choc et de détonation

Il existe de nombreux raffinements et variantes de la dynamique DPD (2.46). En particulier, on peut considérer des modèles DPD où les particules mésoscopiques possèdent des degrés de liberté internes. Ainsi, les modèles de type DPDE (DPD à énergie conservée), proposés indépendamment par Avalos et Mackie [15] et Español [95] sont des dynamiques stochastique préservant exactement l'énergie totale du système. Dans ce cas, chaque particule (q_i, p_i) a une énergie interne ϵ_i, et l'énergie totale du système est $H(q,p) + \sum_{i=1}^{N}\epsilon_i$. L'énergie mécanique dissipée est compensée par des variations des énergies internes.

Dans [P7], j'ai utilisé une dynamique DPDE un peu simplifiée pour proposer un modèle mésoscopique pour la simulation des ondes de choc. L'idée fondatrice de ce modèle, déjà utilisée par Strachan et Holian [326], est de remplacer les molécules complexes que l'on doit simuler par une particule mésoscopique, ayant une énergie interne $\epsilon = N_{\mathrm{red}}k_{\mathrm{B}}T_{\mathrm{int}}/2$, où T_{int} est la température interne de la particule, et N_{red} le nombre de degrés de liberté non représentés explicitement (pour une molécule formée de N_{at} atomes en dimension d, on a $N_{\mathrm{red}} = 2d(N_{\mathrm{at}}-1)$). Les résultats de simulations montrent qu'on peut en effet obtenir un bon accord avec des modèles tous-atomes (voir [P7] et la Section 5.2.2 pour plus de précisions).

Le formalisme DPD permet également de proposer une extension au cas d'ondes de détonation, c'est-à-dire d'ondes de choc qui initient des réactions chimiques à leur passage, ces réactions soutenant et augmentant l'intensité de ladite onde de choc. La modélisation de la détonation au niveau mésoscopique demande d'introduire une nouvelle variable par particule, un taux de réaction fictif λ. La dynamique peut être décomposée selon trois processus physiques :

(i) une dynamique simple sur (q, p, ϵ), analogue à ce qui se passerait pour un matériau inerte ;

(ii) une réaction chimique déterminant l'évolution de λ, et dont il faut préciser l'avancement ;

(iii) enfin, la modélisation de l'exothermicité de la réaction, *i.e.* la répartition de l'énergie libérée par la réaction chimique dans les degrés de liberté du modèle.

Nous avons proposé un tel modèle avec J.-B. Maillet et L. Soulard (voir [P2] et Section 5.2.3), et les premiers résultats numériques qu'il donne sont encourageants.

Diffusion effective pour la coordonnée de réaction

Cette section présente un domaine de recherche intéressant suite aux études précédentes : la détermination d'une dynamique moyenne ou effective de la coordonnée de réaction. En effet, dans la mesure où la coordonnée de réaction est censée représenter une variable macroscopique, ou au moins évoluant lentement par rapport aux autres degrés de liberté du système, il est naturel de chercher une équation effective sur cette variable, où les degrés de liberté non explicitement représentés n'apparaîtraient que par le biais d'effets moyens, plus éventuellement des perturbations stochastiques. On peut ainsi distinguer deux problèmes successifs dans l'obtention d'une dynamique effective pour la coordonnée de réaction : la forme de cette dynamique, selon que la dynamique à réduire soit hamiltonienne ou stochastique, puis, une fois la forme générale de cette dynamique obtenue, l'estimation des paramètres intervenant dans la description. Nous détaillons successivement ces étapes.

Réduction de la dynamique hamiltonienne

Une manière générale de réduire un système déterministe pour obtenir une dynamique effective sur un sous-ensemble des degrés de liberté uniquement, est de recourir à l'opérateur de projection introduit par Mori et Zwanzig [250, 379]. L'idée de cette technique est d'intégrer exactement (quoique formellement) les degrés de liberté non désirés, qui apparaissent alors par le biais d'une mémoire dans la dynamique des degrés de liberté auxquels on s'intéresse, plus éventuellement un terme de forçage aléatoire, lié typiquement à la connaissance imparfaite au temps initial de degrés de liberté non représentés. Nous exposons les grandes lignes de cette dérivation (ainsi que présentée par Givon, Kupferman et Stuart dans [134]), dans le cas où $q = (x, y)$ avec $x \in \mathbb{R}^m$, $y \in \mathbb{R}^{dN-m}$ – c'est-à-dire qu'on ne se soucie pas des problèmes de géométrie rencontrés avec une coordonnée de réaction $\xi : \mathbb{R}^{dN} \to \mathbb{R}^m$ générale (pour ce dernier cas, on pourra consulter [136]). On note également $p = (p_x, p_y)$ l'impulsion associée à q.

De manière générale, pour $(x, y) \in \mathcal{X} \times \mathcal{Y}$ évoluant selon la dynamique

$$\begin{cases} \dot{X} = f(X, Y), \\ \dot{Y} = g(X, Y), \end{cases} \tag{2.47}$$

dont $d\rho(X, Y)$ est une mesure (positive bornée) invariante, on introduit les opérateurs de projection Π et P tels que

$$\Pi(X, Y) = X, \qquad Pf(X) = \frac{\displaystyle\int_{\mathcal{Y}} f(X, Y)\, d\rho(X, Y)}{\displaystyle\int_{\mathcal{Y}} d\rho(X, Y)}.$$

On peut alors récrire la solution de (2.47) sous la forme

$$\dot{X}(t) = Pf(X(t)) + \int_0^t K(X(t-s), s)\, ds + n(X(0), Y(0), t).$$

Le terme de forçage n et le terme de mémoire K sont liés par une relation de fluctuation/dissipation, et sont définis respectivement par l'équation

$$\partial_t n = (\mathrm{Id} - P)\mathcal{L}n, \qquad n(X,Y,0) = f(X,Y) - Pf(X),$$

et la relation

$$K(X,t) = P\mathcal{L}n(X,Y,t),$$

où $\mathcal{L}$ est l'opérateur de Liouville $\mathcal{L} = f(X,Y)\cdot\nabla_X + g(X,Y)\cdot\nabla_Y$.

Dans le cas de la dynamique hamiltonienne, on peut supposer ques les conditions initiales sont distribuées selon la mesure canonique, ce qui fixe le choix de la mesure utilisée pour la projection P. [11] On a alors la dynamique projetée suivante, définie sur $\mathbb{R}^{2m}$:

$$\frac{d}{dt}\begin{pmatrix} x \\ p_x \end{pmatrix} = \begin{pmatrix} M_x^{-1}p_x \\ -\nabla_x \overline{V}(x) \end{pmatrix} + \int_0^t K((x,p_x)(t-s),s)\,ds + n(x(0),p_x(0),y(0),p_y(0),t), \tag{2.48}$$

où M_x est la matrice de masse associée à la variable p_x, et $\overline{V}(x)$ est le potentiel de force moyenne :

$$\overline{V}(x) = -\frac{1}{\beta}\ln\int_{\mathbb{R}^{dN-m}} \mathrm{e}^{-\beta V(x,y)}\,dy. \tag{2.49}$$

Ainsi, la dynamique effective est une dynamique hamiltonienne, avec deux termes supplémentaires, un terme de mémoire, et un terme dû aux conditions initiales. Ce dernier terme est un terme de forçage aléatoire lorsque les conditions initiales sont aléatoires (et dans la limite $N \to +\infty$, voir par exemple [199] pour une preuve rigoureuse dans un cas simple).

Il est important de remarquer toutefois que l'équation limite (2.48) obtenue par cette technique de projection n'est pas plus simple que l'équation hamiltonienne posée dans $\mathbb{R}^{dN}$ (le terme de mémoire n'étant pas explicite). En pratique, elle sert toutefois de point de départ pour proposer des approximations de la dynamique sur la coordonnée de réaction.

Réduction de dynamiques stochastiques

Certains travaux partent d'une dynamique stochastique tous-atomes, et cherchent à en extraire une dynamique réduite pour certains degrés de liberté. On présente ici un cas classique d'une telle réduction dans le cas simple $q = (x,y)$ avec $x \in \mathbb{R}^m$, $y \in \mathbb{R}^{dN-m}$, pour la dynamique

$$dq_t = -\nabla V(q_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t,$$

W_t étant un mouvement brownien standard de dimension dN. Si on a bien choisi une partition du système en variables lentes x et variables rapides y, les variables y ne sont présentes que par le biais d'une action moyenne ressentie par les variables x. On peut rendre cette idée rigoureuse en accélérant artificiellement le temps dans la dynamique en y selon ($\epsilon > 0$)

$$\begin{cases} dx_t^\epsilon = -\nabla_x V(x_t^\epsilon, y_t^\epsilon)\,dt + \sqrt{2\beta^{-1}}\,dW_t^x, \\ dy_t^\epsilon = -\dfrac{1}{\epsilon}\nabla_y V(x_t^\epsilon, y_t^\epsilon)\,dt + \sqrt{\dfrac{2\beta^{-1}}{\epsilon}}\,dW_t^y, \end{cases}$$

où W_t^x, W_t^y sont des mouvements browniens standard indépendants, de dimension respectives m et $dN - m$. Dans la limite $\epsilon \to 0$, on obtient une dynamique effective sur x de la forme

$$dX_t = -\nabla_x \overline{V}(X_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t, \tag{2.50}$$

W_t étant un mouvement brownien standard de dimension m, et $\overline{V}(x)$ le potentiel de force moyenne (2.49) (voir l'article de Papanicolaou [266] et l'ouvrage pédagogique de Pavliotis

[11] Rappelons en effet que la dynamique hamiltonienne préserve d'autres mesures, telles que la mesure de Lebesgue

et STUART [268, Chapitres 10 et 11] pour plus de précisions sur le sens et la validité de cette convergence). Cette approche peut être étendue à des coordonnées de réaction générales (voir [91, Section 10]). Dans ce cas, on trouve une dynamique limite de la forme générale

$$dX_t = f(X_t)\,dt + \sigma(X_t)\,dt, \tag{2.51}$$

les fonctions f et σ dépendant du choix de la coordonnée de réaction.

Une dérivation alternative de la dynamique (2.51) repose sur le travail de GYÖNGY [144]. En effet, pour une coordonnée de réaction $\xi : \mathbb{R}^{dN} \to \mathbb{R}$, on a par le calcul d'Itô, partant de (2.50),

$$d\xi(q_t) = (-\nabla V(q_t)\cdot\nabla\xi(q_t) + \beta^{-1}\Delta\xi(q_t))\,dt + \sqrt{2\beta^{-1}}\,|\nabla\xi(q_t)|\frac{\nabla\xi(q_t)\cdot dW_t}{|\nabla\xi(q_t)|}.$$

Introduisant la dynamique

$$dX_t = f(t,X_t)\,dt + \sigma(t,X_t)\,dB_t, \qquad dB_t = \frac{\nabla\xi(q_t)\cdot dW_t}{|\nabla\xi(q_t)|}$$

avec

$$f(t,z) = \mathbb{E}\left(-\nabla V(q_t)\cdot\nabla\xi(q_t) + \beta^{-1}\Delta\xi(q_t)\;|\,\xi(q_t)=z\right), \quad \sigma(t,z) = \mathbb{E}\left(|\nabla\xi(q_t)|\;|\,\xi(q_t)=z\right),$$

les résultats de [144] montrent que les lois de X_t et de $\xi(q_t)$ sont les mêmes. On retrouve une dynamique de la forme (2.51) si on suppose que les lois conditionnelles de q_t ne dépendent pas du temps, et sont à l'équilibre pour la mesure canonique, auquel cas les espérances conditionnelles ci-dessus peuvent être calculées selon

$$\mathbb{E}(h(q_t)\;|\,\xi(q_t)=z) = \frac{\displaystyle\int_{\xi^{-1}(z)} h(q)\mathrm{e}^{-\beta V(q)}|\nabla\xi(q)|^{-1}\,dq}{\displaystyle\int_{\xi^{-1}(z)} \mathrm{e}^{-\beta V(q)}|\nabla\xi(q)|^{-1}\,dq}.$$

Estimation des paramètres de la dynamique limite

Selon que l'on parte d'un système déterministe ou stochastique, on peut, par une procédure de projection adéquate, obtenir une dynamique effective de type (2.48) (dynamique de Langevin généralisée, *i.e.* à mémoire), ou (2.51) (avec un bruit multiplicatif). Dans les deux cas, une simulation effective de ces dynamiques demande une étape préliminaire d'estimation des paramètres.

Pour les dynamiques de type (2.48), une approche usuelle est de postuler une forme fonctionnelle pour le potentiel de force moyenne $\overline{V}$, le terme de mémoire et le terme de bruit. L'estimation des paramètres correspondants peut alors être faite à partir d'un échantillon de valeurs observées de la coordonnée de réaction en utilisant des techniques statistiques telles que des estimateurs de maximum de vraisemblance (voir par exemple l'article de revue de BIBBY et SORENSEN [30] sur l'estimation de paramètres pour des diffusions elliptiques, ou le travail de POKERN, STUART et WIBERG [271] dans le cas de processus hypoelliptiques). Ces estimations statistiques permettent également de valider ou d'invalider la forme fonctionnelle proposée pour les différents termes.

On peut aussi utiliser les techniques statistiques ci-dessus pour les dynamiques de la forme (2.51) (voir par exemple HUMMER [176]). Pour l'instant, les approches employées reposent toutefois plutôt sur les méthodes dites *sans-équations* (voir, dans le contexte des dynamiques effectives, l'article de KOPELEVICH, PANAGIOTOPOULOS et KEVREKIDIS [196], ainsi que [373]). Le principe de ces méthodes est de partir d'un ensemble de configurations microscopiques indépendantes pour une valeur fixée de la coordonnée de réaction, et de regarder l'évolution en temps courts de la distri-

bution des valeurs de la coordonnée de réaction pour en déduire des approximations du terme de dérive f et du bruit multiplicatif σ dans (2.51).

Part II

Sampling Techniques in Molecular Dynamics

3

Phase-space sampling techniques

3.1 Purely stochastic methods **56**
3.1.1 Rejection method 56
3.1.2 Rejection control 58
3.1.3 Metropolized independence sampler 58
3.1.4 Importance sampling 62
3.2 Stochastically perturbed Molecular Dynamics methods **62**
3.2.1 General framework for NVE Molecular Dynamics 63
3.2.2 Hybrid Monte Carlo 63
3.2.3 Biased Random-Walk 75
3.2.4 Langevin dynamics 78
3.3 Deterministic molecular dynamics sampling **83**
3.3.1 The Nosé-Hoover and Nosé-Hoover chains methods 83
3.3.2 The Nosé-Poincaré and the Recursive Multiple Thermostat methods . . 84
3.4 Numerical illustrations **85**
3.4.1 Description of the linear alkane molecule 86
3.4.2 Discrepancy of sample points 87
3.4.3 Choice of parameters 89
3.4.4 Numerical results 93
3.4.5 Improvement of the convergence rates 94
3.4.6 Computation of correlation functions 96
3.5 Stochastic boundary conditions **96**
3.5.1 Review of some classical stochastic boundary conditions 97
3.5.2 An example of thermal boundary conditions 99
3.6 Some background on continuous state-space Markov chains and processes **105**
3.6.1 Some background on continuous state-space Markov chains 105
3.6.2 Some convergence results for Markov processes. 114

In this chapter, we present and compare, from both a theoretical and a numerical point of view, sampling methods to compute phase space integrals of the form

$$\langle A \rangle = \int_{T^*\mathcal{M}} A(q,p)\, d\mu(q,p), \tag{3.1}$$

or time-dependent properties

$$\langle B \rangle(t) = \int_{T^*\mathcal{M}} B(\Phi_t(q,p),(q,p))\, d\mu, \tag{3.2}$$

where Φ_t is the Hamiltonian flow. In the above expression, $\mathcal{M}$ denotes the position space (also called the *configuration space*), and $T^*\mathcal{M}$ its cotangent space. A generic element of the position space $\mathcal{M}$ will be denoted by $q = (q_1, \cdots, q_N)$ and a generic element of the momentum space $\mathbb{R}^{3N}$ by $p = (p_1, \cdots, p_N)$. The mass matrix is $M = \mathrm{Diag}(m_1, \ldots, m_N)$. The measure μ is the canonical probability measure:

$$d\mu(q,p) = Z^{-1} \exp(-\beta H(q,p))\, dq\, dp, \tag{3.3}$$

where $\beta = 1/k_{\mathrm{B}}T$ (T denotes the temperature and k_{B} the Boltzmann constant) and where H denotes the Hamiltonian of the molecular system:

$$H(q,p) = \frac{1}{2} p^T M^{-1} p + V(q). \tag{3.4}$$

Recall that the measure $d\mu(q,p)$ can be written as $d\mu(q,p) = d\pi(q)\, d\kappa(p)$ with

$$d\kappa(p) = \mathcal{P}(p)\, dp = Z_p^{-1} \exp\left(-\frac{\beta}{2} p^T M^{-1} p\right) dp, \tag{3.5}$$

and

$$d\pi(q) = f(q)\, dq = Z_q^{-1} \mathrm{e}^{-\beta V(q)}\, dq. \tag{3.6}$$

Since it is straightforward to sample from the momentum distribution (3.5) (it is a product of independent Gaussian densities), the actual issue is to sample efficiently from the (position space) measure π given by (3.6).

In this chapter, new convergence results on the Hybrid Monte-Carlo sampling scheme are stated (see Section 3.2.2) and various numerical methods to compute integrals such as (3.1) or (3.2), are reviewed and their efficiencies are compared on a benchmark system (simple alkane molecule). More precisely, we consider the issue of sampling from the canonical measure (3.3).

All the methods considered in this chapter consist in generating a sequence of points $(q^n)_{n\in\mathbb{N}}$ in the position space. These methods can be classified in four categories:

Type 1. $(q^n)_{n\in\mathbb{N}}$ is a sequence of independent realizations of a given random variable of density $f(q) = \dfrac{1}{Z_q} \mathrm{e}^{-\beta V(q)}$; this is the case for the standard Rejection and for the Rejection control methods;

Type 2. $(q^n)_{n\in\mathbb{N}}$ is a realization of a continuous state-space Markov chain, for which π is an invariant measure; this is the case for the Metropolized independence sampler and for the Hybrid Monte Carlo method;

Type 3. $(q^n)_{n\in\mathbb{N}}$ is an approximation of $(q_{t_n})_{n\in\mathbb{N}}$ where $(q_t)_{t\geq 0}$ (resp. $(q_t, p_t)_{t\geq 0}$) is a sample path of a stochastic process on $\mathcal{M}$ (resp. on $T^*\mathcal{M}$), for which π (resp. μ) is an invariant measure; this is the case for the biased Random-Walk (resp. for the Langevin dynamics);

Type 4. $(q^n)_{n\in\mathbb{N}}$ is an approximation of $(q(t_n))_{n\in\mathbb{N}}$ where $(q(t), p(t), x(t))_{t\geq 0}$ is a trajectory of a deterministic extended dynamical system (q and p are the physical variables, while x represents some additional variables; see Section 3.3 for more details); this extended dynamical system is such that it preserves a measure $d\rho$ whose projection on the physical variables q, p is the measure $d\mu$ given by (3.3); this is the case for Nosé-Hoover, Nosé-Poincaré and Recursive Multiple Thermostat methods.

The first two questions we will adress are relevant for all the methods mentioned above:

Question **1.** An observable $A(q)$ on $\mathcal{M}$ being given, does the empirical mean $\frac{1}{N}\sum_{n=0}^{N-1} A(q^n)$ converge to the space average $\int_{\mathcal{M}} A(q)\, d\pi(q)$?

Question **2.** If so, can the speed of convergence be estimated?

For methods of Type 1, the answers to Questions 1 and 2 are obviously positive and are direct consequences of the Law of Large Number (LLN) and of the Central Limit Theorem (CLT) for independent identically distributed (i.i.d.) random variables. For the methods of Type 2, Questions 1 and 2 can be positively answered, at least for compact position spaces $\mathcal{M}$ and under some assumptions on the potential energy V. For Question 1, the point is to check (see Theorem 3.1 and Section 3.6 below) that

$$\pi \text{ is an invariant probability measure of the Markov chain,} \tag{3.7}$$

and that the probability transition kernel $P(q,\cdot)$ of the Markov chain[1] satisfies the accessibility condition

$$\forall q \in \mathcal{M}, \quad \forall B \in \mathcal{B}(\mathcal{M}), \quad \mu^{\mathrm{Leb}}(B) > 0 \;\Rightarrow\; P(q,B) > 0, \tag{3.8}$$

where $\mathcal{B}(\mathcal{M})$ is the Borel σ-algebra of $\mathcal{M}$ and μ^{Leb} is the Lebesgue measure on $\mathcal{M}$. Turning to Question 2, a convergence rate of $N^{-1/2}$ can be obtained when the transition kernel P has some regularity properties, and provided some Lyapunov condition holds true (see Theorem 3.2 and condition (3.11) below).

For the methods of Type 3, analogous results can be stated at the continuous level (for the underlying Markov processes). In computations, discrete-time approximations are used, and one recovers the case of a Markov chain, and the same kind of results as for methods of Type 2 hold true. For methods of Type 4, no general convergence result is known.

In the case when the sequence $(q^n)_{n\in\mathbb{N}}$ originates from a Markov chain on $\mathcal{M}$ or from a discretized stochastic process on $\mathcal{M}$ or on $T^*\mathcal{M}$ (methods of Types 2 and 3), additional questions arise. Indeed, instead of considering *one* realization starting from a given initial data, it is also possible to generate samples with the same computational cost by considering *several* shorter realizations starting either all from the same point or from different points (which constitute a pre-existing initial distribution). In this case, typical convergence results involve weighted total variation norms for the probability measures that are generated. In the sequel, we will often refer to this kind of convergence as the "convergence of densities" since, when the n-step probability transition kernel[2] $P^n(q,\cdot)$ of the Markov chain and the invariant probability measure both admit densities with respect to the Lebesgue measure, the convergence in total variation norm implies the L^1 convergence of the densities. We can thus formulate the following two questions:

Question 3. Does $\|P^n(q,\cdot) - \pi\|$ converge to zero when n goes to infinity for some (weighted) total variation norm?

Question 4. If so, can the speed of convergence be estimated?

Again, if π is an invariant probability measure and if the accessibility condition (3.8) holds true, the answer to Question 3 is positive (see Theorems 3.3 and 3.4 below). A geometric convergence rate in ρ^n for some $\rho \in (0,1)$ in some weighted total variation norm can also be obtained when the

[1] If $q \in \mathcal{M}$ and B is a Borel set of $\mathcal{M}$, $P(q,B)$ is the probability for the Markov chain to be in B when starting from q.

[2] For $q \in \mathcal{M}$ and B a Borel set of $\mathcal{M}$, $P^n(q,B)$ is the probability for the Markov chain to be in B when starting from q after exactly n steps. It is inductively defined from P by $P^0(q,B) = \mathbf{1}_B(q)$ and the induction rule

$$P^n(q,B) = \int_{\mathcal{M}} P(q,dq')P^{n-1}(q',B).$$

transition kernel P has some weak regularity properties and provided some Lyapunov condition holds true (namely condition (3.31) below, see Theorem 3.8). Let us point out that the Lyapunov condition (3.31) providing geometric convergence of the densities is not of the same nature as the condition (3.11) providing a convergence rate of the average along one sample path.

Let us mention that, in some applications, integrals such as (3.1) are sometimes computed using Blue Moon sampling techniques [54, 65, 370]. In this case, integrals over submanifolds (generally hypersurfaces) of $\mathcal{M}$ have to be estimated. For such computations, the theoretical analysis is the same as the one presented here. From the numerical viewpoint, algorithms adapted to the constraint of sampling a hypersurface (and not the whole space) have to be used, namely projected algorithms for stochastic dynamics (see e.g. [66] and Section 4.1.3) and SHAKE or RATTLE algorithms for deterministic evolutions (see [146, Chap. VII.1.4]).

This chapter is organized as follows. We first describe and compare from a theoretical point of view the most popular methods to sample from the canonical distribution. In Section 3.1, we consider purely stochastic methods; stochastically perturbed Molecular Dynamics methods and deterministic thermostatting methods are presented in Section 3.2 and 3.3 respectively. In particular, in Section 3.2.2, we present some new convergence results for the Hybrid Monte Carlo scheme (see Theorems 3.7, 3.9 and 3.10). A summary of the main known results is presented in Table 3.1. We refer to the corresponding sections for notations and further explanations, and to Section 3.6 for some theoretical background on Markov chains and processes.

We then turn to a practical application of those methods in the case of linear alkane molecules in Section 3.4. The fact that some methods may work better than others, and that this depends on the situation at hand, is commonly accepted. However, these beliefs are usually only based on some qualitative comparisons, or on comparison with experimental data. In the latter case, discrepancies between numerical results and experimental results can come both from numerical and modelling approximations, so it is not easy to draw conclusions specifically on the numerical methods. Comparing the methods in a *quantitative* way is one of the main purpose of this study.

Finally, an application of the previous sampling methods to compute time-dependent properties using stochastic boundary conditions is presented in Section 3.5.

3.1 Purely stochastic methods

Purely stochastic methods consist in generating points in the position space according to the measure $d\pi(q) = f(q)\,dq$ given by (3.6), without refering to any physical dynamics of the system.

We briefly recall here four methods, the Rejection, Rejection control, Importance sampling, and Metropolized sampling methods. They all make use of a reference positive probability distribution $g(q)$, such that (i) it is easy to generate samples from g, and (ii) g is a "good" approximation of f, in a sense that will be made precise below.

3.1.1 Rejection method

The Rejection method [215] requires the knowledge of a probability density g which bounds f from above up to a multiplicative factor $c > 0$:

$$f \leq cg, \tag{3.9}$$

and from which it is easy to generate samples. For instance, when $\mathcal{M} = \mathbb{T}^{3N}$ (molecular system with periodic boundary conditions) and the potential energy V is bounded from below, a uniform density g may be used (but its efficiency is likely to be very poor). The idea of the method is to draw proposals according to the density g and to accept them with probability $f/(cg)$.

Table 3.1. Summary of the different sampling methods and their properties. The following shortenings have been used: MH (Metropolis-Hastings scheme), MD (Molecular Dynamics), i.i.d. r.v. (independently and identically distributed random variables), LLN (usual Law of Large Numbers, *i.e.* for i.i.d. variables), MC LLN (LLN for Markov chains), MP LLN (LLN for Markov processes).

Name	Rejection and Rejection control	Metropolized independence sampler (MIS)	Hybrid Monte-Carlo (HMC)	Biased Random-Walk	Langevin dynamics	Deterministic dynamics
Method	Sampling from the true density	MH with independent proposals	MH with MD proposals	Elliptic diffusion	Hypoelliptic diffusion	Extended MD system
Type	i.i.d variables	Markov chain	Markov chain	Markov process	Markov process	ODE
Questions 1, 2	LLN Any textbook	MC LLN (conditions on the proposal function) Section 3.1.3 and [237]	MC LLN (conditions on the potential energy) Section 3.2.2	MP LLN (Lyapunov condition) Section 3.2.3	MP LLN (Lyapunov condition) Section 3.2.4	Open question
Questions 3, 4	-	Uniform ergodicity when a bounding function exists Section 3.1.3	Ergodicity Section 3.2.2	Geometric ergodicity (Lyapunov condition) Section 3.2.3	Geometric ergodicity (Lyapunov condition) Section 3.2.4	Open question
Numerical discretization	-	-	MH with velocity-Verlet Section 3.2.2	Euler-Maruyama or MALA Section 3.2.3	BBK algorithm or higher order schemes Section 3.2.4	Operator splitting Section 3.3
Type	-	-	Markov chain	Markov chain	Markov chain	ODE discretization
Convergence	-	-	Same techniques and results as for the continuous scheme Section 3.2.2	Classical MC techniques Section 3.2.3 and [283]	No result for usual schemes / results for specific schemes Section 3.2.4	Open question
Free parameters	Sampling function g	Proposal function g	Time step Δt, Integration time τ	Time step Δt	Time step Δt, Friction coefficient ξ	Number/values of thermostat masses, time step Δt
Rule	g "close to" f	g "close to" f	"Not too much rejection"	Acceptance rate $\simeq 0.5$	$\xi\Delta t$ "small" (0.01)	

Actually, a bound on the (non-normalized) distribution $\tilde{f}(q) = Z_q f(q) = \mathrm{e}^{-\beta V(q)}$ is sufficient to run the algorithm. Such a bound reads $\tilde{f} \leq \tilde{c}g$, and is much easier to establish in practice since the normalization constant Z_q is unknown and very difficult to estimate. The proposals are then accepted with probability $\tilde{f}/(\tilde{c}g)$.

Finding a function g such that the constant c appearing in (3.9) is as small as possible is very important. It is indeed well-known [215] that, on average, generating one sample point requires c draws, that is c evaluations of the potential energy V, which is by far the most computationally expensive part of the calculation. This constant c is therefore of paramount importance. When the system dimension is small, it is usually possible to find g such that c is not too large, and therefore the method is very efficient. But when c is very large, the method is totally inefficient. In molecular simulation, it is usually very difficult to construct efficient sampling functions g for systems involving more than a few atoms. This can however still be done for some specific systems, such as crystals at low temperature, using Taylor expansions around the equilibrium position, and controlling the relevance of the expansion by Rejection control techniques (see Section 3.1.2 below).

Since the points generated by the Rejection algorithm are independent realizations of some random variable, usual convergence results such as the Law of Large Numbers and the Central Limit Theorem apply [137]. Let A be some observable over the position space, $(q^n)_{0\leq n\leq N-1}$ be the sample generated by the method, and let us set

$$S_N(A) = \sum_{n=0}^{N-1} A(q^n). \tag{3.10}$$

If $\pi(|A|) < +\infty$, then the Law of Large Numbers holds true:

$$\lim_{N\to\infty} \frac{1}{N} S_N(A) = \int_{\mathcal{M}} A(q) f(q)\, dq = \int_{\mathcal{M}} A\, d\pi \quad \text{a.s.}$$

If $\pi(|A|^2) < +\infty$, then the Central Limit Theorem holds true. There exists $\gamma_A > 0$ (in fact,$\gamma_A = \pi(|A|^2) - \pi(|A|)^2$) such the following convergence in law holds:

$$(N\gamma_A)^{-1/2} S_N(\bar{A}) \underset{N\to\infty}{\to} \mathcal{N}(0,1),$$

where $\bar{A} = A - \displaystyle\int_{\mathcal{M}} A\, d\pi$ and $\mathcal{N}(0,1)$ is the standard Gaussian random variable.

3.1.2 Rejection control

It is often tricky to find a function g such that (3.9) is satisfied everywhere in $\mathcal{M}$. However, it is sometimes possible to find a sampling function g for which (3.9) is satisfied for most proposals $\widetilde{q}$ generated from g. In this case, the Rejection method presented in the previous section can be somewhat modified so that the non-global character of the bound is taken into account.

The Rejection control scheme [64,215] allows one to handle proposals that violate the inequality (3.9) by an appropriate *a posteriori* reweighting. Let us just note here that this scheme can be recast [64] as an Importance sampling scheme, a method we will recall in Section 3.1.4.

3.1.3 Metropolized independence sampler

When c is large, the Rejection method may require many evaluations of the potential energy V. As c is unknown in practice, it is difficult to estimate *a priori* the computational efficiency of the method. Therefore, a stochastic method with a fixed computational cost could provide an interesting alternative.

The Metropolized independence sampler (MIS), presented e.g. in [215, Section 5.4.2], is one such method. Basically, it is a Metropolis-Hastings algorithm [153, 238] with i.i.d. proposals. Therefore, the generated sequence of points forms a Markov chain (see [240] for some definitions and properties of continuous state-space Markov chains).

Metropolis-Hastings algorithm

We first recall the general idea of the Metropolis algorithm [238], which was later generalized by Hastings [153] to provide a general purpose sampling method (see also Section 4.3 and Section 6.1.1 for non trivial applications of the Metropolis-Hastings algorithm to the case of path sampling and Variational Monte Carlo respectively). We present it here on the configurational space $\mathcal{M}$, and consider that we have a rule to generate proposal configurations q' starting from the current configuration q, and that this proposal function is characterized by the probability density $\mathcal{P}(q, q')$ (It is also called 'generation probability' or 'transition density' in the field of molecular simulation).

METROPOLIS-HASTINGS ALGORITHM

Algorithm 3.1. Starting from some initial configuration q^0, and for $n \geq 1$,

(1) Propose a move from q^n to $\widetilde{q}^{n+1}$ according to the transition density $\mathcal{P}(q^n, \widetilde{q}^{n+1})$;

(2) Compute the acceptance rate

$$\alpha^n = \min\left(\frac{f(\widetilde{q}^{n+1})\,\mathcal{P}(\widetilde{q}^{n+1}, q^n)}{f(q^n)\,\mathcal{P}(q^n, \widetilde{q}^{n+1})}, 1\right);$$

(3) Draw a random variable U^n uniformly distributed in $[0, 1]$ ($U^n \sim \mathcal{U}[0, 1]$);

(i) if $U^n \leq \alpha^n$, accept the move and set $q^{n+1} = \widetilde{q}^{n+1}$;

(ii) if $U^n > \alpha^n$, reject the move and set $q^{n+1} = q^n$.

(4) go to Step (1).

We denote by P the transition kernel of this Markov chain. It is easily seen that

$$P(q, dq') = r(q, q')\mathcal{P}(q, q')\, dq' + \left(1 - \int r(q, q''')\mathcal{P}(q, q'')\, dq''\right)\delta_q,$$

where the density $r(q, \cdot)$ is given by

$$r(q, q') = \min\left(1, \frac{f(q')\mathcal{P}(q', q)}{f(q)\mathcal{P}(q, q')}\right).$$

By construction, $d\pi(q) = f(q)\, dq$ is an invariant measure [215].

The key point in all Metropolis-Hastings schemes is to find an efficient proposal function. In particular, there is always a trade-off between the acceptance and the decorrelation rate of the Markov chain. Indeed, if the acceptance rate is low, the obtained sample is degenerate, and not statistically confident. On the other hand, to increase the acceptance rate, more correlated iterations can be used. In this case the method is more likely to remain trapped in local minima, and the numerical ergodicity rate may be slow.

Metropolized independence sampler

We assume that the potential energy V is continuous. Considering an everywhere positive probability density g, let us set $\mathcal{P}(q, q') = g(q')$ and $w(q) = \dfrac{f(q)}{g(q)}$. This version of the Metropolis-

Hastings is called the Metropolized independence sampler (MIS). The algorithm we will use is therefore as follows:

METROPOLIZED INDEPENDENCE SAMPLING

Algorithm 3.2. Consider an initial point q^0. For $n \geq 1$,

(1) generate a point $\tilde{q}$ in $\mathcal{M}$ from the density g;

(2) generate a random number $U^n \sim \mathcal{U}[0,1]$;

(3) if $U^n \leq \min\left\{1, \dfrac{w(\tilde{q})}{w(q^n)}\right\}$, set $q^{n+1} = \tilde{q}$, otherwise, set $q^{n+1} = q^n$;

(4) replace n by $n+1$ and go back to step (1).

Convergence of the average along one sample path

Let us now recall some convergence results for Markov chains, which, applied to the specific cases of the Metropolized independence sampling, will provide convergence results. Let us denote by A some observable on the position space and by $(q^n)_{n\in\mathbb{N}}$ one realization of the MIS Markov chain starting from a given q^0. The question under examination is that of the convergence of the empirical mean $\frac{1}{N}S_N(A)$ toward $\int_{\mathcal{M}} A(q)\,d\pi(q)$ where π is the canonical measure defined by (3.6) and $S_N(A)$ is defined by (3.10).

First, π is an invariant measure due to general results on Metropolis-Hastings algorithms [215]. Therefore, condition (3.7) is satisfied. Condition (3.8) is also trivially satisfied whenever the support of f is a subset of the support of g. This is the case here since we have chosen a function g whose support is the whole position space $\mathcal{M}$.

Since conditions (3.7) and (3.8) are satisfied, a Law of Large Numbers (LLN) holds for almost all starting points, and Question 1 can therefore be answered positively. Indeed, recall the following theorem:

Theorem 3.1 ([240, Theorem 17.1.7]). *Suppose conditions (3.7) and (3.8) are satisfied. Then, for any measurable function $A \in L^1(\pi)$,*

$$\lim_{N\to\infty} \frac{1}{N} S_N(A) = \int_{\mathcal{M}} A d\pi \quad \text{a.s.}$$

for almost all starting points $q^0 \in \mathcal{M}$, where $S_N(A)$ is defined by (3.10).

To obtain a convergence rate on $S_N(A)$, an additional condition is needed, such as:

$$\begin{gathered}\text{There exist two measurable functions } L \geq \min\{1, A\} \text{ and } W \geq 0, \text{ a real number } b \\ \text{and a petite set } C \text{ such that} \\ \Delta W(q) \leq -L(q) + b\mathbf{1}_C(q), \quad \pi(W^2) < +\infty,\end{gathered} \tag{3.11}$$

where A is the observable under consideration and $\Delta W(q)$ is defined by

$$\forall q \in \mathcal{M},\ \Delta W(q) = (PW)(q) - W(q) = \int_{\mathcal{M}} P(q, dy)W(y) - W(q). \tag{3.12}$$

The definition of petite sets can be found in [240]. Let us make the following remark, which will be very useful:

Remark 3.1. *Under some regularity conditions that will always be met here (including the fact that the chain is weak Feller [240, Chap. 6]), all compact subsets of $\mathcal{M}$ are petite sets and the Markov chain is Doeblin [89]. As a consequence, when the state space $\mathcal{M}$ is compact, the condition (3.11) holds true (choose $C = \mathcal{M}$, W and L arbitrary smooth functions and take b large enough).*

Condition (3.11) allows one to obtain a Central Limit Theorem (CLT). For a given measurable function A such that $\pi(|A|) < +\infty$, let us formally define the function $\hat{A}$ by the following Poisson equation:

$$-\Delta \hat{A} = A - \pi(A), \tag{3.13}$$

where Δ is defined as in (3.12). It is not clear in general whether $\hat{A}$ is well-defined. This turns out to be the case when condition (3.11) is satisfied, and allows to state a CLT:

Theorem 3.2 ([240, Theorem 17.5.3]). *Assume conditions (3.7), (3.8) and (3.11) hold true, and let A be a function such that $|A| \leq L$. Let $S_N(A)$ be defined by (3.10). There exists a function $\hat{A}$ which satisfies (3.13), and the constant $\gamma_A^2 := \pi(\hat{A}^2 - (P\hat{A})^2)$ is well-defined, non-negative and finite. If $\gamma_A^2 > 0$, then, defining $\bar{A} = A - \pi(A)$,*

$$(N\gamma_A^2)^{-1/2} S_N(\bar{A}) \underset{N\to\infty}{\to} \mathcal{N}(0,1),$$

this convergence being in law.

Since conditions (3.7), (3.8) and (3.11) are satisfied for the MIS chain, Question 2 can be answered positively for almost all starting points q^0.

Convergence of the densities

To handle convergence of densities, it is necessary to introduce the total variation norm for a signed Borel measure ν, defined as

$$||\nu|| = \sup_{h \text{ measurable, } |h|\leq 1} |\nu(h)| = \sup_{A\in\mathcal{B}(\mathcal{M})} \nu(A) - \inf_{A\in\mathcal{B}(\mathcal{M})} \nu(A). \tag{3.14}$$

Notice that convergence in total variation implies weak convergence.

Definition 3.1. *A chain on $\mathcal{M}$ is* ergodic *when*

$$\forall q \in \mathcal{M}, \quad \lim_{n\to\infty} ||P^n(q,\cdot) - \pi|| = 0$$

where π is the invariant measure and P^n is the n-step probability transition kernel.

Recall the following theorem:

Theorem 3.3 ([240, Theorem 13.3.4]). *If conditions (3.7) and (3.8) hold true, then*

$$||P^n(q,\cdot) - \pi|| \to 0 \quad \text{as } n \to \infty$$

for π-almost all starting points q.

The convergence in total variation norm implies convergence of the expectations only for bounded observables A. It is therefore not sufficient in practice. Fortunately, the ergodicity results can be strengthened in a straightforward way. For a given measurable non-negative function $W \geq 1$, let us define the W-total variation norm for a signed Borel measure μ as

$$||\mu||_W = \sup_{h \text{ measurable, } |h|\leq W} |\mu(h)|. \tag{3.15}$$

Then Theorem 3.3 can be readily extended to π-integrable functions A.

Theorem 3.4 ([240, Theorem 14.0.1]). *Suppose that $A \geq 1$ is measurable and $\pi(|A|) < +\infty$. If conditions (3.7) and (3.8) hold true, then for π-almost all $q \in \mathcal{M}$,*

$$||P^n(q,\cdot) - \pi||_{|A|} \to 0 \quad \text{as } n \to \infty.$$

Since conditions (3.7) and (3.8) are satisfied, the MIS Markov chain is ergodic and Theorems 3.3 and 3.4 hold true. This answers Question 3.

Under an assumption which is reminiscent of the Rejection method setting, a simple uniform convergence rate (independent of the starting point q^0) can be obtained:

Theorem 3.5 ([237, Theorem 2.1]). *If the probability density g used in the metropolized independence sampling scheme is such that*

$$\exists c,\ \forall q \in \mathcal{M}, \quad f(q) \leq c g(q),$$

then the scheme is geometrically ergodic with a uniform bound. In this case, for all $q^0 \in \mathcal{M}$,

$$||P^n(q^0,\cdot) - \pi|| \leq (1 - c^{-1})^n.$$

This theorem gives an answer to Question 4. Note that in the particular case when $c = 1$ (that is when $f = g$ since both functions are probability densities), the convergence is already achieved for $n = 1$. This is actually clear since in this case the MIS scheme samples from the true density!

3.1.4 Importance sampling

Importance sampling is a well-known general stochastic integration method. The underlying idea is to recast the integral $\mathbb{E}_\pi(A) = \int_{\mathcal{M}} A(q)\, f(q)\, dq$ as

$$\mathbb{E}_\pi(A) = \int_{\mathcal{M}} \left(A(q)\, \frac{f(q)}{g(q)} \right) g(q)\, dq$$

and to approximate the latter integral through a random sample $(q^n)_{0 \leq n \leq N-1}$ drawn according to the density g (see e.g. [215, Section 2]).

The choice of the trial function g is crucial for the overall efficiency of the method. It should be a good approximation of f or, better, of $f(q)A(q)$. Since f is typically of exponential or Gaussian form, and A is most often bounded by a polynomial, f is usually the most important term in the product $f(q)A(q)$ as far as sampling issues are concerned. Besides, in applications, it is often the case that several integrals have to be computed, with different functions A. So g is often looked for as a good approximation of f.

Let us note that, for the computation of static quantities, the importance sampling method based on a density g outperforms the Rejection method based on the same density g [64].

3.2 Stochastically perturbed Molecular Dynamics methods

We first present in Section 3.2.1 the general framework of deterministic microcanonical (NVE) MD. In Section 3.2.2, we describe the Hybrid Monte Carlo (HMC) method, from both the theoretical and the numerical viewpoints, and give some new convergence results (see Theorems 3.7, 3.9, 3.10). We then present the biased Random-Walk (BRW) in Section 3.2.3, and the Langevin dynamics in Section 3.2.4.

We assume in the sequel that $T^*\mathcal{M}$ is globally diffeomorphic to $\mathcal{M} \times \mathbb{R}^{3N}$, and actually identify the two sets for simplicity. We also assume that $\mathcal{M}$ is globally diffeomorphic to $\mathbb{R}^{3N}$ in Sections 3.2.3

and 3.2.4, and identify the two sets as well. Straightforward modifications allow to handle the other cases (such as systems with periodic boundary conditions or isolated systems parametrized by rigid-body motions and internal coordinates).

3.2.1 General framework for NVE Molecular Dynamics

The equations of motion

$$\begin{cases} \dfrac{dq(t)}{dt} = \dfrac{\partial H}{\partial p}(q(t),p(t)) = M^{-1}p(t), \\ \dfrac{dp(t)}{dt} = -\dfrac{\partial H}{\partial q}(q(t),p(t)) = -\nabla V(q(t)), \end{cases} \tag{3.16}$$

associated with the Hamiltonian (3.4) can be numerically integrated e.g. by the celebrated velocity-Verlet algorithm [360]

$$\begin{cases} p^{n+1/2} = p^n - \dfrac{\Delta t}{2}\nabla V(q^n), \\ q^{n+1} = q^n + \Delta t\ M^{-1}p^{n+1/2}, \\ p^{n+1} = p^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(q^{n+1}), \end{cases} \tag{3.17}$$

where Δt is the time step. The velocity-Verlet scheme is an *explicit* integrator: recall that in Statistical Physics one often considers systems with a large number of particles, making implicit algorithms untractable. The numerical flow associated with the velocity-Verlet algorithm shares two qualitative properties with the exact flow of (3.16): it is *time reversible* and *symplectic*, which are very important properties as far as the long time numerical integration of Hamiltonian dynamics is concerned (see [146, Chap. VIII and IX] and [205]). This algorithm also asks for a unique evaluation of the forces $F = -\nabla V$ per time step. For all these reasons, it is the most commonly used algorithm in molecular dynamics.

The dynamics (3.16) cannot be used to generate points according to the canonical measure, because the energy (3.4) is preserved by the flow. Hence, the trajectory of the system remains on the submanifold of constant energy

$$T^*\mathcal{M}(E_0) = \{(q,p) \in T^*\mathcal{M}; H(q,p) = E_0\}$$

where $E_0 = H(q_0,p_0)$ is the energy of the initial data. Under some assumptions, the dynamics (3.16) can be used to compute microcanonical (NVE) ensemble averages, that is, averages over $T^*\mathcal{M}(E_0)$. The numerical analysis of this method (in the very simple case of completely integrable systems) can be read in [48, 49, 203]. To generate points according to the canonical measure, there is a need for stochastic perturbations to ensure that different energy levels will be explored, and eventually all of them. These considerations straightforwardly extend to the numerical case since symplectic methods such as (3.17) almost preserve the energy over extremely long times [146, Chap. IX].

3.2.2 Hybrid Monte Carlo

Presentation of the method

The Hybrid Monte Carlo method allows one to generate points in the position space distributed according to the canonical measure (3.6). It aims at combining the advantages of molecular dynamics (that approximates the physical dynamics of the system) and of Monte Carlo methods (that explore the position space more globally). It is in fact a Metropolis-Hastings algorithm, in which proposals are constructed using the NVE Hamiltonian flow of the system. This method has

been first introduced by Duane et al. in [88] and partially analyzed from a mathematical viewpoint by Schütte in [301]. This method can be seen as a generalization of the Andersen thermostat method [7]. It has been used in [302, 303] to identify the metastable conformations of some biological systems.

In the standard HMC setting, the sequence of generated positions forms a Markov chain of order one defined as follows:

HYBRID MONTE CARLO

Algorithm 3.3. Consider an initial configuration $q^0 \in \mathcal{M}$ and $\tau > 0$. For $n \geq 0$,

(1) generate momenta p^n according to the canonical distribution (3.5) and compute the energy $E^n = H(q^n, p^n)$ of the configuration (q^n, p^n);

(2) compute $\Phi_\tau(q^n, p^n) = (p^{n,\tau}, q^{n,\tau})$, that is, integrate the NVE equations of motion (3.16) on the time interval $[0, \tau]$ starting from the initial data (q^n, p^n);

(3) compute the energy $E^{n,\tau} = H(q^{n,\tau}, p^{n,\tau})$ of the new phase-space configuration. Accept the proposal $q^{n,\tau}$ with probability

$$\alpha^n = \min\left(1, \mathrm{e}^{-\beta(E^{n,\tau} - E^n)}\right);$$

more precisely, generate a random number $U^n \sim \mathcal{U}[0,1]$, and set $q^{n+1} = q^{n,\tau}$ if $U^n \leq \alpha^n$ and $q^{n+1} = q^n$ otherwise;

(4) replace n by $n+1$ and go back to step (1).

Let us emphasize that the proposal $q^{n,\tau}$ would always be accepted at step (3) if the NVE equations of motion, that are energy conserving, were integrated exactly. In practice, the time-step Δt used in the numerical integrator (3.17) can be chosen larger than in standard applications of MD since the dynamics of the system used to generate proposals is not constrained to accurately reproduce the physical dynamics of the system. On the other hand, it should not be too large; otherwise, the rejection rate would be large and the efficiency of the method would be low.

Let us notice that in the standard HMC method, only the end points of the MD trajectories are part of the sample. It is not completely clear whether taking into account the intermediate points of the generated MD trajectories in the sample would bias the sampling, e.g. if the final point is rejected, should these intermediate points be kept? See [256] for some work in this direction.

Let us also mention that there exist several refinements of the standard HMC scheme. In order to improve the acceptance rate, one could use a criterion based on a shadow Hamiltonian to accept or reject the new point [150, 184]. The idea is that this shadow Hamiltonian is preserved more accurately than the Hamiltonian (3.4) by the numerical trajectory. The bias introduced by this modification is corrected by a convenient reweighting, in the spirit of importance sampling. Another improvement consists in generating, after each NVE trajectory of length τ, some new momenta which are correlated with the previous ones [173, 191]. Of course, both approaches can be combined [2].

Convergence of the average along one realization

As above, let us denote by A some observable on the position space and by $(q^n)_{n\in\mathbb{N}}$ one realization of the HMC Markov chain starting from a given q^0. Let Π_1 be the first coordinate field of the phase-space: $\Pi_1(q,p) = q$.

Convergence results for the HMC scheme have been published by Schütte in [301]. In this proof, the NVE Hamiltonian flow is assumed to satisfy two conditions:

(A) a *mixing* condition, which reads as follows (see [301, Assumption 4.27]): for every pair of open subsets $B, C \subset \mathcal{M}$, there exists $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0, \quad \int_B T^n \mathbf{1}_C(q) f(q)\, dq > 0,$$

where f is given by (3.6) and the function Tu is defined for any function $u : \mathcal{M} \to \mathbb{R}$ by

$$Tu(q) = \int_{\mathbb{R}^{3N}} u\left(\Pi_1 \Phi_{-\tau}(q,p)\right) \mathcal{P}(p)\, dp, \tag{3.18}$$

where Φ_τ is the Hamiltonian flow. This condition amounts to a certain accessibility of the whole position space when starting from any point;

(B) a so-called *momentum invertibility of the flow* condition (see [301, Definition 4.1]). The flow Φ_τ is called momentum-invertible if the two following conditions hold true:

(i) for almost every $q \in \mathcal{M}$, there is an open set $M(q) \subset \mathbb{R}^{3N}$ such that the function $y_q : p \mapsto \Pi_1 \Phi_{-\tau}(q,p)$ is locally invertible in $M(q)$, that is, $\det \nabla_p y_q \neq 0$ for $p \in M(q)$.

(ii) there is an $\eta > 0$ such that

$$\operatorname*{ess-inf}_{q \in \mathcal{M}} \int_{M(q)} \mathcal{P}(p)\, dp = \eta.$$

This condition states that the transition probabilities are bounded from below in some sense.

The following convergence result is given in [301]:

Theorem 3.6 ([301, Lemma 4.31 and Theorem A.24]). *Under the assumptions (A) and (B) recalled above, for any measurable function $A \in L^1(\pi)$, it follows*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} A(q^n) = \int_{\mathcal{M}} A\, d\pi \quad \text{a.s.} \tag{3.19}$$

for almost all starting points $q^0 \in \mathcal{M}$, where $(q^n)_{n \in \mathbb{N}}$ is the sequence of points generated by the HMC Algorithm 3.3 where, at step (2), the NVE equations of motion (3.16) are exactly integrated.

Note that ergodicity results have also been proved [301, Corollary 4.33], as well as convergence results on the numerical flow [301, page 96] (in this latter case, $(q^n)_{n \in \mathbb{N}}$ in (3.19) is the sequence of points generated by the HMC Algorithm 3.3 where the NVE equations of motion (3.16) are now numerically integrated).

The conditions (A) and (B) recalled above are difficult to check in practice, and furthermore, it is not clear whether they are necessary. We present here a new convergence result, that does not require these assumptions.

Let us first consider the case when the NVE equations of motion are integrated exactly. The transition kernel P of the HMC Markov chain is defined by

$$\forall (q, B) \in \mathcal{M} \times \mathcal{B}(\mathcal{M}), \qquad P(q, B) = \int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(q,p) \in B\}} \mathcal{P}(p) dp, \tag{3.20}$$

where the density $\mathcal{P}$ is the canonical distribution on the momentum space given by (3.5).

As the phase-space canonical measure $\mu = \pi \otimes \kappa$ is an invariant measure for Φ_τ, it is clear that the position-space canonical measure π is an invariant measure for the HMC Markov chain (see e.g. [215, Section 9.3] for details). Therefore, condition (3.7) holds true.

We now consider the accessibility condition (3.8). This condition is not satisfied in general, for any potential energy. Consider for example a one-dimensional particle ($\mathcal{M} = \mathbb{R}$) of mass $m = 1$ subjected to the potential energy $V(q) = \frac{1}{2}q^2$. Then the solution $q(t)$ starting from q^0 with momentum p^0 is given by

$$q(t) = q^0 \cos(t) + p^0 \sin(t).$$

As already noticed by Mackenzie in [221], taking $\tau = 2\pi$ leads to $q(\tau) = q^0$ whatever the choice of p^0. The condition (3.8) is therefore clearly not satisfied, and the Markov chain is not ergodic. Of course this spurious effect only arises for special choices of τ. It is also linked to the fact that the period of the trajectory of the harmonic oscillator does not depend on the initial momentum.

To prove the accessibility condition (3.8), a first way is to make the additional assumption that the potential energy is bounded from above. We acknowledge that this assumption is often not satisfied in practice. Nevertheless, for some potential energies that do not satisfy this assumption, it is still possible to prove an accessibility condition by some explicit constructions, specific to the system at hand (especially in the case of a singular central potential energy, see below). We will also consider in Section 3.2.2 another possibility, based on random integration times τ, that can be used for a larger class of potentials.

We now turn to proving the accessibility condition (3.8) under the assumption that V is bounded from above. This is the result of the following Lemmas.

Lemma 3.1 (HMC accessibility - exact flow). *Let $\tau > 0$. Assume that V is in $C^1(\mathcal{M})$ and is bounded from above. Then for any $q, q' \in \mathcal{M}$ and any neighborhood $\mathcal{V}'$ of q', there holds*

$$P(q, \mathcal{V}') > 0.$$

Proof. The proof is based on the least action principle (LAP). Let us denote by

$$S(\phi) = \int_0^\tau \left(\frac{1}{2}\dot{\phi}^T(t) M \dot{\phi}(t) - V(\phi(t)) \right) dt$$

the action associated with the path $\phi \in \mathcal{H} = \{\phi \in H^1([0,\tau], \mathcal{M}) \mid \phi(0) = q,\ \phi(\tau) = q'\}$. Since V is bounded from above, there exists E_0 such that $V(q) \leq E_0$ for all $q \in \mathcal{M}$. Thus, S is bounded from below:

$$S(\phi) \geq -\int_0^\tau V(\phi(t))\, dt \geq -E_0 \tau.$$

Therefore, there exists a minimizing sequence $(\phi_n)_{n\in\mathbb{N}} \in \mathcal{H}$ such that $S(\phi_n) \to \inf_{\phi\in\mathcal{H}} S(\phi) = s > -\infty$. Without restriction, it can be assumed that $s \leq S(\phi_n) \leq s + 1$ for all $n \in \mathbb{N}$. Thus,

$$\int_0^\tau \dot{\phi}_n^T(t) M \dot{\phi}_n(t)\, dt = 2S(\phi_n) + 2\int_0^\tau V(\phi_n(t))\, dt \leq 2S(\phi_n) + 2\tau E_0 \leq 2(s+1) + 2\tau E_0.$$

Therefore, $(\dot{\phi}_n)_{n\in\mathbb{N}}$ is bounded in $L^2([0,\tau], \mathcal{M})$. The sequence $(\phi_n)_{n\in\mathbb{N}}$ is then bounded in the space $H^1([0,\tau], \mathcal{M})$. Let $\phi \in H^1([0,\tau], \mathcal{M})$ such that (up to extraction) $\phi_n \rightharpoonup \phi$ in $H^1([0,\tau], \mathcal{M})$-weak and $\phi_n \to \phi$ almost everywhere. Since $\mathcal{H}$ is convex and closed in $H^1([0,\tau], \mathcal{M})$, the limit ϕ is actually in $\mathcal{H}$. Besides, it is easy to check that $\liminf_{n\to\infty} S(\phi_n) \geq S(\phi)$ (by lower semi-continuity on the kinetic energy and Fatou lemma on the potential energy), and this gives immediately

$$\inf_{\psi\in\mathcal{H}} S(\psi) = \min_{\psi\in\mathcal{H}} S(\psi) = S(\phi).$$

Thus ϕ minimizes S on $\mathcal{H}$. Therefore, the equation

$$M\,\ddot{\phi} = -\nabla V(\phi) \tag{3.21}$$

holds true on $(0,\tau)$ in the distributions sense. By standard regularity results, $\phi \in C^2([0,\tau],\mathcal{M})$ and (3.21) holds true in the sense of continuous functions. Hence the function ϕ is simply the solution of the Hamiltonian dynamics with $\phi(0) = q$, $\phi(\tau) = q'$ and initial velocity $\dot{\phi}(0)$.

Consider eventually a neighborhood $\mathcal{V}'$ of q'. Then $P(q,\mathcal{V}') > 0$ is a straightforward consequence of the continuity of the solutions of (3.21) with respect to the initial velocity $\dot{\phi}(0)$. □

Lemma 3.1 gives accessibility from any point to any *open* set. It is therefore not enough for condition (3.8) to hold true since it requires accessibility from one point to any arbitrary Borel set of positive Lebesgue measure. This asks for some regularity of the transition kernel, and in fact, some regularity of the dynamics, inferred from stronger assumptions on the potential energy V. More precisely, we have the following lemma:

Lemma 3.2 (HMC irreducibility - exact flow). *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is a globally Lipschitz function. Then the transition kernel of the HMC Markov chain satisfies*

$$\forall q \in \mathcal{M},\ \forall B \in \mathcal{B}(\mathcal{M}),\quad \mu^{\mathrm{Leb}}(B) > 0 \Rightarrow P(q,B) > 0.$$

Proof. Consider $B \in \mathcal{B}(\mathcal{M})$ such that $\mu^{\mathrm{Leb}}(B) > 0$, and $q \in \mathcal{M}$. We want to show that $P(q,B) > 0$ for P defined by (3.20). For the sake of simplicity, we assume here that all particle masses are equal to 1.

The proof is based on volume conservation in the phase space: any Borel set of final positions of strictly positive measure can be reached from q and a set of momenta of strictly positive measure. Denote $I_B(q) = \{p \in \mathbb{R}^{3N} \mid \Pi_1\Phi_\tau(q,p) \in B\}$, and consider the function $\theta : I_B(q) \mapsto B$ such that $\theta(p) = \Pi_1\Phi_\tau(q,p)$. This function is surjective according to the proof of the accessibility Lemma 3.1, so that $\theta(I_B(q)) = B$. Moreover, $P(q,B) = \displaystyle\int_{I_B(q)} \mathcal{P}(p)\,dp$. Therefore, since $\mathcal{P}$ is positive and continuous, it is enough to show that $\mu^{\mathrm{Leb}}(I_B(q)) > 0$ in order to get $P(q,B) > 0$.

We proceed by contradiction. Suppose $\mu^{\mathrm{Leb}}(I_B(q)) = 0$. We first note that θ is Lipschitz (of constant $\mathrm{Lip}(\theta)$) since ∇V is continuous and globally Lipschitz by assumption, and $\tau > 0$ is fixed. Indeed, denote C the Lipschitz constant of ∇V and note that a solution of the equations of motion can be written as

$$q(\tau) = q + p\tau - \int_0^\tau (\tau - s)\nabla V(q(s))\,ds.$$

For two different initial momenta p_1 and p_2, we have

$$|q_1(t) - q_2(t)| \le |p_1 - p_2|t + C\int_0^t (t-s)|q_1(s) - q_2(s)|ds.$$

By Gronwall lemma, there exists $c_\tau < +\infty$ such that

$$|q_1(\tau) - q_2(\tau)| \le c_\tau |p_1 - p_2|,$$

hence θ is Lipschitz.

Since the Lebesgue measure and the Hausdorff measure $\mathcal{H}^{3N}$ agree on $\mathbb{R}^{3N}$ (see [101, Section 2.2, Theorem 2]), and since the behavior of the Hausdorff measure under Lipschitz mappings is known [101, Section 2.4, Theorem 1], we obtain

$$\mu^{\mathrm{Leb}}(B) = \mu^{\mathrm{Leb}}(\theta(I_B(q))) = \mathcal{H}^{3N}(\theta(I_B(q))) \le \mathrm{Lip}(\theta)^{3N}\mathcal{H}^{3N}(I_B(q)) = \mathrm{Lip}(\theta)^{3N}\mu^{\mathrm{Leb}}(I_B(q)) = 0.$$

This gives $\mu^{\mathrm{Leb}}(B) = 0$, in contradiction with the assumption $\mu^{\mathrm{Leb}}(B) > 0$. □

Since conditions (3.7) and (3.8) are satisfied, a Law of Large Numbers (LLN) holds true for almost all starting points (see Theorem 3.1). We can therefore answer positively to Question 1:

Theorem 3.7. *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is a globally Lipschitz function. Let $(q^n)_{n\in\mathbb{N}}$ be the sequence of points generated by the HMC Algorithm 3.3 where, at step (2), the NVE equations of motion (3.16) are exactly integrated. Then*

$$\frac{1}{N}\sum_{n=0}^{N-1} A(q^n) \to \int_{\mathcal{M}} A(q)\,d\pi \quad \text{a.s.}$$

for almost all starting points $q^0 \in \mathcal{M}$.

Convergence of the HMC scheme has been established above for smooth potentials, possibly under certain boundedness assumptions on the potential V or its derivatives. However, in many applications, non-globally smooth potentials are used. Central potentials, such as the Lennard-Jones or the Coulomb potential, are some famous examples of singular potentials commonly considered in biology or physics. We present here some results concerning the convergence of the HMC scheme for a single particle in a central potential decaying sufficiently fast at infinity (such as $|q|^{-\alpha}$ for α large enough). Only the accessibility properties of the chain are stated explicitely, the rest of the proof following the same lines as for the usual HMC scheme.

In view of the reversibility of the NVE equations of motion, to show that any point q_2 can be reached in two steps from a point q_1, is equivalent to showing that the end points of the trajectories starting from q_1 and q_2 coincide. This is the following

Proposition 3.1 (HMC accessibility for one particle in a decreasing central potential). *Consider a central potential $V(q) = V(|q|) \in \mathrm{C}^1(\mathbb{R}^3 \setminus \{0\})$ such that $q \cdot V'(q) \le 0$, ∇V is lipschitz on $\mathbb{R}^3 \setminus B_a(0)$ for all $a > 0$ with a constant C_a such that $\lim_{a\to\infty} C_a = 0$, and $|\nabla V|$ is bounded on $\mathbb{R}^3 \setminus B_a(0)$ for all $a > 0$. Consider $q_1, q_2 \in \mathbb{R}^3 \setminus \{0\}$ such that q_1, the singularity 0 and q_2 are not aligned in this order (there is no $\lambda > 0$ such that $q_1^0 = -\lambda q_2^0$). Then there exist p_1, p_2 such that the solutions of the equations of motion*

$$\ddot{z} = -\nabla V(z)$$

starting respectively from q_1, q_2 with momenta p_1, p_2 coincide at the time τ.

The proof is based on an explicit two-step construction. If q_1, 0 and q_2 are aligned in this order, then an additional configuration q_3 not aligned with the previous ones should be considered. Hence, one can go from q_1 to q_2 by four trajectories of time length τ. These results can be extended to more general potentials such as the Lennard-Jones potential in a simple way.

Proof. We consider two points q_1^0, q_2^0 and the corresponding initial momenta p_1^0, p_2^0. The two particles are assumed to be of identical masses 1, the general result following after straightforward modifications. Then,

$$q_1(t) = q_1^0 + p_1^0 t - \int_0^t (t-s)\nabla V(q_1(s))\,ds,$$

and, setting $p_2^0 = p_1^0 - \frac{q_2^0 - q_1^0}{\tau} + p$ (using a "small" parameter p),

$$q_{2,p}(t) = q_2^0 + \left(p_1^0 - \frac{q_2^0 - q_1^0}{\tau}\right) t + pt - \int_0^t (t-s)\nabla V(q_{2,p}(s))\,ds. \tag{3.22}$$

We look for p such that $q_{2,p}(\tau) = q_1(\tau)$. This condition can be rewritten as

$$p = \frac{1}{\tau}\int_0^\tau (\tau - s)[\nabla V(q_{2,p}(s)) - \nabla V(q_1(s))]\,ds = F(p).$$

Under this form, we recognize a fixed-point equation, trivially verified by $p = 0$ in the case $\nabla V = 0$. The idea is then solve this equation for ∇V small. This can be done if the trajectories move away

from the singularity in 0. To this end, the momentum p_1^0 has to be taken large enough, and p has to be small compared to p_1^0.

We now formalize these heuristic considerations. Notice first that the initial momentum p_1^0 can be chosen so that the particle moves out from 0. Indeed, using polar coordinates (r, θ) for the particle position $q \in \mathbb{R}^3$,

$$\partial_{tt}(r^2) = \partial_{tt}(x \cdot x) = 2(|\dot{q}|^2 - q \cdot \nabla V(q)) \geq -2rV'(r).$$

By integration,

$$\partial_t(r^2)(t) - \partial_t(r^2)(0) \geq -\int_0^t 2r(t)V'(r(t))dt \geq 0. \tag{3.23}$$

So, if the initial conditions are such that $\partial_t(r^2)(0) > 0$, the distance r of a particle to the origin is increasing. Let us set

$$M = \sup_{|q| \geq \min(|q_1^0|, |q_2^0|)} |\nabla V(q)|, \quad K = M\tau. \tag{3.24}$$

Since

$$\begin{aligned} \partial_t(q_1 \cdot q_1)(0) &= 2q_1^0 \cdot p_1^0, \\ \partial_t(q_{2,p} \cdot q_{2,p})(0) &= 2q_2^0 \cdot \left(p_1^0 + p - \frac{q_2^0 - q_1^0}{\tau}\right), \end{aligned}$$

and considering p and p_1^0 such that

$$|p| \leq K, \quad q_2^0 \cdot p_1^0 \geq q_2^0 \cdot \frac{q_2^0 - q_1^0}{\tau} + K|q_2^0|, \quad q_1^0 \cdot p_1^0 \geq 0, \tag{3.25}$$

it follows $\partial_t(q_1 \cdot q_1)(0) \geq 0$ and $\partial_t(q_{2,p} \cdot q_{2,p})(0) \geq 0$. Let us note that, because q_1^0, the singularity 0 and q_2^0 are not aligned, such p_1^0 exist. Therefore,

$$\forall t \geq 0, \quad \forall |p| \leq K, \quad |q_{2,p}(t)| \geq |q_2^0|, \quad |q_1(t)| \geq |q_1^0|. \tag{3.26}$$

Next, we show that there exists t_* small enough such that $p \mapsto q_{2,p}(t)$ is Lipschitz with uniform bound on $[0, t_*]$. Indeed, from the expression (3.22), and since ∇V is lipschitz of constant $C = C_{|q_2^0|}$ on $\mathbb{R}^3 \setminus B_{|q_2^0|}(0)$, we obtain

$$|q_{2,p}(t) - q_{2,p'}(t)| \leq |p - p'|t + C\int_0^t (t-s)|q_{2,p}(t) - q_{2,p'}(t)|\, ds.$$

This Gronwall inequality implies

$$|q_{2,p}(t) - q_{2,p'}(t)| \leq |p - p'| \int_0^t s \exp(C(t-s))ds.$$

Taking $t_* \leq \tau$ small enough, we get for all $0 \leq t \leq t_*$,

$$|q_{2,p}(t) - q_{2,p'}(t)| \leq \frac{1}{4C}|p - p'|. \tag{3.27}$$

This time t_* is now fixed in the remainder of this proof.

Thus, p_1^0 being fixed, the distance between two trajectories can be controlled for small times. For larger times, we use the fact that we can go arbitrary far from the origin by an appropriate choice of the initial momentum. Indeed,

$$|q_{2,p}(t)| \geq \left| q_2^0 + \left(p_1^0 + p - \frac{q_2^0 - q_1^0}{\tau} \right) t \right| - \tau^2 \sup_{|q| \geq |q_2^0|} |\nabla V(q)|. \tag{3.28}$$

Let $\epsilon > 0$. Since ∇V is lipschitz on $\mathbb{R}^3 \setminus B_a(0)$ with a constant C_a such that $\lim_{a\to\infty} C_a = 0$, there exists $R(\epsilon)$ such that $C_{R(\epsilon)} < \epsilon$. If view of (3.28), there exists an momentum p_1^0 large enough satisfying (3.25) such that

$$\forall p, |p| \leq K, \quad \forall t \geq t_*, \quad |q_{2,p}(t)| \geq R(\epsilon). \tag{3.29}$$

Considering two momenta $|p|, |p'| \leq K$, a Gronwall inequality can again be obtained. There exists a constant C_τ (that does not depend on $\epsilon \leq 1$) such that

$$\forall t,\ t_* \leq t \leq \tau, \quad |q_{2,p}(t) - q_{2,p'}(t)| \leq C_\tau |p - p'|. \tag{3.30}$$

The proof can now be concluded. Recall that we look for a fixed-point of the function

$$F(p) = \frac{1}{\tau} \int_0^\tau (\tau - s)[\nabla V(q_{2,p}(s)) - \nabla V(q_1(s))] ds.$$

The mapping F maps $B_K = \{|p| \leq K\}$ into itself when p_1^0 satisfies (3.25). Indeed, the bound (3.26) is verified in this case, so that (3.24) implies

$$|F(p)| \leq \frac{1}{\tau} \int_0^\tau (\tau - s) 2M\, ds = M\tau = K.$$

Picard theorem can then be applied provided F is contractive. Choosing momenta such that (3.25) holds true and such that $\epsilon < \min\{1, \frac{1}{4C_\tau \tau}\}$,

$$|F(p) - F(p')| \leq C \int_0^{t_*} |q_{2,p}(s) - q_{2,p'}(s)|\, ds + \int_{t_*}^\tau |\nabla V(q_{2,p}(s)) - \nabla V(q_{2,p'}(s))|\, ds.$$

Using (3.27) for the first term and, (3.29), the fact that ∇V is lipschitz on $\mathbb{R}^3 \setminus B_{R(\epsilon)}(0)$ with a constant $C_{R(\epsilon)} < \epsilon$, and (3.30) for the second term, there holds

$$\forall |p|, |p'| \leq K, \quad |F(p) - F(p')| \leq \frac{1}{2} |p - p'|.$$

The function F is then contractive on the ball $\{|p| \leq K\}$. There is therefore a fixed point $p = F(p)$ with $|p| \leq K$. □

Convergence of the densities

Since condition (3.7) is satisfied, and condition (3.8) holds true under the assumptions of Lemma 3.2 on the potential energy (V is C^1, bounded from above and ∇V is globally Lipschitz), the HMC Markov chain is ergodic (see Theorem 3.3). In particular,

$$\left|\left| P^n(q^0, \cdot) - \pi \right|\right| \to 0$$

for almost all starting points $q^0 \in \mathcal{M}$, where $||\cdot||$ denotes the total variation norm (3.14). We also get convergence in the $|A|$-total variation norm (3.15) provided $\pi(|A|) < +\infty$ and $|A| \geq 1$ (see Theorem 3.4). This answers Question 3.

Convergence rates

We have not been able to state more sophisticated convergence results (Central Limit Theorem, geometric ergodicity) in the general HMC framework since they require stronger results on the Markov chain such as a drift condition (3.11) or a Lyapunov condition such as

$$\begin{gathered}\text{There exist a measurable function } W \geq 1, \text{ real numbers } c > 0 \text{ and } b, \\ \text{and a petite set } C \text{ such that} \\ \forall q \in \mathcal{M},\ \Delta W(q) \leq -cW(q) + b\mathbf{1}_C,\end{gathered} \tag{3.31}$$

where $\Delta W(q)$ is defined by (3.12). Let us however make the following remark:

Remark 3.2. *Under some regularity conditions that will always be met here (including the fact that the chain is weak Feller [240, Chap. 6]), and when $\mathcal{M}$ is compact, condition (3.31) is straightforwardly satisfied with the choice $C = \mathcal{M}$ (in view of Remark 3.1, $\mathcal{M}$ is a petite set and the Markov chain is Doeblin [89]) for any arbitrary smooth function W (taking b large enough).*

When the state space is compact, conditions (3.11) and (3.31) hold true (in view of Remarks 3.1 and 3.2). We thus obtain a positive answer to Question 2 (see Theorem 3.2). We also obtain a positive answer to Question 4, in view of the following theorem:

Theorem 3.8 ([240, Theorem 15.0.1]). *Assume conditions (3.7), (3.8) and (3.31) hold true. Then there exist $\rho < 1$ and $R < +\infty$ such that, for all q satisfying $W(q) < +\infty$,*

$$\|P^n(q,\cdot) - \pi\|_W \leq R\,W(q)\,\rho^n,$$

where P^n is the n-step probability transition kernel and $\|\cdot\|_W$ is the norm defined by (3.15).

Numerical implementation: Method and convergence results

It is standard to use the velocity-Verlet scheme (3.17) to integrate numerically the trajectories over times $\tau = k\Delta t$ for some integer k. Let us point out that the acceptance/rejection step (3) in Algorithm 3.3 ensures that the HMC Markov chain correctly samples the canonical measure π, so that no bias is introduced by the numerical discretization. The situation will be different for the Biased Random-Walk and the Langevin equation (see Sections 3.2.3 and 3.2.4). We denote by $P_{\Delta t}$ the transition kernel of the Markov chain using the velocity-Verlet integrator (3.17) with time-step Δt.

The theoretical proof of convergence for the numerical version of HMC follows the same lines as the proof of convergence for the exact version using the Hamiltonian flow. The only difference lies in the additional acceptance/rejection step which does not modify the structure of the chain (for it does not change the accessibility properties of the chain). We only precise here the changes that have to be considered for the accessibility Lemma.

Lemma 3.3 (HMC accessibility - numerical flow). *Let $\tau > 0$. Assume that V is in $C^1(\mathcal{M})$ and is bounded from above on $\mathcal{M}$, and consider the numerical discretization scheme (3.17). Then for any $q, q' \in \mathcal{M}$ and any neighborhood $\mathcal{V}'$ of q', there holds*

$$P_{\Delta t}(q, \mathcal{V}') > 0.$$

Proof. The proof of Lemma 3.1 is based on the minimization of the action S over some space $\mathcal{H}$. Here, we extend this proof to the discretized case using a convenient approximation of this variational problem. There are several ways to discretize the variational problem, leading to different numerical schemes. In particular, the velocity-Verlet algorithm can be derived by minimizing the discretized action [226]

$$S_{\Delta t}(\Phi) = \Delta t \sum_{i=0}^{k-1} \left[\frac{1}{2} \left(\frac{q^{i+1} - q^i}{\Delta t} \right)^2 - \frac{V(q^{i+1}) + V(q^i)}{2} \right], \tag{3.32}$$

where $\tau = k\Delta t$ (we again assumed here that all particle masses are equal to 1).

The minimization is performed on the sequences $\Phi = \{q^0, q^1, \ldots, q^k\}$ with the constraints $q^0 = q$ and $q^k = q'$. The quantity $S_{\Delta t}$ is still bounded from below for a potential energy bounded from above. Hence, there exists a minimizing sequence $(\Phi_n)_{n\in\mathbb{N}} = (\{q^{0,n}, q^{1,n}, \ldots, q^{k,n}\})_{n\in\mathbb{N}}$. Each difference $q^{i+1,n} - q^{i,n}$ is easily seen to be bounded, thus each component $q^{i,n}$ is in fact bounded. We can consider $\bar{\Phi} = (\bar{q}^0, \ldots, \bar{q}^k)$ such that, upon extraction, we have $q^{i,n} \to \bar{q}^i$ when $n \to \infty$ for each i. Moreover, $S(\bar{\Phi}) = \min_\Phi S(\Phi)$. The optimality conditions then read

$$\bar{q}^{i+1} = 2\bar{q}^i - \bar{q}^{i-1} - \Delta t^2 \nabla V(\bar{q}^i)$$

for $1 \le i \le k-1$. We recognize the Verlet scheme. As in addition $\bar{q}^0 = q$ and $\bar{q}^k = q'$, this shows that given two points q, q', there is a path connecting them using a numerical velocity-Verlet trajectory with initial velocity $\bar{p}^0 = \dfrac{\bar{q}^1 - \bar{q}^0}{\Delta t} + \dfrac{\Delta t}{2} \nabla V(\bar{q}^0)$. By continuity, for initial velocities close to $\bar{p}^0$, the endpoint of the resulting trajectory remains in a neighborhood of q'. □

We can now state a Law of Large Number theorem (see Theorem 3.1):

Theorem 3.9. *Assume that $V \in C^1(\mathcal{M})$ is bounded from above and ∇V is globally Lipschitz. Let $(q^n)_{n\in\mathbb{N}}$ be the sequence of points generated by the HMC Algorithm 3.3 where, at step (2), the NVE equations of motion (3.16) are numerically integrated by (3.17). Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} A(q^n) \to \int_{\mathcal{M}} A(q)\, d\pi \quad \text{a.s.}$$

for almost all starting points $q^0 \in \mathcal{M}$.

Random Time Hybrid Monte Carlo

In order to prove convergence of the classical HMC scheme, we have assumed in the previous section that the potential energy is bounded from above. As explained in the discussion just above Lemma 3.1, another possibility is to modify the HMC scheme as in [221]. The modification consists in transforming the fixed parameter τ into a random variable, distributed with a density $\mathcal{T}(\tau)$. This ensures that resonance effects are avoided. We call this scheme "Random Time Hybrid Monte Carlo" (RTHMC).

The only property required on $\mathcal{T}$ is that $\mathcal{T}$ is continuous and positive on $\mathbb{R}_+$. The corresponding Markov transition kernel reads, for $q \in \mathcal{M}$ and $B \in \mathcal{B}(\mathcal{M})$,

$$P(q, B) = \int_{\mathbb{R}^{3N} \times \mathbb{R}_+} \mathbf{1}_{\{\Pi_1 \Phi_\tau(q,p) \in B\}} \mathcal{P}(p) \mathcal{T}(\tau)\, dp\, d\tau. \tag{3.33}$$

Notice that π is still an invariant probability measure for this Markov chain, so condition (3.7) holds true. Therefore, to get convergence results, we only need to show condition (3.8). This is done in two steps, as for the classical HMC scheme.

The first lemma states that there is a positive probability to go from one state q to any neighborhood of any state q' in one RTHMC iteration.

Lemma 3.4 (RTHMC accessibility). *Assume that $V \in C^1(\mathcal{M})$ and $D^2V \in L^\infty(\mathbb{R}^3)$. Then for any $q_0, q_1 \in \mathcal{M}$, and there exists $\tau^* > 0$ such that, for all $0 < \tau \le \tau^*$, there exists $p \in \mathbb{R}^{3N}$ with $\Pi_1 \Phi_\tau(q_0, p) = q_1$.*

Proof. A similar idea is used in [301] in a slightly different context. If V is identically equal to zero, then going from q_0 to q_1 is possible through the choice of (say) the initial momenta $p^* = M(q_1 - q_0)/\tau$ for some evolution time $\tau > 0$. We then consider the rescaled equation

$$M\ddot{q}_\epsilon(t) = -\epsilon \nabla V(q_\epsilon(t)) \tag{3.34}$$

and the associated flow ϕ_ϵ. Setting

$$F(\epsilon, p) = \phi_\epsilon(\tau, q_0, p) - q_1,$$

the function F is $C^1(\mathbb{R} \times \mathbb{R}^{3N})$ (we use here the assumption $D^2V \in L^\infty(\mathbb{R}^3)$), $F(0, p^*) = 0$ and $\partial_p F(0, p^*) = \tau M^{-1}$ is invertible. In view of the implicit function theorem, there exists $\epsilon^* > 0$ such that for all $0 \le \epsilon \le \epsilon^*$, there exists p_ϵ such that $F(\epsilon, p_\epsilon) = 0$.

This shows (by the change of variables $t \to \epsilon t$ in (3.34) for $0 < \epsilon \le \epsilon^*$) that $\Pi_1 \Phi_{\epsilon\tau}(q_0, p_\epsilon/\epsilon) = q_1$. □

Condition (3.8) can then be obtained in the same way as for the classical HMC scheme, the proof following the same lines as for Lemma 3.2.

Lemma 3.5 (RTHMC irreducibility). *Provided that $V \in C^1(\mathcal{M})$ and $D^2V \in L^\infty(\mathcal{M})$, the transition kernel (3.33) of the RTHMC Markov chain satisfies condition (3.8).*

Proof. Consider $B \in \mathcal{B}(\mathcal{M})$ such that $\mu^{\mathrm{Leb}}(B) > 0$, and $q \in \mathcal{M}$. We want to show that $P(q, B) > 0$ for P defined by (3.33). For the sake of simplicity, we assume here that all particle masses are equal to 1.

The proof relies on the fact that, for a given q and for $\tau > 0$ small enough, the mapping $p \mapsto \Pi_1 \Phi_\tau(q, p)$ is invertible. Denote $J_B(q, \tau) = \{p \in \mathbb{R}^{3N} \mid \Pi_1 \Phi_\tau(q, p) \in B\}$, and consider $\psi_\tau : J_B(q, \tau) \to B$ such that $\psi_\tau(p) = \Pi_1 \Phi_\tau(q, p)$.

We first show that ψ_τ is an injective function for $\tau > 0$ small enough. From the equations of motion,

$$\psi_\tau(p) = q + p\tau - \int_0^\tau (\tau - s) \nabla V(\psi_s(p))\, ds.$$

Hence

$$\nabla_p \psi_\tau(p) = \tau\, \mathrm{Id} - \int_0^\tau (\tau - s) D^2 V(\psi_s(p)) \cdot \nabla_p \psi_s(p)\, ds. \tag{3.35}$$

Set $\alpha_R(s) = \sup_{|p| \le R} ||\nabla_p \psi_s(p) - s\mathrm{Id}||_\infty$. Since ∇V is a globally Lipschitz function, we have

$$\alpha_R(\tau) \le C \left(\int_0^\tau (\tau - s) \alpha_R(s)\, ds + \frac{\tau^3}{6} \right) \tag{3.36}$$

with $C = \|D^2V\|_{L^\infty(\mathcal{M})}$. We now consider $\tau_c^R = \sup\{\tau';\ \alpha_R(\tau) \le \tau/2 \text{ for all } \tau \in [0, \tau']\}$. From (3.36), we obtain that $\tau_c^R \ge \sqrt{2/C}$. Hence, we have

$$\forall \tau \in \left[0, \sqrt{2/C}\right],\ \alpha_R(\tau) \le \frac{\tau}{2}.$$

Inserting this inequality in (3.36), we also obtain that

$$\forall \tau \in \left[0, \sqrt{2/C}\right],\ \alpha_R(\tau) \le C\frac{\tau^3}{4}.$$

It follows that

$$\alpha(s) = \sup_{p \in \mathbb{R}^{3N}} ||\nabla_p \psi_s(p) - s\mathrm{Id}||_\infty \le C\frac{\tau^3}{4}. \tag{3.37}$$

Now,

$$\begin{aligned}(\psi_\tau(p_1) - \psi_\tau(p_2)) \cdot (p_1 - p_2) &= \int_0^1 (p_1 - p_2) \cdot \nabla_p \psi_\tau(p_2 + s(p_1 - p_2)) \cdot (p_1 - p_2)\, ds \\ &= \int_0^1 (p_1 - p_2) \cdot (\nabla_p \psi_\tau(p_2 + s(p_1 - p_2)) - \tau \mathrm{Id}) \cdot (p_1 - p_2)\, ds \\ &+ \quad \tau |p_1 - p_2|^2\end{aligned}$$

Let us suppose that $\psi_\tau(p_1) = \psi_\tau(p_2)$. Then

$$\tau |p_1 - p_2|^2 \leq \alpha(\tau)\ |p_1 - p_2|^2 \leq \frac{\tau}{2}\ |p_1 - p_2|^2$$

and we obtain $p_1 = p_2$. Hence, the mapping $J_B(q, \tau) \ni p \mapsto \psi_\tau(p) \in B$ is an injective function for $\tau \leq \sqrt{2/C}$.

We now show that this mapping is onto. We consider, for $q' \in B$, the C^1 function

$$G(\tau, p, q') = \psi_\tau(p) - q'.$$

Let us fix $q^* \in B$ such that, for all $\epsilon > 0$, $\mu^{\mathrm{Leb}}(B \cap B_\epsilon(q^*)) > 0$. Lemma 3.4 shows that there exists $\tau^* > 0$ such that

$$\forall \tau,\ 0 < \tau < \min(\tau^*, \sqrt{2/C}), \quad \exists p \in \mathbb{R}^{3N} \text{ s.t. } G(\tau, p, q^*) = 0.$$

Since $\partial_p G = \partial_p \psi_\tau$ is invertible (using (3.35) and the bound (3.37)), we obtain from the implicit function theorem that there exists a neighborhood $\mathcal{V}_\tau(p)$ of p and a neighborhood $\mathcal{V}_\tau(q^*)$ of q^* such that, for any $q' \in \mathcal{V}_\tau(q^*)$, there exists $p' \in \mathcal{V}_\tau(p)$ with $G(\tau, p', q') = 0$. This gives the desired result.

Thus, for $0 < \tau < \min(\tau^*, \sqrt{2/C})$, the mapping ψ_τ is one-to-one from $\mathcal{V}_\tau(p)$ onto $\mathcal{V}_\tau(q^*)$. Using (3.37), we also have $\mathrm{Det}(\nabla_p \psi_\tau(p)) = \tau^{3N}(1 + \mathrm{o}(1))$ uniformly in p. Hence, the mapping ψ_τ is invertible and $\mathrm{Det}(\nabla_p \psi_\tau^{-1}(q)) = \tau^{-3N}(1 + \mathrm{o}(1))$.

We are now in position to show that $P(q, B) > 0$. By contradiction, assume $P(q, B) = 0$. Then $\int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(x,p) \in B\}} \mathcal{P}(p)\, dp = 0$ for almost all τ. Therefore, for almost all $0 < \tau < \min(\tau^*, \sqrt{2/C})$, we have $\int_{\mathbb{R}^{3N}} \mathbf{1}_{\{\Pi_1 \Phi_\tau(x,p) \in B \cap \mathcal{V}_\tau(q^*)\}} \mathcal{P}(p)\, dp = 0$. Thus, a change of variable shows that

$$\int_{B \cap \mathcal{V}_\tau(q^*)} \mathcal{P}(\psi_\tau^{-1}(q))\, |\mathrm{Jac}(\psi_\tau^{-1}(q))|\, dq = 0$$

for almost all $0 < \tau < \min(\tau^*, \sqrt{2/C})$. This is however not possible since $\mathcal{P}$ is continuous and positive, $\mu^{\mathrm{Leb}}(B \cap \mathcal{V}_\tau(q^*)) > 0$, and $|\mathrm{Jac}(\psi_\tau^{-1}(q))| \sim \tau^{-3N}$ when $\tau \to 0$ so that $|\mathrm{Jac}(\psi_\tau^{-1}(q))| > 0$ for τ small enough.

We then get convergence of the average along a sample path (see Theorem 3.1):

Theorem 3.10. *Assume that $V \in C^2(\mathcal{M})$ and $D^2 V \in L^\infty(\mathcal{M})$. Let $(q^n)_{n \in \mathbb{N}}$ be the sequence of points generated by the RTHMC algorithm where the NVE equations of motion (3.16) are exactly integrated. Then*

$$\frac{1}{N} \sum_{n=0}^{N-1} A(q^n) \to \int_{\mathcal{M}} A(q)\, d\pi \quad \text{a.s.}$$

for almost all starting points $q^0 \in \mathcal{M}$.

We also obtain ergodicity and convergence of the densities as for the classical HMC scheme under the assumptions of Lemma 3.5 (see Theorem 3.3).

For the numerical discretization, we have to consider times $\tau_n = n\Delta t$, and a probability $\mathcal{T}$ on $\mathbb{N}$ such that $\mathcal{T}(n) > 0$ for all n (a Poisson law for instance). The time-step Δt has to be chosen small enough such that no resonance effect can appear.

3.2.3 Biased Random-Walk

The so-called biased Random-Walk, also known as the Brownian dynamics, or the overdamped Langevin dynamics, is defined by the fictitious dynamics

$$dq_t = -\nabla V(q_t)\, dt + \sigma\, dW_t, \tag{3.38}$$

where $(W_t)_{t\geq 0}$ is a $3N$-dimensional standard Wiener process and $\sigma = (2/\beta)^{1/2}$. The term "biased" refers to the fact that the brownian trajectories are affected by the drift term $-\nabla V$ which tends to draw them toward the local minima of V. The infinitesimal generator $\mathcal{A}$ associated with the biased Random-Walk (3.38) is defined by

$$\mathcal{A}g = -\nabla V \cdot \nabla g + \frac{\sigma^2}{2}\Delta g, \tag{3.39}$$

for $g \in C^2(\mathbb{R}^{3N})$. We denote by P^t the Markov semigroup associated with (3.38). Trajectorial existence and uniqueness for (3.38) is classical for globally Lipschitz force-fields [152, 224], namely for potential energies V satisfying for some positive constant L

$$\forall (x,y) \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}, \quad |\nabla V(x) - \nabla V(y)| \leq L\, |x-y|. \tag{3.40}$$

When this condition is not satisfied, it is possible to conclude to trajectorial existence and uniqueness for locally Lipschitz force-fields under the following hypothesis [152, 224]: there exist a function $W(q) \in C^2(\mathbb{R}^{3N})$ that goes to infinity at infinity and a positive constant c such that

$$\mathcal{A}W \leq cW. \tag{3.41}$$

Besides, under assumption (3.40) or (3.41), one can prove that the Markov process (3.38) is Feller [241].

From the Fokker-Planck equation associated with (3.38), it is easy to check that

$$\pi \text{ is an invariant probability measure of (3.38)}, \tag{3.42}$$

where π is the canonical position space distribution (3.6).

Convergence of the time average along one sample path

Let us consider the time average

$$S_T(A) = \frac{1}{T}\int_0^T A(q_t^x)\, dt, \tag{3.43}$$

where q_t^x is a sample path of (3.38) with the deterministic initial condition $q_0 = x$. Convergence results analogous to the results obtained for Markov chains can be extended to Markov processes, with an average (3.43) still taken only over one realization of the process (see [335] for a seminal contribution (that also considers discretization issues), [336, 337] for improvements and refinements, and [265] for a recent review).

To obtain an almost sure convergence of $S_T(A)$ to the position space average (and thus a positive answer to Question 1), the following theorem can be used:

Theorem 3.11 ([241, Theorem 8.1]). *Assume that the process q_t defined by (3.38) is Feller, that condition (3.42) holds true as well as the following condition:*

$$\text{for all } t, \text{ for all } q \in \mathbb{R}^{3N} \text{ and all open sets } \mathcal{O} \subset \mathbb{R}^{3N}, \quad P^t(q, \mathcal{O}) > 0. \tag{3.44}$$

Then, for π-almost every $q \in \mathbb{R}^{3N}$ and for any $A \in L^1(\pi)$,

$$\lim_{T \to \infty} S_T(A) = \int_{\mathbb{R}^{3N}} A(q) d\pi \quad a.s.$$

If ∇V is globally Lipschitz, then (3.44) holds true by standard results [287]. In other cases, a simple way to check condition (3.44) is to use a controllability argument inspired from [231, Lemma 3.4]. Central Limit Theorems (which would provide a convergence rate of $S_T(A)$ towards its limit and thus provide an answer to Question 2) can also be stated. We refer for example to [172].

Convergence of the densities

Ergodicity holds true whenever conditions (3.42) and (3.44) are satisfied (see [241, Theorem 6.1]). Question 3 can therefore be answered positively. To get an exponential convergence rate (in the W-total variation norm (3.15)), that is, to answer Question 4, one needs to show the stronger condition

$$\mathcal{A}W(q) \leq -cW(q) + b\mathbf{1}_C(q), \tag{3.45}$$

where $W \geq 1$ is a measurable function going to infinity at infinity, $c > 0$, $b \in \mathbb{R}$ and C is a compact set (compare this condition with condition (3.31) for Markov chains). We do not address this question in the present here (see [231, 336, 337] for examples of such studies).

Numerical implementation

The Euler-Maruyama numerical scheme associated to (3.38) reads, when taking integration steps $h = \Delta t^2/2$:

$$q^{n+1} = q^n - \frac{\Delta t^2}{2} \nabla V(q^n) + \beta^{-1/2} \Delta t R^n, \tag{3.46}$$

where $(R^n)_{n \in \mathbb{N}}$ is a sequence of i.i.d. $3N$-dimensional standard Gaussian random vectors.

For globally Lipschitz force-fields, the Euler-Maruyama scheme (3.46) converges: if the process q_t defined by (3.38) is ergodic, then the numerical Markov chain is ergodic and its invariant measure is close to the invariant measure of the original process (for Δt small enough) [231, Theorem 7.3].

However, for non-globally Lipschitz force-fields, it is not sufficient to consider the discretization (3.46) of the diffusion process alone. Indeed, examples of non-globally Lipschitz force-fields are known for which the Euler-Maruyama scheme fails [231, 283]. There are two ways out of this situation. First, convenient discretizations of (3.46) using some implicit integration can be used. Under some assumptions on the potential energy V, the corresponding numerical scheme converges: (i) there exists an invariant probability measure for the Markov chain formalizing the algorithm; (ii) empirical averages of observables (with at most polynomial growth) converge to position space averages up to $\mathrm{O}(\Delta t)$ terms (see [337]). However, implicit methods become untractable for large systems. Another approach may then be considered, the so-called "Metropolis-adjusted Langevin[3] algorithm" (MALA), proposed by Roberts and Tweedie in [283], which corrects the

[3] The term "Langevin" does not refer here to the Langevin dynamics as known in the Physics literature (see Section 3.2.4). In the Probability and Statistics fields, it is, for some authors, the name for the biased Random-Walk.

Euler-Maruyama discretization (3.46) by an additional acceptance/rejection step in a Metropolis-Hastings fashion. Therefore, there is no bias in the measure sampled. The algorithm consists in generating proposal steps using (3.46), and accepting or rejecting them according to a Metropolis-Hastings rule with the proposal density

$$\mathcal{P}(q,q') = \sqrt{\frac{\beta}{2\pi\Delta t^2}} \exp\left(-\frac{\beta}{2\Delta t^2}\left|q' - q + \frac{\Delta t^2}{2}\nabla V(q)\right|^2\right).$$

In the case of the MALA algorithm, using a potential energy $V \in C^1(\mathbb{R}^{3N})$ is enough to satisfy condition (3.8). Since π is by construction an invariant probability measure (and therefore condition (3.7) holds true), the Markov chain formalizing the algorithm is ergodic for almost all starting points, and the convergence results stated in Theorems 3.1 and 3.3 apply. On the other hand, conditions ensuring the Central Limit Theorem and geometric ergodicity (conditions (3.11) and (3.31), see Theorems 3.2 and 3.8) are not easy to check. We refer to [283,285] for such studies.

The only adjustable parameter of the algorithm is the time-step Δt. The rejection rate is a good indicator of efficiency. It is indeed well-known that a good sampling is a trade-off between decorrelation (to this end, larger time-steps are required) and acceptance rate (the larger the time-step, the larger the rejection rate). We refer for example to [284] where it is shown that, for tensorized distributions, the asymptotical optimal acceptance rate, when the dimension of the position space $\mathcal{M}$ goes to infinity, is 0.574. This theoretical result does not extend to more complicated situations. However, numerical experiments show that an acceptance/rejection rate about 50% leads to a rather efficient method.

In Section 3.4, we present numerical results obtained both with the Euler-Maruyama scheme and with the MALA scheme.

Comparison of MALA and the one-step HMC scheme

Note that choosing the time step h of the MALA algorithm such that $h = \Delta t^2/2$ makes the comparison between the MALA algorithm and the one-step Hybrid Monte Carlo methods easier since both schemes use (3.46) to generate a proposal. Indeed, when $\tau = \Delta t$ and $M = I_{3N}$, the HMC velocity is randomized every time-step and thus formally reads $p^n = \sqrt{\frac{1}{\beta}} M R^n$, where R^n is a $3N$-dimensional standard Gaussian random variable. Notice however that the acceptance/rejection steps differ since the HMC acceptance/rejection step involves the comparison of total energies and the Biased Random-Walk acceptance/rejection step involves the comparison of the potential energies alone. As far as the acceptance/rejection step is concerned, the MALA scheme uses the acceptance rate

$$r_{\mathrm{MALA}}(q^n, \tilde{q}^{n+1}) = \min\left\{1, \exp\left(-\beta\left[\frac{1}{2}\Delta t(\tilde{q}^{n+1} - q^n)\cdot(\nabla V(q^n) - \nabla V(\tilde{q}^{n+1})) + \mathrm{O}(\Delta t^2)\right]\right)\right\},$$

and $\tilde{q}^{n+1} - q^n = \sqrt{\frac{1}{\beta}} R^n + \mathrm{O}(\Delta t^2)$. On the other hand, considering Algorithm 3.3, the acceptance/rejection rate of the hybrid Monte Carlo algorithm reads

$$r_{\mathrm{HMC}}(q^n, \tilde{q}^{n+1}) = \min\left\{1, \frac{\exp\left(-\beta\tilde{H}^{n+1}\right)}{\exp\left(-\beta H^n\right)}\right\},$$

where H^n is the initial energy and $\tilde{H}^{n+1}$ is the energy at the end of the trajectory. If a Velocity-Verlet scheme is used to compute the trajectory,

$$r_{\mathrm{HMC}}(q^n, \tilde{q}^{n+1}) = \min\left\{1, \exp\left(-\beta\left[\frac{\Delta t}{2} M^{-1} p^n \cdot (\nabla V(q^n) - \nabla V(\tilde{q}^{n+1})) + \mathrm{O}(\Delta t^2)\right]\right)\right\}.$$

So the acceptance/rejection steps of the MALA algorithm on the one hand and of the HMC algorithm on the other are the same up to second order terms.

3.2.4 Langevin dynamics

The paradigm of Langevin dynamics is to introduce in the Newton equations of motion (3.16) some fictitious brownian forces modelling fluctuations, balanced by viscous damping forces modelling dissipation. More precisely, the equations of motion read here

$$\begin{cases} dq_t = & M^{-1}p_t\,dt, \\ dp_t = & -\nabla V(q_t)\,dt - \xi M^{-1}p_t\,dt + \sigma\,dW_t, \end{cases} \tag{3.47}$$

where $(W_t)_{t\geq 0}$ is a $3N$-dimensional Wiener process. The parameters ξ and σ represent the magnitude of the fluctuations and of the dissipation respectively, and are linked by the fluctuation-dissipation relation:

$$\sigma = (2\xi/\beta)^{1/2}, \tag{3.48}$$

where $\beta = 1/k_BT$. Therefore, there remains one adjustable parameter in the model. Let us remark that the biased Random-Walk (3.38) is obtained from the Langevin dynamics (3.47) by letting the mass matrix M go to zero and by setting $\xi = 1$, which amounts here to rescaling the time.

The infinitesimal generator $\mathcal{A}$ associated to the SDE (3.47) reads:

$$\mathcal{A}g(q,p) = M^{-1}p\cdot\nabla_q g(q,p) - (\xi M^{-1}p + \nabla V(q))\cdot\nabla_p g(q,p) + \frac{\sigma^2}{2}\Delta_p g(q,p), \tag{3.49}$$

for $g \in C^2(\mathbb{R}^d \times \mathbb{R}^{3N})$. The proof of trajectorial existence and uniqueness follows the same lines as for the biased Random-Walk case, with the same kind of assumptions (globally Lipschitz force fields ∇V or a Lyapunov condition analogous to (3.41)). It is straightforward to show that the canonical probability measure (3.3) is a steady state of the Fokker-Planck equation associated with (3.47).

Convergence results

The same results hold true for the Langevin process as the ones stated in Sections 3.2.3 and 3.2.3 for the biased Random-Walk, the proofs following the same lines. We refer to [231] for further details concerning condition (3.44) (where $\mathbb{R}^{3N}$ is to be replaced by $\mathbb{R}^{3N}\times\mathbb{R}^{3N}$ and P^t is now the Markov semigroup associated with the Langevin dynamics). We also refer to [159] for a remarkable work allowing, under some assumptions of local regularity and growth at infinity on the potential energy V, to obtain geometrical convergence of the density $P^t(q,\cdot)$ toward the invariant measure, in some weighted Sobolev norms. In particular, estimates of the convergence rate involving M, ξ, β and V, can be explicitely derived.

Questions 1 and 3 can therefore be answered positively. Question 4 can also be answered positively when a convenient drift condition can be stated (condition (3.45) where $\mathcal{A}$ is now the infinitesimal generator associated to (3.47)).

Numerical implementation

There are several ways to compute numerically an invariant distribution using a Langevin dynamics:

(i) with a Metropolized scheme as for the biased Random-Walk case (see [298] and Section 6.1.2 for an application to Variational Monte-Carlo);

(ii) with convenient discretizations and a step-size Δt sufficiently small ensuring the sampling from an invariant measure close to the canonical measure (3.3);

(iii) by extending usual NVE schemes used in deterministic MD simulations to the case of the Langevin dynamics (the quasi-symplectic schemes of [242]);
(iv) by using splitting ideas borrowed from integration methods for deterministic flows (see e.g. [146]).

It is not completely understood which integration scheme is the most efficient [244, 311, 369], especially because the comparison benchmarks vary from one field to another. The last two ways are the most convenient in many applications, and allows usually to take larger time steps than for pure NVE simulations since the scheme is intrinsically more stable in view of its dissipative properties. Unfortunately, to our knowledge, there is no theoretical proof of convergence for the resulting schemes. Let us now detail successively the last three approaches.

First-order schemes with invariant probability

General results of error analysis hold true for the numerical discretization of the Langevin equation for globally Lipschitz force fields [231]. In this case, the resulting numerical Markov chain is ergodic for usual discretization schemes (including the Euler-Maruyama disretization) and their invariant measures are close to the invariant measure of the original process (for Δt small enough).

The results are not the same for only locally Lipschitz force fields. Some classes of discretized schemes however behave properly under additional assumptions on the potential energy. This is the case for the so-called split-step Backward Euler-method proposed in [231]. Applied to the Langevin equation (3.47), this algorithm reads

$$\begin{cases} q^{n+1} = q^n + \Delta t\ M^{-1}p^* \\ p^* \ = p^n - \xi \Delta t M^{-1}p^* - \Delta t \nabla V(q^{n+1}) \\ p^{n+1} = p^* + \sigma\sqrt{\Delta t}G^n \end{cases} \tag{3.50}$$

where $(G^n)_{n\in\mathbb{N}}$ is a sequence of $3N$-dimensional i.i.d. Gaussian random vectors. Unfortunately, this method is implicit (see the first two equations, to be solved for (q^{n+1}, p^*)), therefore not convenient for MD simulations of large systems. The following *explicit* scheme is therefore prefered

$$\begin{cases} p^* \ = p^n - \xi \Delta t\, M^{-1}p^* - \Delta t \nabla V(q^n) \\ q^{n+1} = q^n + \Delta t\, M^{-1}p^* \\ p^{n+1} = p^* + \sigma\sqrt{\Delta t}G^n \end{cases} \tag{3.51}$$

where $(G^n)_{n\in\mathbb{N}}$ is a sequence of $3N$-dimensional i.i.d. Gaussian random vectors.

We now turn to the numerical analysis of (3.51). Let us denote by $\mathcal{F}_n$ the σ-algebra of events up to and including the n-th iteration. We need to prove condition (3.7) and condition (3.8) to state a Law of Large Number theorem (see Theorem 3.1). The accessibility condition (3.8) is easily seen to be satisfied (by arguments similar to those of Section 3.2.3 in this time discrete case). We now prove condition (3.7), that is, the existence of an invariant probability measure. For this purpose, we need to make some assumptions on the potential energy V, similar to those of [231], to state a Lyapunov inequality for the discretized process. Indeed, we want to make use of the following theorem:

Theorem 3.12 ([231, Theorem 2.5]). *Denote by P the transition kernel associated with the Markov chain formalizing (3.51), assumed to be Feller. Assume that (3.8) is satisfied and that there exist a function $W_{\Delta t}(q,p) \geq 1$, going to infinity at infinity, and two real numbers $b \in (0,1)$ and $c > 0$ such that*

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) \,|\, \mathcal{F}_n) \leq b\ \mathbb{E}(W_{\Delta t}(q^n, p^n)) + c, \tag{3.52}$$

where (q^n, p^n) is the discrete trajectory given by (3.51). Then there exists an invariant probability measure $\mu_{\Delta t}$, and condition (3.7) holds true.

The numerical scheme then converges (with respect to the measure $d\mu_{\Delta t}$) in the sense of Questions 1 to 4. The question of estimating the distance between $\mu_{\Delta t}$ and the canonical measure μ has been addressed in e.g. [231, 337].

Let us now find $W_{\Delta t}$, b and c satisfying (3.52). We assume that the potential energy V is in $C^2(\mathbb{R}^{3N})$ and satisfies a one-sided Lipschitz condition: there exists $C > 0$ such that

$$\forall a, b \in \mathbb{R}^{3N},\ (\nabla V(a) - \nabla V(b)) \cdot (a - b) \leq C|a - b|^2. \tag{3.53}$$

We also assume that there exist $A, B > 0$ such that

$$\forall q \in \mathbb{R}^{3N},\ -\nabla V(q) \cdot M^{-1} q \leq A - B\left(V(q) + \frac{\xi^2}{4} q^T M^{-1} q\right). \tag{3.54}$$

These conditions are satisfied for example for potential energies growing quadratically at infinity. The following result, strongly inspired from [231], can then be stated:

Lemma 3.6. *Let (q^n, p^n) be the discrete trajectory given by (3.51). Let us assume that V is bounded from below and let us set $m = \max\{m_1, \ldots, m_N\}$,*

$$W(q,p) = 1 + \frac{1}{2} p^T M^{-1} p + \frac{\xi^2}{4} q^T M^{-1} q + V(q) - \inf V + \frac{\xi}{2} p^T M^{-1} q \tag{3.55}$$

and $W_{\Delta t}(q,p) = W(q,p) + \dfrac{\xi}{4m^2}\Delta t|p|^2$. When (3.53) and (3.54) are satisfied, and that

$$0 \leq \Delta t \leq \frac{\xi}{\xi^2/m + 4C}. \tag{3.56}$$

Then $W_{\Delta t}$ satisfies (3.52) for some $c > 0$, $0 < b < 1$.

Proof. Consider the numerical scheme (3.51). Some computations give

$$\begin{aligned} W(q^{n+1}, p^*) - W(q^n, p^n) \leq & -\frac{\xi \Delta t}{2m^2}\left(1 - \frac{\xi \Delta t}{2m}\right)|p^*|^2 - \frac{\xi \Delta t}{2}\nabla V(q^n) \cdot M^{-1} q^n \\ & + V(q^n + \Delta t\, M^{-1} p^*) - V(q^n) - \Delta t \nabla V(q^n) \cdot M^{-1} p^*. \end{aligned}$$

The one-sided Lipschitz condition (3.53) allows to handle the term $V(q^n + \Delta t M^{-1} p^*) - V(q^n) - \Delta t \nabla V(q^n) \cdot M^{-1} p^*$. The condition (3.54) allows to handle the term $-\dfrac{\xi \Delta t}{2}\nabla V(q^n) \cdot M^{-1} q^n$. When (3.56) is satisfied, it then follows

$$W(q^{n+1}, p^*) - W(q^n, p^n) \leq A\frac{\xi \Delta t}{2} - B\frac{\xi \Delta t}{2}\left(V(q^n) + \frac{\xi^2}{4} q^n \cdot M^{-1} q^n\right) - \frac{\xi}{4m^2}\Delta t|p^*|^2. \tag{3.57}$$

Recalling $W_{\Delta t}(q,p) = W(q,p) + \dfrac{\xi}{4m^2}\Delta t|p|^2$, we obtain

$$\begin{aligned} W_{\Delta t}(q^{n+1}, p^*) - W_{\Delta t}(q^n, p^n) &\leq A\frac{\xi \Delta t}{2} - B\frac{\xi \Delta t}{2}\left(V(q^n) + \frac{\xi^2}{4} q^n \cdot M^{-1} q^n\right) - \frac{\xi}{4m^2}\Delta t|p^n|^2 \\ &\leq A\frac{\xi \Delta t}{2} - B' W_{\Delta t}(q^n, p^n) \end{aligned}$$

for some $B' > 0$. The final step $p^{n+1} = p^* + \sigma\sqrt{\Delta t} G^n$ leads to

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(W_{\Delta t}(q^{n+1}, p^*)) + \mathbb{E}|\sigma\sqrt{\Delta t} G^n|^2,$$

so that

$$\mathbb{E}(W_{\Delta t}(q^{n+1}, p^{n+1}) \mid \mathcal{F}_n) \leq b\, \mathbb{E}(W_{\Delta t}(q^n, p^n)) + c \tag{3.58}$$

for some $c > 0$, $0 < b < 1$.

Algorithms derived from the Verlet scheme

Let us now turn to the second approach, and describe algorithms generalizing the Verlet algorithm, and therefore widely used in practice; on the other hand, there are no convergence results at this date to our knowledge (only consistency results are known). One such algorithm is the BBK algorithm, proposed by Brünger, Brooks and Karplus [45]. Another example is the quasi-symplectic algorithm of [242].

We focus in the sequel on the BBK algorithm, which is well-suited only for small values of ξ [244, 299] (otherwise, algorithms from [4] or the Langevin impulse scheme [310] (see below) should be used). It is a modification of the usual velocity-Verlet scheme obtained by adding a term $-\xi \dfrac{p_i}{m_i} + \dfrac{\sigma_i}{\sqrt{\Delta t}} G_i^n$ to the force f_i exerted on particle i (the relation between ξ and σ_i will be made precise below). This may explain its popularity since it only asks for slight modifications of standard MD codes. The random forcing terms G_i^n ($i \in \{1, \ldots, N\}$ is the label of the particles, n is the iteration index) are standard i.i.d. Gaussian random variables. The scheme reads:

$$\begin{cases} p_i^{n+1/2} = p_i^n + \dfrac{\Delta t}{2}\left(-\nabla_{q_i} V(q^n) - \xi \dfrac{p_i^n}{m_i} + \dfrac{\sigma_i}{\sqrt{\Delta t}} G_i^n\right), \\ q_i^{n+1} = q_i^n + \Delta t\, \dfrac{p_i^{n+1/2}}{m_i}, \\ p_i^{n+1} = \dfrac{1}{1 + \frac{\xi \Delta t}{2m_i}} \left(p_i^{n+1/2} - \dfrac{\Delta t}{2} \nabla_{q_i} V(q^{n+1}) + \sigma_i \dfrac{\sqrt{\Delta t}}{2} G_i^{n+1}\right). \end{cases} \tag{3.59}$$

We now make precise the relation between ξ and σ_i by considering the case when there are no forces. When $\nabla V = 0$, the BBK algorithm reads

$$\left(1 + \frac{\xi}{2m_i}\Delta t\right) p_i^{n+1} = \left(1 - \frac{\xi}{2m_i}\Delta t\right) p_i^n + \sigma_i \frac{\sqrt{\Delta t}}{2}\left(G_i^n + G_i^{n+1}\right). \tag{3.60}$$

We see that, if $\mathbb{E}(p_i^n) = 0$, then $\mathbb{E}(p_i^{n+1}) = 0$. Choosing p_i^0 such that $\mathbb{E}(p_i^0) = 0$, we have $\mathbb{E}(p_i^n) = 0$ for all n. Let us now denote by $K_i^n = \mathbb{E}((p_i^n)^2)$ the variance of p_i^n. Setting $\gamma_i = \dfrac{\xi \Delta t}{2m_i}$, one has

$$K_i^{n+1} = \left(\frac{1 - \gamma_i}{1 + \gamma_i}\right)^2 K_i^n + \frac{3\sigma_i^2 \Delta t}{(1 + \gamma_i)^3}.$$

The above recursion is of the general form $x_{n+1} = a x_n + b$, and has a fixed point provided $a < 1$, which is always the case here since $\gamma_i > 0$. This fixed point K_i^∞ is such that

$$\frac{1}{m_i} K_i^\infty = \frac{3\sigma_i^2}{2\xi(1 + \gamma_i)}. \tag{3.61}$$

Setting σ_i to the value

$$\sigma_i^{\Delta t} = \sqrt{\frac{2\xi(1 + \gamma_i)}{\beta}} = \sqrt{\frac{2\xi}{\beta}\left(1 + \frac{\xi \Delta t}{2m_i}\right)}, \tag{3.62}$$

we see that $K_i^\infty = \dfrac{3m_i}{\beta}$, which is indeed the expected value (the kinetic temperature is correct). Note that (3.62) gives the magnitude of the random forcing that should be used in numerical

simulations if one wants the kinetic temperature to be correct. Otherwise, if σ is chosen according to (3.48), the time-averaged kinetic temperature is lower than the target temperature T, and the error is of order Δt, as can be seen from (3.61). This is consistent with the results obtained in [369] from a modified equation approach. Note that using (3.62) instead of (3.48) does not improve the configurational sampling accuracy (the error on the configurational sampling is of order Δt with both choices (3.48) and (3.62)).

Another modification of the BBK algorithm has been proposed in [306]. It amounts to using the same Gaussian random variables in the first and the third lines of (3.59). In this case, there is no bias on the kinetic temperature with the choice (3.48).

Schemes based on splitting

A third approach, more recent, is to design algorithms based on a operator splitting method. The Langevin Impulse algorithm, proposed in [310], is such an algorithm. When $\nabla V = 0$ and $M = \mathrm{Id}$, the Langevin dynamics

$$\begin{cases} dq_t = p_t\, dt, \\ dp_t = -\gamma p_t\, dt + \sigma\, dW_t, \end{cases} \tag{3.63}$$

can be integrated explicitely by integrating first the Ornstein-Uhlenbeck process on the momentum, and integrating once again to obtain the evolution of the positions. It holds

$$p_t = \mathrm{e}^{-\gamma t} p_0 + \sigma \int_0^t \mathrm{e}^{-\gamma(t-s)}\, dW_s = \mathrm{e}^{-\gamma t} p_0 + P_t,$$

where P_t is a gaussian process such that

$$\mathbb{E}(P_t^2) = \sigma^2 \int_0^t \mathrm{e}^{-2\gamma(t-s)}\, ds = \frac{1-\mathrm{e}^{-2\gamma t}}{\beta}.$$

Then,

$$q_t = q_0 + \int_0^t p_s\, ds = q_0 + \frac{1-\mathrm{e}^{-\gamma t}}{\gamma}\, p_0 + \sigma \int_0^t \int_0^s \mathrm{e}^{-\gamma(s-u)}\, dW_u\, ds = q_0 + \frac{1-\mathrm{e}^{-\gamma t}}{\gamma}\, p_0 + Q_t,$$

where the random variable Q_t can be rewritten as

$$Q_t = \int_0^t \int_u^t \sigma \mathrm{e}^{-\gamma(s-u)}\, ds\, dW_u.$$

Therefore, Q_t is a centered gaussian process of variance

$$\mathbb{E}(Q_t^2) = \int_0^t \left[\int_u^t \sigma \mathrm{e}^{-\gamma(s-u)}\, ds \right]^2 du = \frac{\sigma^2}{\gamma^2} \int_0^t \left(1 - \mathrm{e}^{-\gamma(t-u)}\right)^2 du = \frac{1}{\beta\gamma} \left[2t - \frac{3 - 4\mathrm{e}^{-\gamma t} + \mathrm{e}^{-2\gamma t}}{\gamma} \right].$$

However, the variables Q_t and P_t are correlated since

$$\mathbb{E}(P_t Q_t) = \mathbb{E}\left[\left(\int_0^t \frac{\sigma}{\gamma} \left(1 - \mathrm{e}^{-\gamma(t-u)}\right)\, dW_u \right) \left(\int_0^t \mathrm{e}^{-\gamma(t-u)}\, dW_u \right) \right].$$

Therefore

$$\mathbb{E}(P_t Q_t) = \frac{\sigma}{\gamma} \int_0^t \left(1 - \mathrm{e}^{-\gamma(t-u)}\right) \mathrm{e}^{-\gamma(t-u)}\, du = \frac{1}{\gamma\beta} \left(1 - \mathrm{e}^{-\gamma t}\right)^2.$$

Combining the integration of the flow (3.63) with the straightforward integration of the flow

$$\begin{cases} dq_t = 0, \\ dp_t = -\nabla V(q_t)\, dt, \end{cases}$$

the discretization proposed in [311] is recovered. Other discretizations of Langevin dynamics were obtained using splitting ideas (see e.g. [107, 280] and Section 4.3.1 for a precise statement of the corresponding scheme).

This approach is more rigorous than other classical algorithms to integrate the Langevin dynamics such as the ones described in [4]. The idea of those algorithms is to exactly integrate the dynamics when the forces vary linearly with respect to time. In practice, forces are interpolated in time between two successive time steps.

3.3 Deterministic molecular dynamics sampling

We now turn in this section to purely deterministic methods. These methods rely on the following idea: a system in the canonical ensemble can be considered as a system interacting with an external heat bath, the interaction being such that, at equilibrium, the physical system variables are distributed according to the canonical measure (3.3). Thus, the idea is to consider an extended system composed of the physical variables and some additional variables modelling the bath. Various dynamics have been proposed in this vein.

In this section, we first consider the Nosé-Hoover dynamics and its generalization to the Nosé-Hoover chains [171, 229, 260, 346]. Then, we consider the Nosé-Poincaré method [35] and the Recursive Multiple Thermostats method, which has been recently proposed in [206].

3.3.1 The Nosé-Hoover and Nosé-Hoover chains methods

The Nosé-Hoover (NH) method, proposed by Hoover, consists in describing the heat bath by two scalar variables, its "position" η and its "momentum" ξ, and to *postulate* the following dynamics for the extended set of variables [171, 260]:

$$\begin{cases} \dfrac{dq_i}{dt} = \dfrac{p_i}{m_i}, \\ \dfrac{dp_i}{dt} = -\nabla_{q_i} V - \dfrac{p_i \xi}{Q}, \\ \dfrac{d\eta}{dt} = \dfrac{\xi}{Q}, \\ \dfrac{d\xi}{dt} = \displaystyle\sum_{i=1}^{N} \dfrac{p_i^2}{m_i} - g k_{\mathrm{B}} T, \end{cases} \tag{3.64}$$

where V is the potential energy of the system, g is a parameter we will fix later and T is the target temperature. The parameter Q represents the mass of the thermostat; it is a free parameter that the user has to choose. The quantity

$$\tilde{H}_{\mathrm{NH}} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(q) + \frac{\xi^2}{2Q} + g k_{\mathrm{B}} T \eta \tag{3.65}$$

is an invariant of the dynamics (3.64), which also preserves the measure

$$d\mu_{\mathrm{NH}} = \exp(3N\eta)\, dq\, dp\, d\eta\, d\xi. \tag{3.66}$$

We refer to [113] for details on the origin of this dynamics. Let us just note here that (3.64) is not a Hamiltonian dynamics[4]. Since the dynamics preserves (3.65), it cannot be ergodic with respect to $d\mu_{\rm NH}$. Let us introduce the manifold $\mathcal{M}_{\rm NH}(E_0) = \left\{(q,p,\eta,\xi) \in \mathbb{R}^{6N+2} \mid \tilde{H}_{\rm NH}(q,p,\eta,\xi) = E_0\right\}$ and the measure

$$d\rho_{\rm NH} = \frac{d\sigma_{\rm NH}}{\|\nabla \tilde{H}_{\rm NH}\|_2}, \tag{3.67}$$

where $d\sigma_{\rm NH}$ is the area measure induced on $\mathcal{M}_{\rm NH}(E_0)$ by the measure (3.66), $\nabla \tilde{H}_{\rm NH}$ is the gradient of (3.65) with respect to all variables and $\|\cdot\|_2$ is the Euclidian norm. Then $d\rho_{\rm NH}$ is an invariant measure for the Nosé-Hoover dynamics (3.64).

Suppose now that the dynamics is ergodic with respect to $d\rho_{\rm NH}$ (note that this implies that $\tilde{H}_{\rm NH}$ is the unique invariant of (3.64)). Let us set $g = 3N$, where N is the number of particles. An easy computation (see [204, 346]) shows that the dynamics $(q(t), p(t))$ is ergodic with respect to the canonical measure (3.3), and thus provides a sampling of the phase space according to the canonical measure (at least before numerical discretization).

We emphasize the fact that, to the best of the authors knowledge, there is no rigorous proof in the literature showing that (3.64) is ergodic with respect to $d\rho_{\rm NH}$. Furthermore, it has been numerically observed that, for some systems, the dynamics $(q(t), p(t))$ does not seem to sample the phase space according to the canonical measure. For instance, this is the case with the one-dimensional harmonic oscillator, for which it is actually observed that the trajectory stays in a ring, namely that there exist $c, C > 0$ such that $c \leq q^2(t) + p^2(t) \leq C$ for all t (see [229, 346]). Some mathematical analysis of this fact can be read in [204].

To circumvent this difficulty, a generalization of the Nosé-Hoover dynamics (3.64) has been proposed by Martyna et al. in [229]. The idea consists in coupling the physical variables with a first thermostat as in (3.64), and to couple this thermostat with a second one, which can be coupled to a third one, and so on. The variables now include $2M$ additional scalar variables η_j and ξ_j, $j = 1, \ldots, M$, where the number M of thermostats is arbitrary. The corresponding dynamics is the so-called Nosé-Hoover chain dynamics (NHC) [229], in which there are M free parameters, $Q_1, \ldots, Q_M$, representing the masses of the M thermostats. The dynamics preserves an invariant $\tilde{H}_{\rm NHC}$ and a measure $d\mu_{\rm NHC}$ (which are the generalization of (3.65) and (3.66)).

As for the Nosé-Hoover dynamics, if the NHC dynamics is ergodic with respect to a measure $d\rho_{\rm NHC}$ built in the same way as $d\rho_{\rm NH}$, then the dynamics $(q(t), p(t))$ is ergodic with respect to the canonical measure. Provided that the number M of thermostats is large enough ($M \geq 3$ or 4 in practice), numerical simulations seem to show that this dynamics samples the phase space according to the canonical measure, even for systems such as the harmonic oscillator. Again, there is no rigorous proof showing that the NHC dynamics is *actually* ergodic with respect to $d\rho_{\rm NHC}$.

Regarding numerical integration, it seems interesting to work with algorithms that preserve the qualitative structure of the dynamics, that is *time reversibility* and *measure preservation*. Reversible-in-time and measure-preserving algorithms have been proposed in [230] (let us just mention here that they are based on a splitting of the dynamics). Simulation results discussed in Section 3.4 have been obtained with these algorithms.

3.3.2 The Nosé-Poincaré and the Recursive Multiple Thermostat methods

Both the Nosé-Hoover and the Nosé-Hoover chain dynamics suffer from not being Hamiltonian dynamics. As a consequence, the quasi-conservation by the numerical flow of the invariants $\tilde{H}_{\rm NH}$

[4] The Nosé-Hoover dynamics can be recast, after changing variables and time, as a Hamiltonian dynamics, the so-called Nosé dynamics [259]. However, the time of this dynamics does not correspond anymore to the physical time.

(see (3.65)) and $\tilde{H}_{\mathrm{NHC}}$ is not guaranted. On the contrary, when working with a Hamiltonian dynamics, it is known that the energy can be preserved by the numerical flow over very long times, provided symplectic algorithms are used (see [146, Chap. IX] and [278]). Another problem with Nosé-Hoover chains is the choice of the number of thermostats as well as their masses Q_j, which seem to have an influence on the results.

The Recursive Multiple Thermostat method (RMT) has been recently proposed by Leimkuhler and Sweet [206] to solve the difficulties that have just been highlighted. It is a Hamiltonian dynamics which, like the Nosé-Hoover or Nosé-Hoover chains dynamics, couples the physical variables with a heat bath. This dynamics is a generalization of the Nosé-Poincaré (NP) method [35], which is also a Hamiltonian method. The Nosé-Poincaré method consists in adding a single thermostat, whereas the RMT method consists in adding an arbitrary number M of thermostats, which are all coupled together and to the physical particles. This is not the case in the Nosé-Hoover chain dynamics, where only the first thermostat is coupled to the physical particles (and not the other thermostats).

The Nosé-Poincaré method is based on the following Hamiltonian:

$$H_{\mathrm{NP}}(q,p,\eta,\xi) = \eta\left(H\left(q,\frac{p}{\eta}\right) + \frac{\xi^2}{2Q} + gk_{\mathrm{B}}T\ln\eta - H_0\right), \tag{3.68}$$

where H is given by (3.4), H_0 is chosen such that $H_{\mathrm{NP}} = 0$ for the initial conditions, and where Q is some free parameter. Sampling properties and numerical algorithms are discussed in [35]. Let us just mention here that, as for the Nosé-Hoover dynamics, one has to set $g = 3N$ if the only invariant of the dynamics is H_{NP}.

The motivation for introducing the RMT method is the observation that, at least for some systems, numerical results seem to depend much less on the thermostat masses (which are user-chosen parameters) than with the Nosé-Poincaré method (see [206, 333]).

The numerical results that are presented in Section 3.4 have been obtained with the algorithms proposed in [35] and [206]. Let us note that different algorithms may have different numerical stabilities, and so different abilities to adequately sample the phase space with a trajectory of a given number of time steps. A new algorithm for the RMT dynamics has been proposed very recently in [20].

3.4 Numerical illustrations

The different methods presented above can be used to compute numerical approximations of phase space integrals. In some cases, theoretical convergence rates can be obtained. Typically, when a CLT holds true, the error is bounded by $Cn^{-1/2}$ (where n is the number of evaluations of the potential energy and/or of the forces; see the Central Limit Theorem 3.2) for some unknown prefactor C, depending on both the system and the observable A. An important issue is the value of the prefactor in numerical computations, which can greatly vary from one method to another one.

However, since this prefactor depends on A, it is not easy to compare the different methods in a general way. After a brief description of the alkane model in Section 3.4.1, we present in Section 3.4.2 an abstract criterion defined without any explicit dependence on an observable A. The criterion measures the deviation between the empirical distributions and the canonical distribution. This comparison can be performed for a fixed sample size (bearing in mind the computation of autocorrelation functions with a fixed computational cost for example), or, more fairly, at a fixed computational cost. Some improvements can also be achieved when combining different sampling techniques, or when resorting to strategies different from the computation of a single long trajectory. This is made precise in Section 3.4.5. In Section 3.4.6, we consider a specific case of a time-dependent observable A, which corresponds to a correlation function. The numerical results

that are obtained with this physical choice illustrate the conclusions drawn from the abstract criterion in Section 3.4.2.

3.4.1 Description of the linear alkane molecule

Linear alkanes are chemical compounds of the form CH_3-$(CH_2)_n$-CH_3. In this study, the so-called united-atom model [294] is used, in which the conformation of the molecule is completely characterized by the positions of the Carbon atoms. The presence of the Hydrogen atom is implicitely taken into account in the definition of the interaction potential energy the Carbon atoms are subjected to. The Carbon atoms of the linear alkane molecule are indexed from 1 to N, and their positions are described by the vector $q = (q_1, \ldots, q_N) \in \left(\mathbb{R}^3\right)^N$. We set $r_{i,j} = q_j - q_i$ and we denote by $d_{i,j} = |r_{i,j}|$ the distance between the Carbon atoms i and j.

In the model presented here, the interatomic potential energy involves two-, three-, and four-body interactions :

(1) two Carbon atoms connected by a covalent bond interact *via* a harmonic potential energy

$$V_2(d) = \frac{1}{2}k_0(d - d_0)^2; \tag{3.69}$$

(2) two Carbon atoms that are separated by three covalent bonds or more interact *via* a Lennard-Jones potential energy

$$V_{\mathrm{LJ}}(d) = 4\epsilon\left(\left(\frac{\sigma}{d}\right)^{12} - \left(\frac{\sigma}{d}\right)^{6}\right).$$

The parameters ϵ and σ depend on the atoms that interact, and can have three values: $\epsilon_{\mathrm{CH_3-CH_3}}$ and $\sigma_{\mathrm{CH_3-CH_3}}$ when two CH_3 groups interact (the end groups), $\epsilon_{\mathrm{CH_3-CH_2}}$ and $\sigma_{\mathrm{CH_3-CH_2}}$ when an interior group interacts with an end group, and $\epsilon_{\mathrm{CH_2-CH_2}}$ and $\sigma_{\mathrm{CH_2-CH_2}}$ when two CH_2 groups interact;

(3) three consecutive Carbon atoms C_i-C_{i+1}-C_{i+2} interact *via* the three-body interaction potential energy

$$V_3(\theta_i) = \frac{1}{2}k_\theta(\theta_i - \theta_0)^2, \tag{3.70}$$

where

$$\theta_i = \arccos\left(\frac{r_{i,i+1} \cdot r_{i+1,i+2}}{|r_{i,i+1}| \cdot |r_{i+1,i+2}|}\right) \tag{3.71}$$

is the bending angle of the C_i-C_{i+1}-C_{i+2} chain;

(4) lastly, four consecutive Carbon atoms C_i-C_{i+1}-C_{i+2}-C_{i+3} experience the four-body interaction potential energy

$$V_4(\phi_i) = u_{\mathrm{tors}}(\cos\phi_i), \tag{3.72}$$

where ϕ_i is the dihedral angle defined by

$$\cos\phi_i = -\frac{(r_{i,i+1} \times r_{i+1,i+2}) \cdot (r_{i+1,i+2} \times r_{i+2,i+3})}{|(r_{i,i+1} \times r_{i+1,i+2})| \cdot |(r_{i+1,i+2} \times r_{i+2,i+3})|} \tag{3.73}$$

and where the function u_{tors} is given by

$$u_{\mathrm{tors}}(x) = c_1(1 - x) + 2c_2(1 - x^2) + c_3(1 + 3x - 4x^3).$$

The potential energy of the linear alkane molecule eventually reads

$$V(q) = \sum_{i=1}^{N-1} V_2(d_{i+1,i}) + \sum_{i=1}^{N-2} V_3(\theta_i) + \sum_{i=1}^{N-3} V_4(\phi_i) + \sum_{i=1}^{N-4}\sum_{j=i+3}^{N} V_{\mathrm{LJ}}(d_{i,j}), \tag{3.74}$$

where the term V_{LJ} depends on the type of interaction considered.

The values of the parameters d_0, ϵ, σ, k_θ, θ_0, c_1, c_2 and c_3 are taken from [228]. In the system of units where the length unit is $l_0 = 1.53 \times 10^{-10}$ m and the energy unit is such that $k_B T = 1$ at $T = 300$ K, the time unit is $\bar{t} = 364$ fs, and the numerical values of the parameters are $d_0 = 1$, $\epsilon_{\mathrm{CH_3-CH_3}} = 0.294$, $\epsilon_{\mathrm{CH_3-CH_2}} = 0.241$, $\epsilon_{\mathrm{CH_2-CH_2}} = 0.198$, $\sigma_{\mathrm{CH_3-CH_3}} = \sigma_{\mathrm{CH_3-CH_2}} = \sigma_{\mathrm{CH_2-CH_2}} = 2.55$, $k_\theta = 208$ rad^{-2}, $\theta_0 = 1.187$ rad, $c_1 = 1.18$, $c_2 = -0.23$ and $c_3 = 2.64$. Notice that for these values of the parameters c_i, the function u_{tors} has a unique global minimum (at $\phi = 0$) and two local non-global minima. As far as the parameter k_0 is concerned, we set $k_0 = 1000$ (another possibility [228] is to constrain the C-C covalent bond length to be equal to d_0). We set the unit of mass such that the mass of each particle is equal to 1.

We note that $\sum_{i=1}^N \nabla_{q_i} V = 0$, and that $\sum_{i=1}^N q_i \times \nabla_{q_i} V = 0$. As a consequence, the Newton equations (3.16) not only preserve the energy, but also preserve the linear momentum $\sum_{i=1}^N p_i$ and the angular momentum $\sum_{i=1}^N q_i \times p_i$. Similarly, the Nosé-Hoover dynamics (3.64) also has additional invariants: besides (3.65), it preserves $e^\eta \sum_{i=1}^N p_i$ and $e^\eta \sum_{i=1}^N q_i \times p_i$. As a consequence, it cannot be ergodic with respect to (3.67). One can nevertheless recover correct sampling properties in the q variables by

- starting from an initial condition that satisfies $\sum_{i=1}^N p_i(0) = 0$ and $\sum_{i=1}^N q_i(0) \times p_i(0) = 0$, so that the linear and angular momenta are always equal to 0;
- setting $g = 3N - N_c$, where N_c is the number of conservation laws (besides the energy (3.65)).

In the case under study here, $N_c = 6$. The same kind of remarks also hold true for the Nosé-Hoover chain dynamics, the Nosé-Poincaré dynamics and the RMT method. The simulation results that we present below have been obtained with these choices. Note that there is no need for any modification for the stochastically perturbed MD methods.

The linear pentane $\mathrm{CH_3}$-$(\mathrm{CH_2})_3$-$\mathrm{CH_3}$ is the shortest linear alkane for which a two-body Lennard-Jones interaction (coupling the variables $d_{i,i+1}$, θ_i and ϕ_i all together) has to be taken into account. In addition, it involves only two dihedral angles and these two angles essentially determine the conformation of the molecule. Indeed, the covalent stretching and bending potential energies (namely, V_2 and V_3) are stiff and consequently the bond lengths and bending angles are statistically close to their equilibrium values at room temperature. Therefore, the linear pentane molecule is a good test case for it allows a simple reduced representation of the conformation while being a non-trivial model in which the internal degrees of freedom are coupled all together. For completeness, tests on longer molecules are performed in order to investigate the robustness of the numerical methods with respect to increasing configurational space dimensions.

Some reference empirical densities for the dihedral angles obtained through Importance sampling techniques are presented in Figure 3.1. They correspond to pentane, with $N = 10^9$ sample points.

3.4.2 Discrepancy of sample points

In order to quantitatively assess the quality of the samples generated by the various methods described above, we use a discrepancy criterion. Recall that the discrepancy D_n of a sequence $x = \{x_m\}_{0 \le m \le n-1}$ with values in $[0,1]^d$ is defined as (see [200])

$$D_n(x) = \sup_{y \in [0,1]^d} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{x_m \in [0,y]\}} - \mathrm{Volume}([0,y]) \right|, \tag{3.75}$$

where, for d-dimensional vectors y, z, we write $y \le z$ when $y_i \le z_i$ for all $1 \le i \le d$, and note $[0,y] = \{z \in [0,1]^d,\ z \le y\}$. The fact that $D_n(x) \to 0$ when $n \to \infty$ is equivalent (see [200, p.15]) to the fact that, for any Riemann integrable function A defined on $[0,1]^d$,

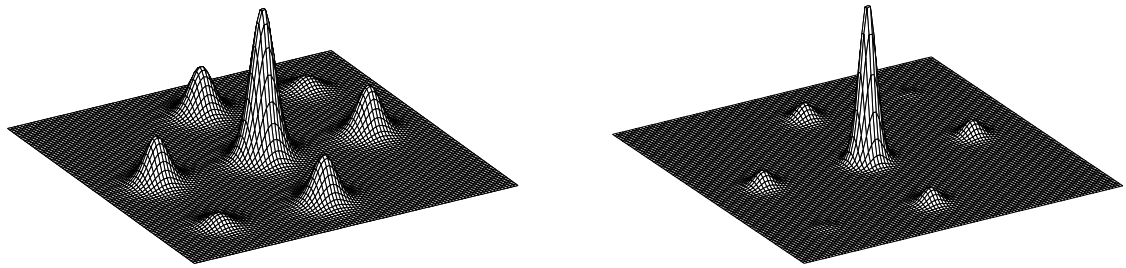


Fig. 3.1. Empirical probability distribution of the dihedral angles (ϕ_1, ϕ_2) of the pentane molecule generated with Importance sampling, for $\beta = 1$ (Left) and $\beta = 2$ (Right), with sample size $N = 10^9$ and $\epsilon_{\mathrm{CH_3-CH_3}} = 0.29$, $\epsilon_{\mathrm{CH_3-CH_2}} = 0$.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} A(x_m) = \int_{[0,1]^d} A(x)\, dx.$$

In addition, for functions A which have bounded variations $V_{\mathrm{HK}}(A)$ in the sense of Hardy and Krause [257], the following error estimate holds true:

$$\left| \frac{1}{n} \sum_{m=0}^{n-1} A(x_m) - \int_{[0,1]^d} A(x)\, dx \right| \leq V_{\mathrm{HK}}(A) D_n(x). \tag{3.76}$$

If $A \in C^d([0,1]^d)$, then its variation $V_{\mathrm{HK}}(A)$ has a simple expression (see [257, page 19]). If $d = 2$, which is the case we will be interested in below, then

$$V_{\mathrm{HK}}(A) = \int_{[0,1]^2} \left| \frac{\partial^2 A}{\partial x_1 \partial x_2} \right| dx + \int_0^1 \left| \frac{\partial A}{\partial x_1}(x_1, 1) \right| dx_1 + \int_0^1 \left| \frac{\partial A}{\partial x_2}(1, x_2) \right| dx_2.$$

As a consequence of (3.76), the convergence of $D_n(x)$ toward 0 implies the Law of Large Numbers, and the rate of convergence of $D_n(x)$ gives information about the convergence rate of the observable average.

In this framework, we intend for example to characterize the repartition of sample points in the subset $[-\pi, \pi]^2$ of the (ϕ_i, ϕ_j)-plane for two of the dihedral angles ϕ_i, ϕ_j. This can be achieved by considering the marginal ν_{ij} of the canonical density π with respect to the other degrees of freedom. Unfortunately, there is no simple exact expression of this marginal. We therefore consider the situation when all $\epsilon = 0$ (that is when the Lennard-Jones interactions are all turned off), in which case the marginal has the simple expression

$$d\nu_{ij}(\phi_i, \phi_j) = Z_\phi^{-2} \mathrm{e}^{-\beta V_4(\phi_i)} \mathrm{e}^{-\beta V_4(\phi_j)}\, d\phi_i d\phi_j, \tag{3.77}$$

with V_4 given by (3.72).

We then introduce the discrepancy criterion

$$D_n(\{q^m\}) = \sup_{(\phi_i,\phi_j)\in[-\pi,\pi]^2} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{\phi_i^m \leq \phi_i, \phi_j^m \leq \phi_j\}} - \int_{\{\psi_i \leq \phi_i, \psi_j \leq \phi_j\}} d\nu_{ij}(\psi_i, \psi_j) \right|, \tag{3.78}$$

which provides a bound on the L^∞ distance between the empirical distribution functions and the exact ones. Notice that the second integral factorizes as

$$\int_{\{\psi_i \le \phi_i, \psi_j \le \phi_j\}} d\nu_{ij}(\psi_i, \psi_j) = Z_\phi^{-2} \int_{\psi_i \le \phi_i} \mathrm{e}^{-\beta V_4(\psi_i)}\, d\psi_i \int_{\psi_j \le \phi_j} \mathrm{e}^{-\beta V_4(\psi_j)}\, d\psi_j,$$

and can therefore easily be computed using standard numerical techniques.

Numerically, we compute an approximate value of D_n as follows. Suppose that we have partioned the (ϕ_i, ϕ_j)-plane into K^2 boxes $B_{kl} = [\Phi_k, \Phi_{k+1}[\times[\Phi_l, \Phi_{l+1}[$ with $\Phi_k = -\pi + \frac{2k\pi}{K}$ for $0 \le k \le K-1$. The supremum in (3.78) is now taken over a finite set of elements:

$$D_n^K(q) = \sup_{1 \le k, l \le K} \left| \frac{1}{n} \sum_{m=0}^{n-1} \mathbf{1}_{\{\phi_i^m \le \Phi_k, \phi_j^m \le \Phi_l\}} - \int_{\{\psi_i \le \Phi_k, \psi_j \le \Phi_l\}} d\nu_{ij}(\psi_i, \psi_j) \right|. \tag{3.79}$$

We then compute the discrepancies for the sample points obtained by different methods with a fixed computational cost. The computational cost measures here the number of force or energy evaluations.

3.4.3 Choice of parameters

We describe here how we choose the parameters of the numerical methods for a fixed computational cost in the case of pentane. The cost has to be understood with respect to forces or energies evaluations. Notice that there is no parameter to tune for purely stochastic method such as the Rejection method and Importance sampling. For the Metropolized independence sampler, the only improvement that could be done is an undersampling. However, the quality of the samples is not changed by some reasonable undersampling (in the range $1 - 100$).

Stochastic methods

For the purely stochastic methods, we have worked with $g(q) = \tilde{Z}_q^{-1} \exp(-\beta \tilde{V}(q))$, where

$$\tilde{V}(q) = \sum_{i=1}^{N-1} V_2(d_{i+1,i}) + \sum_{i=1}^{N-2} V_3(\theta_i)$$

and $\tilde{Z}_q$ is a normalization constant. When expressed in internal coordinates (with the change of variables $R = (d_{2,1}, \ldots, d_{N,N-1}, \theta_1, \ldots, \theta_{n-2}) = h(q)$), the functions V_2 and V_3 are quadratic (see (3.69) and (3.70)), which makes it possible to actually sample from $g(R)\, dR$ (and so, from $g(q)\, dq$ up to a Jacobian term).

Hybrid Monte Carlo

The only relevant parameters are the time $\tau = k\Delta t$ and the time-step Δt. We generate several samples of size N with a computational cost equal to 10^6 forces or energies evaluations. Therefore, the product kN is a constant equal to 10^6. We compute the discrepancy (3.79) for each parameter values, averaging over 10 realizations (see Table 3.2). We found no systematic improvement using an undersampling procedure. We present the results under the form m (σ) where m is the mean of the discrepancies and σ the square-root of the variance.

The optimal choice within this set of parameters is $\Delta t = 0.025$ and $\tau = 10$. This corresponds to an acceptance rate of 0.7. When $\beta \neq 1$ and/or the molecule is longer, we choose a new time step Δt such that the acceptance/rejection rate is still around 0.7. Actually, the choice $\Delta t = 0.025$ remains convenient (though maybe not optimal) for a broad range of temperatures and sizes.

Table 3.2. Discrepancy results for the HMC algorithm.

Δt	τ	Discrepancy ($\epsilon = 0$)
0.02	1	0.106 (0.0310)
	5	0.0750 (0.0143)
	10	0.0532 (0.0141)
	20	0.400 (0.0107)
	50	0.0389 (0.00869)
	100	0.0550 (0.0163)
0.025	1	0.103 (0.0406)
	5	0.467 (0.0249)
	10	0.0389 (0.0183)
	20	0.0447 (0.0114)
	50	0.0481 (0.0201)
	100	0.0524 (0.0181)

Δt	τ	Discrepancy ($\epsilon = 0$)
0.01	1	0.0224 (0.0894)
	10	0.0692 (0.0352)
	100	0.0690 (0.0242)
0.03	1	0.0860 (0.0322)
	5	0.0486 (0.00875)
	10	0.503 (0.00704)
	20	0.410 (0.0111)
	50	0.0563 (0.0176)
	100	0.0540 (0.0157)
0.035	1	0.130 (0.0458)
	10	0.0478 (0.195)
	100	0.561 (0.347)

Biased Random-Walk

The only relevant parameter is Δt. We study the quality of the sampling for different values of this parameter for samples of size $N = 10^6$ (there is one computation of forces and energies per time step), see Table 3.3. We found no systematic improvement using an undersampling procedure.

Table 3.3. Discrepancy results for the biased random-walk.

Δt	Rejection rate	Discrepancy ($\epsilon = 0$)
0.01	0.022	0.190 (0.466)
0.02	0.18	0.125 (0.0298)
0.025	0.33	0.0920 (0.0362)
0.028	0.45	0.104 (0.0446)
0.03	0.53	0.110 (0.0362)
0.035	0.73	0.112 (0.0544)

The choice $\Delta t = 0.025$ or $\Delta t = 0.028$ seem reasonable. Notice that according to the discussion in Section 3.2.3, the optimal choice of Δt at $\beta = 1$ (giving the best symmetry estimate and the lowest discrepancy) is indeed expected to correspond to a rejection rate close to to the asymptotic optimal rejection rate for tensorized distributions (which is 0.426 [284]). When $\beta \neq 1$ and/or the molecule is longer, we choose a new time step Δt such that the acceptance/rejection rate is still around 0.5. Actually, the choice $\Delta t = 0.025$ remains convenient (though maybe not optimal) for a broad range of temperatures and sizes.

Discretized Langevin process

The only relevant parameters are the friction coefficient ξ and the time-step Δt. We study the quality of the sampling for different values of this parameter for samples of size $N = 10^6$ (there is one computation of forces and energies per time step), see Table 3.4. We found no systematic improvement using an undersampling procedure.

The results show that too small values of ξ have to be avoided (the random fluctuations are not large enough to cross barriers) as well as large values of ξ (where the stochasticity prevents the system to follow the physical dynamics). We set $\xi = 1$ and $\Delta t = 0.02$ in the sequel. This choice remains convenient (though maybe not optimal) for a broad range of temperatures and sizes.

Table 3.4. Discrepancy results for the Langevin dynamics.

Δt	ξ	Discrepancy ($\epsilon = 0$)	Δt	ξ	Discrepancy ($\epsilon = 0$)	Δt	ξ	Discrepancy ($\epsilon = 0$)
0.01	0.1	0.0582 (0.0175)	0.02	0.1	0.0529 (0.0144)	0.03	0.1	0.0487 (0.0134)
	0.5	0.0580 (0.0208)		0.5	0.0354 (0.00740)		0.5	0.0376 (0.00937)
	1	0.0689 (0.0219)		1	0.0339 (0.0142)		1	0.0311 (0.0120)
	5	0.0548 (0.0232)		5	0.0350 (0.0106)		5	0.0488 (0.0140)
	10	0.0427 (0.00849)		10	0.0441 (0.0161)		10	0.0575 (0.0155)

Nosé-Hoover chains

The parameters are the number M of thermostats, their masses, and the integration time step Δt. We set $\Delta t = 0.003$, which ensures a conservation of the energies up to a few percents in general. We use the two above statistical indicators of the quality of the sampling, as well as the time average of

$$A_2 = \frac{1}{3N} \sum_{i=1}^{N} \sum_{\alpha=x,y,z} p_{i,\alpha}^2, \quad A_4 = \frac{1}{3N} \sum_{i=1}^{N} \sum_{\alpha=x,y,z} p_{i,\alpha}^4.$$

In the long time limit, they should converge to $1/\beta$ and $3/\beta^2$. We also display $\triangle \tilde{H}/\tilde{H}$, which is the relative conservation of energies. We have observed that, in the case $\epsilon = 0$, the invariant is preserved with a much better accuracy than in the case $\epsilon = 0.29$ (this is due to the fact that, when $\epsilon \neq 0$, the end atoms of the chain should not be too close; we thus have to handle collisions, which lower the energy conservation accuracy). The results are presented in Table 3.5 for $N = 1,000,000$ and $\beta = 1$ (the values for $\triangle \tilde{H}/\tilde{H}$, $\langle A_2 \rangle$ and $\langle A_4 \rangle$ have been computed in the case $\epsilon = 0.29$).

Table 3.5. Discrepancy results for the Nosé-Hoover dynamics.

M	Q	$\triangle \tilde{H}/\tilde{H}$	$\langle A_2 \rangle$	$\langle A_4 \rangle$	Discrepancy ($\epsilon = 0$)
1	0.1	6 %	0.999981	3.06987	0.127
	1.0	4 %	0.999962	3.01696	0.074
	10.0	0.3 %	0.999922	4.37835	0.238
2	0.05; 0.05	1.5 %	1.00007	2.95343	0.080
	0.1; 0.1	1.2 %	1.00009	2.91847	0.143
	0.3; 0.3	3 %	1.00043	2.95486	0.169
	1.0; 1.0	0.4 %	0.999555	2.88511	0.232
	10.0; 10.0	0.1 %	0.997356	2.92125	0.189
	0.15; 0.01	3.7%	0.998261	2.92262	0.217
	0.75; 0.05	3.3%	0.998902	2.95794	0.163
	1.5; 0.1	0.1 %	0.993824	2.92667	0.242
	4.5; 0.3	0.2 %	0.995765	2.89965	0.277
	15.0; 1.0	0.2 %	0.971896	2.80145	0.338
	150.0; 10.0	0.15 %	0.988531	2.89529	0.352

We first see that the Nosé-Hoover chain dynamics is more stable than the Nosé-Hoover dynamics (for a given time step and given values of the thermostats, the drift of the invariant is smaller). The best results in term of discrepancy and closeness of $\langle A_2 \rangle$ and $\langle A_4 \rangle$ to their target values (1 and 3 here) are obtained here for $M = 1$ with $Q = 1$ or $M = 2$ with $Q_1 = Q_2 = 0.05$. We choose to work with the latter choice because the conservation of the invariants is better in this case. Note that different initial conditions lead to different discrepancy results. However, making

again the same test with different initial conditions (but still with $\Delta t = 0.003$), we have observed that the choice $Q_1 = Q_2 = 0.05$ seems to give better results than other choices.

On the other hand, if we set the time step to $\Delta t = 0.001$, it seems that the best choices are now $Q_1 = Q_2 = 0.1$ and $Q_1 = 0.15, Q_2 = 0.01$. In the following, when appropriate, we will comment the results obtained with these two different choices. Unless otherwise stated, we work with $Q_1 = Q_2 = 0.05$.

The Nosé-Poincaré and RMT methods

The parameters are the number M of thermostats, their masses, and the integration time step Δt. We set $\Delta t = 0.001$, which ensures a conservation of the hamiltonian up to a few percents in general. Note that we have decreased the time step in comparison to the Nosé-Hoover type method. This decrease is not due to energy conservation problems (the hamiltonian is preserved with a reasonnable accuracy when $\Delta t = 0.003$), but because it is quite hard, from the numerical results at $\Delta t = 0.003$, to select parameter values. In particular, discrepancy results vary in a large range for different initial conditions, so it is hard to assess that one parameter choice is better than another one. Selecting parameters has proved to be easier when working with $\Delta t = 0.001$.

We use the two above statistical indicators of the quality of the sampling, as well as the time average of A_2 and A_4 given above. As with the NHC method, we have observed that, in the case $\epsilon = 0$, the invariant is preserved with a much better accuracy than in the case $\epsilon = 0.29$. The results are presented in Table 3.5 for $N = 1,000,000$ and $\beta = 1$ (the values for $\triangle\tilde{H}/\tilde{H}$, $\langle A_2 \rangle$ and $\langle A_4 \rangle$ have been computed in the case $\epsilon = 0.29$).

Table 3.6. Discrepancy results for the Nosé-Poincaré dynamics.

M	Q	$\triangle\tilde{H}/\tilde{H}$	$\langle A_2 \rangle$	$\langle A_4 \rangle$	Discrepancy ($\epsilon = 0$)
1	0.1	0.02 %	0.999981	3.21418	0.269
	1.0	0.08 %	1.0	2.69515	0.304
	10.0	0.2 %	1.00024	4.98638	0.350
2	0.05; 0.05	0.15 %	1.0059	2.46228	0.320
	0.1; 0.1	0.2 %	1.00905	2.63986	0.460
	0.3; 0.3	0.3 %	1.01655	3.35365	0.360
	1.0; 1.0	0.06 %	1.01059	3.03896	0.373
	10.0; 10.0	4 %	1.0292	2.85634	0.328
	0.15; 0.01	1 %	1.00538	3.09675	0.344
	0.75; 0.05	0.3 %	1.00799	2.82565	0.297
	1.5; 0.1	0.1 %	1.01253	3.00398	0.281
	4.5; 0.3	0.1 %	0.996809	2.84965	0.225
	15.0; 1.0	0.6 %	1.03506	3.16739	0.377
	150.0; 10.0	0.03 %	1.02456	3.26963	0.310
	0.05; 0.1	1 %	1.00577	2.91749	0.277
	0.1; 0.2	1 %	1.00094	2.87149	0.292
	0.3; 0.6	2 %	1.02247	3.34102	0.347
	1.0; 2.0	0.03 %	0.999142	2.73679	0.263
	10.0; 20.0	1.2 %	1.02031	3.15916	0.341

The best result in terms of discrepancy leads to select $Q_1 = 4.5, Q_2 = 0.3$. This choice seems robust with respect to the initial condition. Depending on the numerical results at hand, other choices could be made. For a trajectory length of 10^6 steps, $Q_1 = 1.0, Q_2 = 2.0$ seems to give also good results. However, when the trajectory length is increased to 10^7 steps, the two more robusts choices seem to be $Q_1 = 4.5, Q_2 = 0.3$, that we selected above, and $Q_1 = 0.1, Q_2 = 0.2$. We will

comment in the following the results obtained with the latter choice. Unless otherwise stated, we work now with $Q_1 = 4.5, Q_2 = 0.3$.

3.4.4 Numerical results

The results are presented in Tables 3.7 to 3.9. For each method, 10 different simulations have been performed, and we give in the Tables the mean and the square-root of the variance (in brackets) of the 10 different results.

Table 3.7. Numerical results for the discrepancy (3.79) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$.

Method	Parameters	Discrepancy for 10^6 evaluations	Discrepancy for 10^7 evaluations
Importance sampling	-	0.00428 (0.00114)	0.00115 ($1.60.10^{-4}$)
Rejection	-	0.00856 (0.00204)	0.00256 ($4.98.10^{-4}$)
MIS	-	0.0228 (0.00416)	0.0225 ($7.75.10^{-4}$)
HMC	$\tau = 10\Delta t,\ \Delta t = 0.025$	0.0389 (0.0183)	0.0119 ($4.87.10^{-4}$)
BRW (Euler-Maruyama)	$\Delta t = 0.028$	0.0791 (0.0265)	0.0231 (0.00619)
BRW (MALA)	$\Delta t = 0.028$	0.104 (0.0446)	0.0343 (0.0139)
Langevin	$\Delta t = 0.02,\ \xi = 1$	0.0339 (0.0142)	0.0157 (0.00393)
NHC	$Q_1 = Q_2 = 0.05,\ \Delta t = 0.0025$	0.103 (0.036)	0.0456 (0.0117)
RMT	$Q_1 = 5, Q_2 = 7.5,\ \Delta t = 0.0025$	0.196 (0.142)	0.178 (0.177)

Table 3.8. Numerical results for the discrepancy (3.79) for the (ϕ_1, ϕ_3) distribution for C_9H_{20} in the case $\beta = 1$ and $K = 100$. The computational cost is fixed to 10^7 force or energy evaluations.

Method	Parameters	Discrepancy
Importance sampling	-	0.0205 (0.00544)
Rejection	-	0.192 (0.0379)
MIS	-	0.521 (0.0151)
HMC	$\tau = 10\Delta t,\ \Delta t = 0.02$	0.0261 (0.00846)
BRW (Euler-Maruyama)	$\Delta t = 0.025$	0.0402 (0.0229)
BRW (MALA)	$\Delta t = 0.025$	0.0477 (0.0129)
Langevin	$\Delta t = 0.025,\ \xi = 1$	0.0144 (0.00544)
NHC	$Q_1 = 0.15, Q_2 = 0.01,\ \Delta t = 0.0025$	0.0292 (0.0102)
NP	$Q = 5,\ \Delta t = 0.0025$	0.0386 (0.0095)

One can see that purely stochastic methods are very efficient for small alkane chains, but rapidly loose their efficiency when the length of the chain increases. Thus, the Langevin dynamics and the HMC method seem to be the most efficient methods, although other non purely stochastic methods also give good results. The Langevin, the HMC and the BRW (with Euler-Maruyama algorithm) methods keep the same efficiency whatever the length of the chain. This seems also to be the case for the NHC method. The efficiency of the BRW (with the MALA algorithm) decreases when the chain length increases. There seems to be a problem with the RMT method applied to the pentane molecule. A careful analysis of the results show that the numerical dihedral angle distribution corresponds to (3.77) but with a temperature significantly different from the target temperature. If longer chains are considered, this problem disappears and the RMT method results are of the same order of magnitude as the results from other methods (see Tables 3.8 and 3.9).

Table 3.9. Numerical results for the discrepancy (3.79) for the (ϕ_1, ϕ_3) distribution for $C_{12}H_{26}$ in the case $\beta = 1$ and $K = 100$. The computational cost is fixed to 10^7 force or energy evaluations.

Method	Parameters	Discrepancy
Importance sampling	-	0.102 (0.0436)
Rejection	-	1.0 (0.0)
MIS	-	0.493 (0.222)
HMC	$\tau = 10\Delta t,\ \Delta t = 0.02$	0.0207 (0.00730)
BRW (Euler-Maruyama)	$\Delta t = 0.023$	0.0312 (0.0102)
BRW (MALA)	$\Delta t = 0.023$	0.0610 (0.0201)
Langevin	$\Delta t = 0.025,\ \xi = 1$	0.0173 (0.00726)
NHC	$Q_1 = 0.15, Q_2 = 0.01,\ \Delta t = 0.0025$	0.0350 (0.00865)
RMT	$Q_1 = 5, Q_2 = 7.5,\ \Delta t = 0.0025$	0.0428 (0.0194)

We can also see that, for short chains, the biased Random-Walk (MALA) is more efficient than the NHC method. However, for chains of 9 and 12 particles, the NHC method is more efficient. The biased Random-Walk with the Euler-Maruyama algorithm always seems to be a little more efficient than the biased Random-Walk with the MALA algorithm.

3.4.5 Improvement of the convergence rates

Convergence rate improvements using several shorter realizations

We already mentionned that, instead of running a single long trajectory, it might be more efficient, for a given computational cost, to run several shorter trajectories. This can be done for methods of Type 2 to 4. For methods of Type 2 and 3, this strategy relies on the following numerical approximation. Assuming that the methods are ergodic, it follows

$$\mathbb{E}_x(A(q^{N_1})) \to \int_{\mathcal{M}} A(q)\, d\pi \tag{3.80}$$

when $N_1 \to +\infty$. In some cases, this convergence is exponentially fast. The term $\mathbb{E}_x(A(q^{N_1}))$ is the expectation of the realizations of the chain conditioned at starting from $x \in \mathcal{M}$. It can be approximated by N_2 independent realizations of the Markov chain. Each realization is labelled by an index $k \in \{1, \ldots, N_2\}$, and the associated sample path is $(q^{0,k}, \ldots, q^{N-1,k})$. Notice that, for all samples, $q^{0,k} = x$. An approximation of $\mathbb{E}_x(A(q^{N_1}))$ is then obtained as

$$\mathbb{E}_x(A(q^{N_1})) \simeq I_{N_2}^{N_1}(x) = \frac{1}{N_2} \sum_{k=1}^{N_2} A(q^{N_1,k}). \tag{3.81}$$

Notice that we expect the error between $I_{N_2}^{N_1}(x)$ and the space average $\int_{\mathcal{M}} A(q)\, d\pi$ to be of the form $C(x)\rho^{N_1} + C(x, N_1)N_2^{-1/2}$ for some $0 < \rho < 1$.

When a short trajectory of length N_1 is computed for N_2 realizations starting from a given initial point x, we can also consider the following approximation of the position space average

$$\int_{\mathcal{M}} A(q)\, d\pi \simeq \frac{1}{N_1} \sum_{m=0}^{N_1-1} I_{N_2}^{m}(x), \tag{3.82}$$

where the right hand side is the Cesaro average of (3.81).

The results are presented in Table 3.10 in the case of a Langevin sampling for the pentane molecule at $\beta = 1$. As can be seen, there is a slight improvement when generating several shorter

trajectories, provided these trajectories remain long enough. Note however that such an improvement is not always observed. But we emphasize that there is no degradation of the results either. This is an interesting point since it allows a straightforward parallelization of the method.

Table 3.10. Numerical results for the discrepancy (3.79) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$, using a Langevin method with $\xi = 1$ and $\Delta t = 0.02$. The discrepancy has been computed with all points appearing in (3.82) (that is all points of the N_2 trajectories of length N_1), with a computational cost fixed to 10^7 force or energy evaluations.

Number N_2 of realizations	Discrepancy
1	0.0157 (0.00393)
5	0.0117 (0.00388)
10	0.0132 (0.00210)
20	0.0149 (0.00701)
50	0.0120 (0.00330)
100	0.0112 (0.00263)
200	0.0130 (0.00419)
500	0.0308 (0.00834)
1000	0.0528 (0.00740)

Convergence rate improvements at fixed computational cost, using an appropriate initial distribution

Another improvement is as follows. Instead of considering a fixed initial point, we can make a first approximation of the canonical distribution. Let us denote by π^{N_3} the following approximation of π:

$$\pi^{N_3} = \frac{1}{N_3} \sum_{i=1}^{N_3} \delta_{x^i}.$$

For each initial point x^i $(1 \leq i \leq N_3)$, an approximation (3.82) can be computed, for N_2 realizations of the Markov chain with trajectories of length N_1. The total number of points generated in this way is therefore $N_1 N_2 N_3$. The important issue is then to optimize the choices of N_1, N_2 and N_3 in order to have the best accuracy for a given total cost.

For the method to be efficient, the empirical measure π^{N_3} has to be a good approximation of π. To this end, the points x^i are chosen as follows. We first generate N^{tot} points $(y^1, \ldots, y^{N^{\text{tot}}})$ with weights $(w_1, \ldots, w_{N^{\text{tot}}})$, using (say) an Importance sampling method. We then generate N_3 points from this list with replacement with probabilities $\left(\frac{w_1}{W}, \ldots, \frac{w_{N^{\text{tot}}}}{W}\right)$ where $W = \sum_{i=1}^{N^{\text{tot}}} w_i$, and run one or several trajectories for each starting point. This can improve the rate of convergence of some methods. An example is the biased Random Walk at $\beta = 1$ with $\Delta t = 0.028$ for 10^6 operations. We consider $N^{\text{tot}} = 10^4$, $N_3 = 99$, $N_1 = 10^4$ and $N_2 = 1$. The discrepancy is lowered from 0.104 (0.0446) (with $N_1 = 10^6$, $N_2 = 1$ and $N_3 = 1$, see Table 3.7) to 0.0430 (0.0144). In general, it is observed that convergence occurs faster when starting from an approximate distribution.

Effect of undersampling

As a final improvement, we can test the influence of a systematic undersampling, which consists in picking only some of the points generated instead of considering all of them. Indeed, some techniques generate points $(q^0, \ldots, q^{N-1})$ that may be very much correlated, and it can happen that the sequence $(q^0, q^r, \ldots, q^{sr})$, the undersampling rate r being such that $N - 1 = rs$, is better distributed than the original sequence.

The results are presented in Table 3.11 in the case of a Langevin sampling for pentane at $\beta = 1$. As can be seen, the efficiency of the method remains stable when undersampling the data. This is particularly interesting when computing autocorrelation functions or time-dependent integrals of the form (3.2) since a NVE trajectory has to be computed for each starting point generated from the canonical distribution.

Of course, it is still possible to try to improve the quality of a single realization by filtering out the corresponding sequence of configurations, as is done for NVE simulations in [48, 49], but we will not detail this strategy any further.

Table 3.11. Numerical results for the discrepancy (3.79) for the pentane (ϕ_1, ϕ_2) distribution in the case $\beta = 1$ and $K = 100$, using a Langevin method with $\xi = 1$ and $\Delta t = 0.02$. The computational cost is fixed to 10^6 force or energy evaluations.

Undersampling rate	Discrepancy
1	0.0339 (0.0142)
5	0.0369 (0.0121)
10	0.0350 (0.00996)
50	0.0391 (0.0194)
100	0.0385 (0.0169)
500	0.0343 (0.0102)
1000	0.0539 (0.0173)

3.4.6 Computation of correlation functions

We present, as a final application, the computation of some correlation function, namely the transition rate from the set $\mathcal{A} = \{q \in \mathcal{M} \; ; \; |\phi_1| \geq 1, |\phi_2| \geq 1\}$ (both dihedral angles are not in their ground states) to the set $\mathcal{B} = \{q \in \mathcal{M} \; ; \; |\phi_1| \leq 1, |\phi_2| \leq 1\}$ (both dihedral angles are in their ground states). This transition rate is expressed as

$$C(t) = \frac{\langle \mathbf{1}_{\mathcal{A}}(q^0) \; \mathbf{1}_{\mathcal{B}}(\Pi_1 \Phi_t(q,p)) \rangle}{\langle \mathbf{1}_{\mathcal{A}}(q^0) \rangle}. \tag{3.83}$$

We proceed as follows. We first sample $M = 10^4$ initial conditions according to the canonical measure $d\mu$ (at $\beta = 1$) using 10^6 force evaluations and the parameters given in Table 3.7 (*i.e.* in all cases except for the HMC algorithm, we undersample at rate 100 a single trajectory that always starts from the same equilibrium position; the HMC trajectory is undersampled at rate 10 only since $\tau = 10\Delta t$). We then integrate the Newton equations of motion from each initial condition using the velocity Verlet scheme (3.17), for a time $t = 100$ (with $\Delta t = 0.005$). This procedure is repeated 100 times. The results are presented in Figure 3.2, and are compared with a reference result obtained starting from 10^6 initial conditions sampled with a rejection method.

As can be seen from the results, the methods yielding large discrepancies (such as Nosé-Hoover and BRW) predict a correlation $C(t)$ quite different from the reference result. On the other hand, the HMC and Langevin methods give much better results, especially HMC.

3.5 Stochastic boundary conditions

The vast majority of molecular dynamics simulations use periodic boundary conditions to simulate bulk conditions (see Section 2.2.1). When averages at fixed temperature are computed, Newton's equation of motion (associated with constant energy simulations) are modified so that

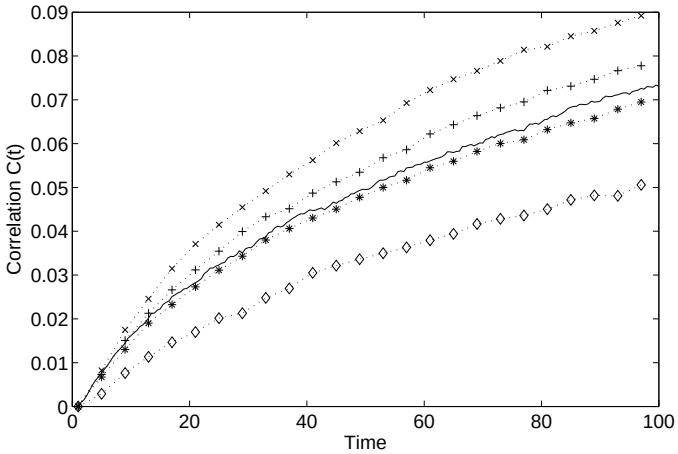


Fig. 3.2. Plot of the correlation function $C(t)$ starting from initial conditions generated with the rejection method (solid line), BRW/EM (x), Langevin/BBK (+), HMC (*) and Nosé-Hoover chain ($\diamond$).

the resulting dynamics is (hopefully) ergodic with respect to the canonical measure. Examples of such modifications are the Nosé-Hoover or the Langevin dynamics (see respectively Section 3.3 and 3.2.4). However, the quantities to compute may be time-dependent quantities, such as correlation functions:

$$\langle B \rangle(t) = \int_{T^*\mathcal{M}} B(\Phi_t(q,p),(q,p))\, d\mu,$$

where μ is the canonical measure and Φ_t the flow of the dynamics. It is not clear which dynamics should be used in this definition. It turns out that the results depend in general of the specifities of the chosen dynamics. For instance, the response of the system to an increased thermostat temperature depends on the parameters chosen for the Nosé-Hoover dynamics [113].

The system under study is usually a small system which should be embedded in a much larger microcanonical system. The larger system acts as an energy reservoir which ensures that the temperature is correct (this is actually the usual derivation of the canonical ensemble [61]). Some ways to obtain such a coupling between the simulated subsystem and the ideal energy reservoir (which should not be explicitely simulated, due to its size), present through some mean action, have been proposed. Section 3.5.1 reviews the most important ones (to our knowledge). In Section 3.5.2, a very simple model of stochastic boundary conditions (already used in [82], but only roughly described) is presented precisely: the core region of the simulated system is governed by NVE dynamics, while the parts of the system close to the boundary follow a Langevin dynamics with random perturbations decreasing as the distance to the boundary increases. In this way, a seamless coupling can be achieved.

3.5.1 Review of some classical stochastic boundary conditions

The first steady-state nonequilibrium molecular dynamics simulations were performed in the 70s by Ashurst and Hoover (see e.g. [12]). Their model uses perturbations limited to the boundary of the system (external force field or thermal fluctuations). This idea of partitioning the system between inner region (governed by Newton's equation of motion) and outer region (the surface of the system, or some small region around the surface), where the effects of the environment are taken into account, has been widely used. It is possible to propose a somehow arbitrary classification of stochastic boundary conditions:

- thermal boundary conditions;
- mechanical boundary conditions;

– mixed thermal and mechanical boundary conditions;
– "grand-canonical" boundary conditions to model system whose number of particles may vary.

Let us also notice that some directions of the system can still be modelled using periodic boundary conditions, while the remaining ones are treated with stochatic boundary conditions.

Thermal boundary conditions

The methods presented in this section take into account the thermal fluctuations of a system through its exchanges with its environment. These exchanges can be modelled

– by constraining the kinetic temperature in the regions close to the boundaries;
– by using "thermal walls", which lead, mathematically speaking, to jump processes (perturbations of the momenta of the impacting particles);
– by using a Langevin dynamics for the region of the system close to the boundary, and the usual Hamiltonian dynamics elsewhere, so that the resulting process is a diffusive process, which is (hopefully, but not trivially) hypoelliptic.

Velocity renormalization

In the first studies [12], the kinetic temperature in the regions close to the boundaries was kept fixed. This was done by velocity rescaling. Some refinings were proposed (see e.g. [27, 133]), rescaling only some components of the velocities (in one direction, typically), or by including the renormalization step directly in the equations of motion. This method is not used anymore nowadays.

Thermal walls

Following a work of Lebowitz and Spohn [201], Ciccotti and Tenenbaum introduce thermal walls modelling the contact of impacting particles with a heat reservoir [67]. The system has free boundary conditions, but when a particle leaves the simulation domain, another one enters at the same place where the leaving particle went out, with a momentum generated from the probability distribution $C^{-1}(e \cdot p) f_T(p) \mathbf{1}_{e \cdot p > 0}$, where e is the local normal vector, f_T the distribution of the momenta at equilibrium at the temperature T (maxwellian distribution) and C is a normalization constant. Therefore, the momenta of the entering particles are not drawn according to a maxwellian distribution of momenta. A numerical study for an ideal gas or a hard sphere gas confirms that the model of [67, 201] is indeed the right strategy [339].

The first simulations relying on thermal walls [67, 340] with different temperatures on both sides of the system have shown that dynamical properties could be computed, but that surface effects were important near the thermal walls (especially the local density and the temperature). This is why such a strategy asks for additional mechanical boundary conditions (see Section 3.5.1) to limit surface effects.

Coupling with a Langevin dynamics

One of the first simulation coupling a Hamiltonian and Langevin dynamics is due to Adelman and Doll [1]. The aim of this coupling was to reduce the number of degrees of freedom in the simulation by replacing the environing particles by some mean action, modelled by a random forcing term and a friction with memory (in the Mori-Zwanzig way). The first study were only a part of the system is governed by a Langevin dynamics, whereas the remaining part obeys Hamiltonian dynamics was proposed by Berkowitz et MacCammon [28], with a mechanical forcing to confine the system (some slices of a crystalline lattice at rest). To reduce surface effects, the idea of coupling Langevin and Hamiltonian dynamics was refined by Brooks et Karplus [43, 45], using especially some averaged confining force. Some studies also mention the use of a Langevin dynamics with a friction depending on the distance to the boundary of the system [82]. Similar ideas were

used in the framework of Nosé-Hoover dynamics [165, 209]; a seamless coupling is however less clear (Nosé masses depending on the distance to the boundary should be considered). These ideas were developed in the field of biology and the reference textbooks for condensed matter molecular dynamics (such as [113]) do not mention it.

Mechanical boundary conditions

Free boundary conditions and some thermal boundary conditions (such as thermal walls) may create surface effects (local density variations, or temperature differences). Periodic boundary conditions are a convenient way to reduce surface effects, though numerical studies [223], and then theoretical studies [273, 274], have shown that periodic boundary conditions also have spurious effects, especially for small systems. More importantly, PBC are problematic when long-range interactions are considered - such as coulombic forces for non-neutral systems (charged defects in solids) or solvant effects (dipole corrections) for biological systems. As an alternative to PBC to confine free boundary systems, one may consider

- forces or constraints arising from short-ranged interactions;
- mean-force effects arising from avarges over a large number of (non-simulated) degrees of freedom.

The second approach was developed in the field of biology. For example, in [192], the system is split into three regions, a core region (Hamiltonian dynamics and averaged electrostatic potential), a buffer region (thermal fluctuations through some Langevin dynamics, forces on the boundaries and averaged electrostatic potential), and an outer region (not explicitely simulated) which determines the averaged electrostatic potential. Such a modelling is refined in [181].

The first approach, more used for mechanical studies of solids, can be implemented in several ways. For instance, a given (macroscopic) displacement can be modelled by layers of surface atoms following rigidly the displacement, and kept fixed for the simulation [68, section II.2.C].

"Grand-canonical" boundary conditions

There are two general strategies to deal with systems whose number of particles varies:

- consider that the system is open and specify a flux of ingoing particles to compensate particle losses;
- use grand-canonical sampling techniques.

The first approach is used in [123] for a model case of non-interacting particles, in which case particle fluxes can be derived. The extension to interacting particles requires additional forcing terms on the boundaries, as well as density-dependent ingoing particle fluxes.

The second approach was presented in [182], for a model system of ionic channel, and refined in [372] to deal with protein solvatation. In a buffer region aroung the boundary, particles are inserted and deleted according to the local chemical potential, using standard grand-canonical sampling techniques [113]. Therefore, the number of particles is preserved in average, and the core region is not perturbed.

3.5.2 An example of thermal boundary conditions

We present more precisely in this section a seamless coupling between a Langevin and a Hamiltonian dynamics (in the spirit of [28, 43, 45, 82]), with periodic boundary conditions. The aim of this coupled model is therefore only to provide interesting thermal boundary conditions, so that time-dependent observables can be computed by averages performed in the core region of the system.

Description of the model

We consider a simulation box $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) with periodic boundary conditions (the configuration space therefore has the geometry of a torus). The simulation box Ω is decomposed into two non-overlapping domains Ω_i and Ω_e (see Figure 3.3), the outer region Ω_e being for example the set

$$\Omega_e = \{x \in \Omega \mid \mathrm{d}(x, \partial\Omega) < r_c\},$$

where $\mathrm{d}(x, \partial\Omega)$ is the distance from $x \in \Omega$ to the boundary $\partial\Omega$, and r_c some positive cut-off radius.

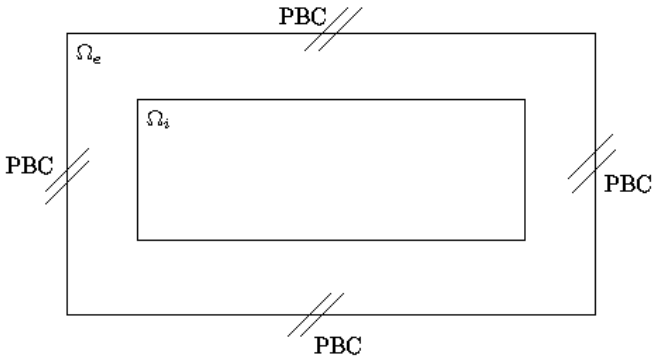


Fig. 3.3. Decomposition of the simulation box Ω into two non-overlapping domains Ω_i and Ω_e.

The dynamics we propose is as follows. The particles that are located in Ω_i are only subjected to the forces that derive from the interaction potential V, whereas the particles that are located in Ω_e also experience some random forcing. More precisely, we consider the dynamics

$$\begin{cases} dq_t &= M^{-1} p_t \, dt, \\ dp_t &= -\nabla V(q_t)\, dt - \Gamma(q_t) M^{-1} p_t \, dt + \Sigma(q_t)\, dW_t, \end{cases} \tag{3.84}$$

where $(W_t)_{t \geq 0}$ is a dN-dimensional Wiener process, and where the matrices Σ and Γ represent the magnitude of the fluctuations and of the dissipation respectively. They are linked by the fluctuation-dissipation relation:

$$\Sigma(q_t)\Sigma(q_t)^T = \frac{2}{\beta}\Gamma(q_t). \tag{3.85}$$

In this expression, $\beta = (k_{\mathrm{B}}T)^{-1}$ is the inverse temperature of the bath. In the sequel, we choose a diagonal matrix for $\Gamma(q)$:

$$\Gamma(q) = \mathrm{Diag}(\gamma(q_1), \ldots, \gamma(q_N)),$$

where the function γ is taken to be a smooth decreasing function of $\mathrm{d}(x, \partial\Omega)$ such that $\gamma(x) = 0$ in Ω_i and $\gamma(x) > 0$ in Ω_e. We also consider

$$\Sigma(q) = \mathrm{Diag}(\sigma(q_1), \ldots, \sigma(q_N)), \qquad \text{with } \sigma(\cdot) = \sqrt{\frac{2\gamma(\cdot)}{\beta}}. \tag{3.86}$$

It is easy to check that the canonical probability measure (3.3) is an invariant probability measure for (3.84) since it is a stationary solution of the associated Fokker-Planck equation.

It is not clear whether the stochastic differential equation (3.84) is ergodic since $\Sigma = 0$ in Ω_i. However, in the following numerical simulations, it is observed that, whatever the starting distribution, the correct kinetic temperature is quickly attained.

In the numerical examples presented in Section 3.5.2 and 3.5.2, we have used the following numerical implementation of (3.84), inspired from the classical BBK scheme used to integrate the Langevin equation [45]:

$$\begin{cases} p_i^{n+1/2} &= p_i^n + \dfrac{\Delta t}{2}\left(-\nabla_{q_i} V(q^n) - \dfrac{\gamma(q_i^n)}{m_i} p_i^n + \dfrac{\sigma(q_i^n)}{\sqrt{\Delta t}} G_i^n\right) \\ q_i^{n+1} &= q_i^n + \dfrac{\Delta t}{m_i} p_i^{n+1/2} \\ p_i^{n+1} &= p_i^{n+1/2} + \dfrac{\Delta t}{2}\left(-\nabla_{q_i} V(q^{n+1}) - \dfrac{\gamma(q_i^{n+1})}{m_i} p_i^{n+1} + \dfrac{\sigma(q_i^{n+1})}{\sqrt{\Delta t}} G_i^{n+1}\right) \end{cases} \tag{3.87}$$

where σ is still given by (3.86), and $\{G_i^n\}_{1\le i\le N,\, n\in\mathbb{N}}$ are identical and independently distributed (i.i.d.) standard gaussian random variables.

Thermal conductivity of Lennard-Jones systems

We first describe the Lennard-Jones system and the thermalization procedure we have considered. The NVE-NVT heating and cooling processes are then dealt with in Section 3.5.2, and alternative approaches to determine the thermal conductivity are briefly reviewed. Some simulation results are finally provided.

Description of the system

We consider a three-dimensional ($d = 3$) Lennard-Jones system, with standard periodic boundary conditions. The potential energy is given by

$$V(q) = \sum_{1\le i<j\le N} V_{\mathrm{LJ}}(|q_i - q_j|) + \frac{1}{2}\sum_{i,j=1}^{N} \sum_{k\in\mathcal{R}\setminus\{0\}} V_{\mathrm{LJ}}(|q_i - q_j + k|), \tag{3.88}$$

where $\mathcal{R}$ is the Bravais lattice and V_{LJ} the usual Lennard-Jones potential

$$V_{\mathrm{LJ}}(r) = 4\epsilon\left(\left(\frac{a}{r}\right)^{12} - \left(\frac{a}{r}\right)^{6}\right) , \tag{3.89}$$

with $\epsilon > 0$ and $a > 0$.

The system is first thermalized at an inverse temperature β using a *full* Langevin dynamics (that is, $\Gamma(q) = \gamma_0 \mathrm{I}_{3N}$ in (3.84)) for a time t_{init} large enough, starting from an equilibrium position such as a FCC lattice for solid state simulations, or a square lattice for liquid phase simulations,[5] and generating the momenta of the particles from the kinetic part of the canonical measure.

Computation of the thermal conductivity

The thermal conductivity λ of a system can be computed either at equilibrium, using a Green-Kubo formula [113], or in a non-equilibrium setting. The former method relies on the integration of the heat flux correlation function, and often requires long simulation times for the time integral to converge. Non-equilibrium molecular dynamics (NEMD) approaches assume a linear response regime, so that the heat flux depends linearly on the temperature gradient. To specify this linear relation, external fictitious mechanical forces can be added [100, 128] to the NVE dynamics, or a temperature gradient can be specified, while the heat flux is then measured. Since these methods also suffer from slow convergence, a different approach has been proposed, where the heat flux is specified, and the temperature field is measured [251].

[5] This initial configuration is much less stable than a FCC lattice, and thermalization is therefore expected to occur faster.

A recent interesting alternative method [175] relies on transient simulations. A small fraction of the system is instantaneously heated, and the kinetic temperature relaxation is monitored. The thermal conductivity can then be computed by comparison with the Fourier law. However, the approach of [175] is based on NVE simulations of relatively small systems, so that complete relaxation toward the canonical ensemble cannot be observed.

We now show that the NVE-NVT model (3.84) is fairly suited for thermal conductivity computations. Let us consider a Lennard-Jones system modeled by (3.84) initially at thermal equilibrium with temperature T_1 (such an equilibrium state is obtained as described in Section 3.5.2) and let us suddenly change the temperature of the thermostat to T_2. The inner system Ω_i is then heated or cooled down through energy exchanges with Ω_e, itself thermostated by the environing heat-bath, and the kinetic temperature of Ω_i as a function of time can be monitored. To reduce statistical errors, several independent relaxations must be performed, starting from initial configurations sampled independently from the canonical measure.

The thermal conductivity can then be recovered as follows. Assuming that the Fourier law holds in the domain $\Omega_i =]0, L[^3$, the local temperature obeys the heat equation

$$\rho C_v \partial_t T = \lambda \Delta T,$$

where ρ denotes the density of the system (expressed in mol/m^3), C_v the specific heat capacity (in J/K/mol), and λ the thermal conductivity (in W/m/K). For variations in a small temperature range, it can indeed be assumed that C_v and λ remain constant in space and time. The specific heat capacity can be found in thermodynamic tables, or computed as a time-independent canonical average according to

$$C_v = \frac{\mathcal{N}_a}{N k_B T^2}(\langle H^2 \rangle - \langle H \rangle^2),$$

where $\mathcal{N}_a$ is the Avogadro number and $\langle \cdot \rangle$ denotes a canonical average.

Setting $\sigma = \dfrac{\lambda}{\rho C_v}$, it follows

$$\partial_t T = \sigma \Delta T.$$

Consider the heating or cooling of the sytem from T_1 to $T_2 = T_1 + \delta T$ with $|\delta T| \ll T_1, T_2$. Setting $u = (T_2 - T)/\delta T$, the evolution of u is governed by the Cauchy problem

$$\left\{ \begin{array}{rcll} \partial_t u & = & \sigma \, \Delta u & \text{in } \Omega_i, \\ u|_{t=0} & = & 1 & \text{in } \Omega_i, \\ u & = & 0 & \text{on } \partial\Omega_i. \end{array} \right. \tag{3.90}$$

The initial condition u_0 can be expanded on the Fourier modes

$$\phi_{klm}(x, y, z) = \left(\frac{2}{L}\right)^{3/2} \sin\left(\frac{k\pi x}{L}\right) \sin\left(\frac{l\pi y}{L}\right) \sin\left(\frac{m\pi z}{L}\right)$$

as

$$u_0(x, y, z) = \frac{16\sqrt{2}L^{3/2}}{\pi^3} \sum_{k,l,m \geq 0} \frac{1}{(2k+1)(2l+1)(2m+1)} \phi_{2k+1,2l+1,2m+1}(x, y, z).$$

Let us denote by

$$h(t, x) = \sum_{k \geq 0} \frac{1}{(2k+1)} \exp\left(-\sigma \frac{(2k+1)^2 \pi^2}{L^2} t\right) \sin\left(\frac{(2k+1)\pi x}{L}\right).$$

Since $\Delta \phi_{klm} = -\dfrac{(k^2 + l^2 + m^2)\pi^2}{L^2} \phi_{klm}$, it follows,

$$u(t,x,y,z) = \frac{64}{\pi^3} h(t,x)\, h(t,y)\, h(t,z).$$

The deviation to the target temperature T_2 is therefore, on average on the domain Ω_i,

$$\bar{u}(t) = \frac{1}{L^3} \int_{]0,L[^3} u(t,x,y,z)\; dx\, dy\, dz = \frac{512}{\pi^6} k(t)^3,$$

where, setting $A = \sigma \pi^2 L^{-2}$,

$$k(t) = \sum_{k \geq 0} \frac{1}{(2k+1)^2} \exp\left(-\sigma \frac{(2k+1)^2 \pi^2}{L^2} t\right) = \mathrm{e}^{-At} \left(1 + \frac{1}{9} \mathrm{e}^{-8At} + \frac{1}{25} \mathrm{e}^{-24At} + \dots\right). \quad (3.91)$$

It then holds

$$\frac{\bar{u}(t)}{\bar{u}(t_0)} = \left(\frac{k(t)}{k(t_0)}\right)^3 \sim \mathrm{e}^{-3A(t-t_0)}$$

for $t \geq t_0$ and t_0 large enough. Therefore, the value of A (and thus of λ provided C_v is known) can be computed by fitting $\bar{u}(t)/\bar{u}(t_0)$ to an exponential function.

Numerical results

The kinetic temperature for a given number N_i of particles is defined as

$$T_{\mathrm{kin}} = \frac{2}{3 N_i k_{\mathrm{B}}} \sum_{n=1}^{N_i} \frac{p_n^2}{2 m_n}.$$

We also define, in analogy with the previous section, $u_{\mathrm{kin}} = (T_2 - T_{\mathrm{kin}})/\delta T$.

Figure 3.4 shows a plot of the instantaneous kinetic temperature in Ω_i in the case of a heating process for fluid Argon from T_1 to T_2, and the corresponding plot of $\bar{u}_{\mathrm{kin}}/\bar{u}_{\mathrm{kin}}(t_0)$ (with $t_0 = 5$ ps), averaged over 30 realizations of the heating process conducted from independent initial conditions. The parameters of the model are $N = 64,000$, $\epsilon/k_{\mathrm{B}} = 119.8$ K, $a = 3.405 \times 10^{-10}$ m, $T_1 = 400$ K, $T_2 = 420$ K, $\Delta t = 2.5 \times 10^{-15}$ s. We use a truncated Lennard-Jones potential with a cut-off radius $r_{\mathrm{c}} = 2.5\, a$. The molar mass is $M = 39.95 \times 10^{-3}$ kg/mol, and the density is $\rho = 35044$ mol/m^3. The simulation cell Ω is then a cubic box of edge length $L = 37.51\, a$. The parameters used for the thermalization are $\gamma_0/m = 10^{12}$ s^{-1} and $t_{\mathrm{init}} = 20$ ps. Then, the independent initial configurations are obtained from this thermalized configuration by running an additional Langevin dynamics for 15 ps before each realization of the heating process.

For the coupled NVE-NVT dynamics, we have used

$$\gamma(\cdot) = \gamma_1 \cos\left(\frac{\pi \cdot}{2 r_{\mathrm{c}}}\right) \quad (3.92)$$

with $\gamma_1/m = 5 \times 10^{12}$ s^{-1}. We have checked that the thermal response is not sensitive to the specific shape of the friction function nor to the value of γ_1 in a broad range.

As can be seen from Figure 3.4 (Left), the kinetic temperature in the inner region of the system converges toward the target value determined by the temperature of the thermostat. The function $\bar{u}_{\mathrm{kin}}/\bar{u}_{\mathrm{kin}}(t_0)$ is plotted on the time interval $[t_0, t_1]$ with $t_0 = 5$ ps and $t_1 = 75$ ps. Notice that, as we discard the initial relaxation, the higher order exponential terms in (3.91) can be neglected, so that we can indeed approximate $\bar{u}_{\mathrm{kin}}/\bar{u}_{\mathrm{kin}}(t_0)$ by $\mathrm{e}^{-3A(t-t_0)}$. A least-square fit gives $A = 0.01438$ s^{-1}. A numerical computation of C_v at $T = 400$ K (using a Langevin NVT sampling with 6×10^5 time-step as described in [51]) gives $C_v = 18.01$ J/K/mol, in good

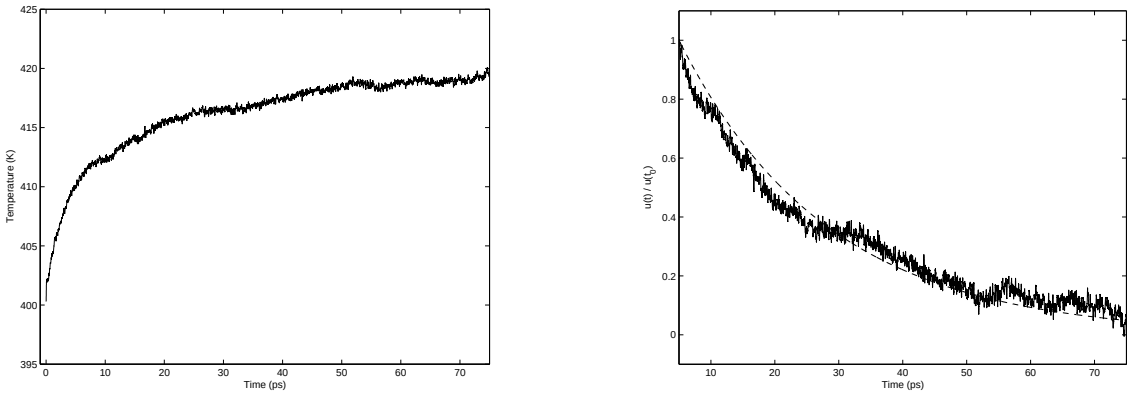


Fig. 3.4. Left: Kinetic temperature in Ω_i as a function of time. Right: Plot of $\bar{u}_{\mathrm{kin}}/\bar{u}_{\mathrm{kin}}(t_0)$ as a function of time with $t_0 = 5$ ps (solid line), as well as its exponential fitting function (dashed line). Notice that the exponential approximation seems to be justified.

agreement with the experimental value $C_v = 18.12$ J/K/mol[6]. Therefore, the computed value of λ is $\lambda = 0.1509$ W/m/K, which is in good agreement with the experimental value $\lambda = 0.1557$ W/m/K at $T = 400$ K.

Thermal relaxation of a displacement cascade in Pu

We finally present in this section some simulation results on the irradiation induced displacement cascades in metallic crystals. When an atom of a crystal ('the primary knock-on atom', PKA) undergoes a nuclear reaction or is hit by a high-energy particle, its kinetic energy is dramatically increased. This will give rise to a cascade of collisions between the neighboring atoms, together with a sudden increase of the local kinetic temperature. These cascades result in the production of numerous defects in the lattice (such as interstitial atoms or vacancies), the so-called 'primary damage state'. A large fraction of the defects quickly disappear due to the recombination between interstitial atoms and vacancies, while the system returns to its original temperature (the kinetic energy in excess is dissipated). This first stage of relaxation lasts about a nanosecond. An experimental investigation of these phenomena is difficult, since the time and length scales involved are too small for a direct observation, but it can be simulated by MD. The remaining defects created by the various cascade relaxations will then interact on much larger time scales (from a second to several years) to form clusters of defects, that will alter the macroscopic mechanical behavior of the material. This is the source of the ageing of radioactive and irradiated materials. Kinetic Monte-Carlo (KMC) models [77] are necessary to deal with such long time scales; these models can be parametrized by the results of MD simulations of the first stage of the cascade relaxation.

Our purpose is to model the thermalization occuring in this first stage. It is important to describe correctly this process, since it has an influence on the distribution of the remaining defects, hence on the parametrization of the KMC model. More specifically, we focus on the example of a FCC Pu crystal (recall that Pu undergoes alpha decay). Since the PKA is launched with a large kinetic energy, the kinetic temperature of the system increases at the beginning of the simulation. Therefore, unless the system is infinitely large (in which case the temperature increase is negligible, and the initial energy excess concentrated in the center of the crystal diffuses over the whole system), there is a need for some dissipation, in order to ensure thermal relaxation. The MD model of [77] considers a crystal with PBC, where the atoms in the unit cells close to the boundary obey a full Langevin dynamics, while the other atoms experience a pure NVE

[6] The experimental values used in this section are taken from the NIST Chemistry Webbook, `http://webbook.nist.gov/chemistry/fluid/`

dynamic. We propose here to consider a Langevin forcing of decreasing magnitude as explained in Section 3.5.2. This can heuristically account for the finite size of the crystal, dissipation being then understood as energy transfer from the simulated box to the rest of the crystal.

Simulations have been carried out for a FCC Pu lattice of 13,500 atoms at $T_0 = 300$ K, using a MEAM potential [21, 22, 24] for Pu [23]. An initial thermalization is performed for a time $t_0 = 10$ ps, using a full Langevin dynamics. The PKA is then launched with an energy of 100 eV in the direction $\langle 5\ 1\ 3 \rangle$. The first stage of the simulation is performed during the time $t_1 = 4$ ps with the time step $\Delta t = 5.10^{-5}$ ps. The second part is performed during the time $t_2 = 35$ ps. The friction function used in this simulation is still given by (3.92), with $\gamma_0/m = 2 \times 10^{12}$ s^{-1} and $r_{\text{cut}} = 4.5 \times 10^{-10}$ m (this is the cut-off range used for the MEAM potential). The evolutions of the kinetic energy of the whole system as a function of the iteration step are displayed in Figure 3.5 for both simulation stages.

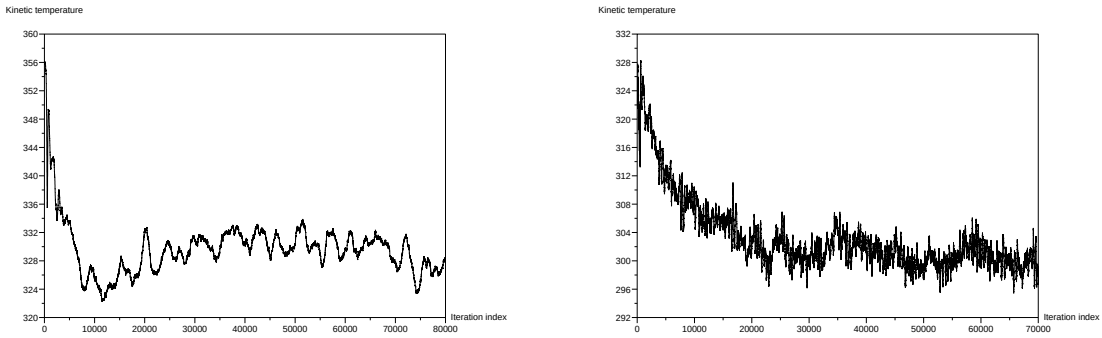


Fig. 3.5. Kinetic temperature as a function of the iteration step for a FCC Pu system experiencing a self-decay-induced cascade of 100 eV. The time-step is $\Delta t = 5 \times 10^{-5}$ ps for the picture on the left (first stage of the simulation), and $\Delta t = 5 \times 10^{-4}$ ps for the picture on the right (second stage of the simulation).

At the end of the second stage of the simulation, the kinetic temperature of the system has returned to the desired value $T = T_0$.

3.6 Some background on continuous state-space Markov chains and processes

3.6.1 Some background on continuous state-space Markov chains

This section is intented to give a quick overview of the most important notions and results for continuous state-space Markov chains. We refer the interested reader to [240], and to [127, Chapter 4] for a simple short introduction to continuous state-space Markov chains. The article [349] is also a beautiful introduction to the topic, making remarkable parallels between the countable case and the continuous state-space case.

Different levels of stability for Markov chains.

We first present in an informal manner the spirit of the characterization of stability for Markov chains $\{\Phi_n\}_{n\in\mathbb{N}}$ on a general state space X (in particular, we do not restrict ourselves to countable spaces). This general introduction is strongly inspired from [240, Section 1.3]. A useful concept is the first hitting time from a point to a set. Define

$$\tau_B = \inf \{n \geq 1 \mid \Phi_n \in B\},$$

the first time when the chain reaches the set B. The weakest form of stability is that the space accessible to the chain does not dramatically change when taking another initial condition, so that all "reasonably sized" sets can be reached from any starting point. This is the concept of *ϕ-irreducibility*, which can be stated as follows, for $x \in X$,

$$\phi(B) > 0 \Rightarrow \mathbb{P}_x(\tau_B < \infty) > 0,$$

where $\mathbb{P}_x$ is the probability induced by the Markov chain starting at x (*i.e.* the probability of events conditional on the chain starting from x). The measure ϕ precises the class of sets that can be "reasonably" reached.

A strengthening of this condition is that not only all sets can be reached, but in fact they are attained almost surely, in the sense that

$$\forall x \in X, \quad \phi(B) > 0 \Rightarrow \mathbb{P}_x(\tau_B < \infty) = 1.$$

This can be further strenghtened by requiring the expected hitting time to be finite:

$$\phi(B) > 0 \Rightarrow \mathbb{E}_x(\tau_B) < \infty,$$

where $\mathbb{E}_x$ is the expectation under $\mathbb{P}_x$. This level of stability is refered to as *recurrence*. Heuristically, it ensures that the chain does not drift, but returns often enough to "central" parts of the space. This kind of behaviour already implies some convenient behaviour along sample paths $(\Phi_0, \Phi_1, \dots)$, leading to a Law of Large Numbers (LLN).

The last level of stability is relevant for recurrent chains, and deals with convergence to a limiting regime independently of the initial condition. This is known as *ergodicity*, and is linked to the convergence of the distribution of the chain. In this case, Central Limit Theorems (CLT) can be stated to precise the behaviour along one sample path.

The different levels of stability introduced are summarized in Figure 3.6, together with conditions ensuring them. Denoting by $\mathcal{B}(X)$ the Borel σ-algebra of X and by μ^{Leb} the Lebesgue measure on X, these conditions read

$$\text{(C1)} \quad \forall x \in X, \quad \forall B \in \mathcal{B}(X), \quad \mu^{\text{Leb}}(B) > 0 \Rightarrow P(x, B) > 0,$$

$$\text{(C2)} \quad \pi \text{ is an invariant probability measure,}$$

$$\text{(C3)} \quad \begin{array}{c} \text{There exist measurable functions } L \geq \min\{1, A\},\ W \geq 0, \text{ a real number } b \\ \text{and a petite set } C \text{ such that} \\ \displaystyle\int_X P(x, dy) W(y) - W(x) \leq -L(x) + b\mathbf{1}_C(x), \quad \pi(W^2) < +\infty. \end{array}$$

$$\text{(C4)} \quad \begin{array}{c} \text{There exist a measurable function } W \geq 1, \text{ real numbers } c > 0 \text{ and } b, \\ \text{and a petite set } C \text{ such that} \\ \Delta W(x) \leq -cW(x) + b\mathbf{1}_C. \end{array}$$

The notion of petite set C will be precised below. Notice that Conditions (C1) and (C2) are usually quite easy to show in a MD setting, already giving ergodicity (without convergence rate however). Conditions (C3) and (C4) can be easily shown when the state space X is compact (when it is a d-dimensional torus for example, as in MD with periodic boundary conditions), under certain regularity conditions on the transition kernel.

These concepts are precised below, and presented in a more rigorous way. We end this section with a simple example, the Random Walk on a (half-)line, in order to see the theory of general state-space Markov chains at work.

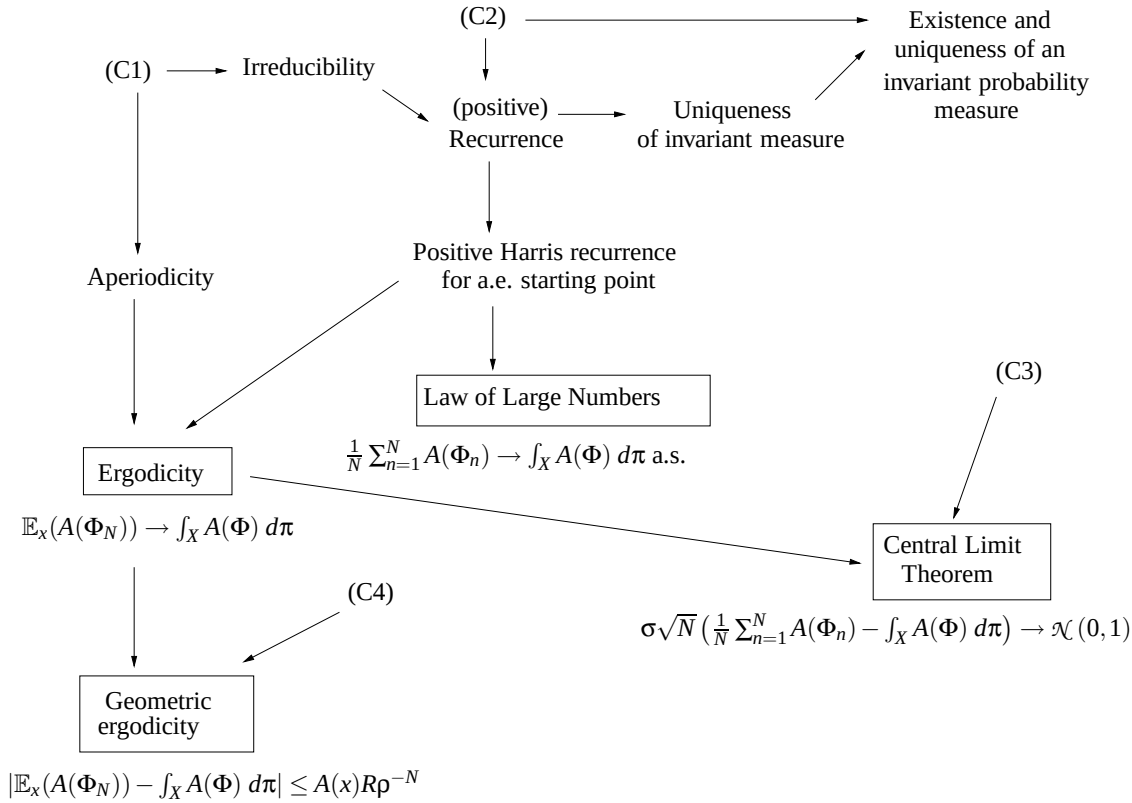


Fig. 3.6. The different levels of stability for Markov chains.

Some fundamental results.

We first precise the probability structure induced by a Markov chain on the state space. We consider a continuous state-space Markov chain given by its transition probability kernel

$$P = \{P(x, B),\ x \in X,\ B \in \mathcal{B}(X)\}$$

where $\mathcal{B}(X)$ is the set of Borel sets of X. The transition probability kernel is such that $P(\cdot, B)$ is a non-negative measurable function on X for all $B \in \mathcal{B}(X)$, and $P(x, \cdot)$ is a probability measure on $\mathcal{B}(X)$ for all $x \in X$. Given a transition probability kernel, one can define a time-homogeneous Markov chain $\Phi = (\Phi_0, \Phi_1, \dots)$ with initial distribution μ. This chain is defined on $\Omega = \prod_{i=1}^{\infty} X_i$ (where each X_i is a copy of X), and is measurable with respect to the product σ-field $\mathcal{F} = \otimes_{i=1}^{\infty} \mathcal{B}(X_i)$. There exists a probability measure $\mathbb{P}_\mu$ on $\mathcal{F}$ such that, for any any $n \in \mathbb{N}$ and any measurable $B_i \in \mathcal{B}(X_i)$ $(1 \le i \le n)$,

$$P_\mu(\Phi_0 \in B_0, \dots, \Phi_n \in B_n) = \int_{y_0 \in B_0} \dots \int_{y_{n-1} \in B_{n-1}} \mu(dy_0) P(y_0, dy_1) \dots P(y_{n-1}, B_n).$$

If an event occurs $\mathbb{P}_x = \mathbb{P}_{\delta_x}$-a.s. for all $x \in X$, we say that it occurs $\mathbb{P}_*$-a.s. We also inductively define P^n, the n-step transition probability by $P^0(x, B) = \delta_x(B)$ and the induction rule

$$P^n(x, B) = \int_X P(x, dy) P^{n-1}(y, B).$$

We then successively turn to the three important notions presented in the introduction of this section.

Irreducibility.

Definition 3.2. *The chain Φ is said to be* ϕ-irreducible *if there exists a measure ϕ on $\mathcal{B}(X)$ such that, for all $x \in X$ and $B \in \mathcal{B}(X)$ such that $\phi(B) > 0$, there exists some n (possibly depending on x and B) such that $P^n(x, B) > 0$.*

Notice that the Condition (C1) above implies μ^{Leb}-irreducibility. When a chain is ϕ-irreducible, there exists a maximal irreducibility measure ψ (see [240, Theorem 4.2.2]). The maximality is to be understood with respect to the domination relation for two measures, denoted as $\phi \prec \psi$, and defined through $\psi(B) = 0 \Rightarrow \phi(B) = 0$. Any other irreducibility measure is absolutely continuous with respect to ψ. The equivalence of maximal irreducibility measures allows then to define $\mathcal{B}^+(X) = \{B \in \mathcal{B}(X) \mid \psi(B) > 0\}$.

Definition 3.3. *A set B is full if $\psi(B^c) = 0$ and absorbing if $P(x, B) = 1$ for all $x \in B$.*

Recurrence.

As in the countable case, irreducible continuous state-space chains have essentially two possible behaviours: they may drift to infinity (transient behaviour) or remain almost always in a bounded region of space (recurrence). The occupation time η_B is defined as the number of visits of Φ to a set $B \in \mathcal{B}(X)$:

$$\eta_B = \sum_{n=1}^{\infty} \mathbf{1}_{\{\Phi_n \in B\}}.$$

Recall that $\mathbb{E}_x$ denotes the expectation under $\mathbb{P}_x = \mathbb{P}_{\delta_x}$, that is, the expectation under the probability generated by the chain starting from x.

Definition 3.4. *A chain Φ is called* recurrent *if it is ψ-irreducible and $\mathbb{E}_x(\eta_B) = \sum_{n=1}^{\infty} P^n(x, B) = +\infty$ for all $x \in B$ and $B \in \mathcal{B}^+(X)$.*

Let us precise some criteria ensuring that a Markov chain is recurrent. A simple case is when an invariant probability measure exists for the system. Let us emphasize that the existence of a (non-normalized) invariant measure is not sufficient, since this measure may be non-normalizable (see an example below).

Definition 3.5. *A ψ-irreducible chain Φ is said to be positive if it admits an invariant probability measure π.*

It is heuristically clear in this case that the chain cannot be transient. The following proposition holds:

Proposition 3.2 ([240], Proposition 10.1 and Theorem 10.4.9). *If a chain Φ is positive then it is recurrent and admits a unique invariant probability measure equivalent to ψ.*

Notice that Conditions (C1) and (C2) above imply positive recurrence for the chain. When no invariant probability measure is known, stronger conditions are needed to get recurrence, such as drift criteria [240, Chapter 8]. In statistical physics however, it is often the case that an invariant probability measure is known.

Law of Large Numbers.

The concept of recurrence can (and has to) be somewhat strengthened to get convergence results such as the Law of Large Numbers (LLN).

Definition 3.6. *A set $B \in \mathcal{B}(X)$ is called* Harris recurrent *if $\mathbb{P}_x(\eta_B = \infty) = 1$ for all $x \in B$. A set B is called* maximal Harris *if it is a maximal absorbing set such that Φ restricted to B is Harris recurrent. A chain Φ is called* Harris recurrent *if it is ψ-irreducible and if every set in $\mathcal{B}^+(X)$ is Harris recurrent. A Harris recurrent and positive chain Φ is called a* positive Harris *chain.*

Actually, any recurrent chain is already almost a Harris recurrent chain. Indeed, the following theorem holds:

Theorem 3.13 ([240], Theorem 9.1.5). *If Φ is recurrent, then $X = H \cup N$ where H is a non-empty maximal Harris set, and N is ψ-null.*

Therefore, starting from an initial value $x \in H$, a positive chain remains in H and is positive Harris on H. This amounts to replacing the whole space X by its full subset H. Note that π is also an invariant measure for the chain on H.

We now turn to the convergence of the average along one sample path. Consider the sum $S_N(A) = \sum_{i=1}^N A(\Phi_n)$. We recall a Law of Large Numbers (LLN) result:

Theorem 3.14 ([240], Theorem 17.1.7). *Suppose Φ is positive Harris. Then, for any measurable function $A \in L^1(\pi)$,*

$$\lim_{n\to\infty} \frac{1}{N} S_N(A) = \int_X A\, d\pi \quad \text{a.s. } [\mathbb{P}_*].$$

Remark 3.3. *Therefore, since the chain starting from H remains in H and is positive Harris on H, the LLN holds true for any chain $\{\Phi_n\}_{n\in\mathbb{N}}$ starting from $\Phi_0 = x \in H$. Therefore, it holds for a.e. starting point, H being a subset of full measure by Theorem 3.13. This result can actually be extended to all starting points [239, 241]. It holds whenever Conditions (C1) and (C2) are verified.*

Small sets and petite sets

The following definitions of small and petite sets are used for the convenience of other definitions and are particularly well-suited for general proofs in the Markov chain setting. However, they will not be used as such in this chapter, for we will be able to work with compact sets, that are small or petite under certain regularity conditions on the Markov transition kernel. We also warn the reader that the terms 'small' and 'petite' do not refer to the size of the spaces involved. They merely refer to some useful uniform lower bounds on the transition kernel.

Definition 3.7. *A set $C \in \mathcal{B}(X)$ is called a ν_m-*small set *if there exist $m > 0$ and a non-trivial measure ν_m such that for all $x \in C$ and $B \in \mathcal{B}(X)$,*

$$P^m(x, B) \geq \nu_m(B).$$

Though it is far from obvious from this definition, any ψ-irreducible chain has small sets $C \subset B$ for any $B \in \mathcal{B}(X)^+$ (see [240, Theorem 5.2.2]). In fact, the whole space X can be recovered by a countable union of small sets (see [240, Proposition 5.2.4]). This allows many properties of continuous state space Markov chains to be stated in the same manner as for countable state space Markov chains.

The notion of small sets is generalized with the notion of *petite sets*. Setting $K_a(x, B) = \sum_{n=0}^\infty P^n(x, B)a(n)$ for $x \in X, B \in \mathcal{B}(X)$ and with $a = \{a(n)\}_{n\in\mathbb{N}}$ a probability measure on $\mathbb{N}$, the expression K_a defines a transition kernel.

Definition 3.8. *Let ν_a be a non-trivial measure on $\mathcal{B}(X)$. A set $C \in \mathcal{B}(X)$ is ν_a-petite if*

$$K_a(x, B) \geq \nu_a(B)$$

for all $x \in C$ and all $B \in \mathcal{B}(X)$.

Notice that a ν_m-small set is ν_{δ_m}-petite. We will now see that compact sets are petite, under certain regularity conditions on the transition kernel.

Definition 3.9. *If $x \mapsto P(x, \mathcal{O})$ is a lower semi-continuous function for any open set $\mathcal{O} \in \mathcal{B}(X)$, then the chain is said to be weak Feller.*

Notice that the lower semi-continuity condition is usually easy to check in practice. It will even often be the case that $P(\cdot, B)$ is a continuous function for any Borel set B. We then have the following

Theorem 3.15. *If the ψ-irreducible chain Φ is weak Feller and if* supp ψ *has a non-empty interior, then all compact subsets of X are petite.*

Ergodicity.

We first introduce the total variation norm for a signed Borel measure μ. It is given by

$$||\mu|| = \sup_{h \text{ measurable, } |h| \leq 1} |\mu(h)| = \sup_{\{A \in \mathcal{B}(X)\}} \mu(A) - \inf_{\{A \in \mathcal{B}(X)\}} \mu(A).$$

Notice that convergence in total variation implies weak convergence.

Definition 3.10. *A chain Φ is* ergodic *when*

$$\forall x \in X, \quad \lim_{n \to \infty} ||P^n(x, \cdot) - \pi|| = 0.$$

In particular, ergodicity implies $\mathbb{E}_x(A(\Phi_n)) \to \int_X A(\Phi)\, d\pi$ when $n \to +\infty$ for any bounded measurable function A.

Ergodicity is actually quite easy to get once the chain has been shown to be recurrent. It is sufficient to show that the chain is aperiodic. We need here the notion of small and petite sets to state the definition of aperiodicity, though in practice much simpler criteria will be used. We introduce the set E_C associated with a ν_M small set C:

$$E_C = \{n \geq 1 \mid \text{the set } C \text{ is } \nu_n\text{-small with } \nu_n = \kappa_n \nu_M \text{ for some } \kappa_n > 0\}.$$

We see that $M \in E_C$. Let us denote by d the greatest common divisor of the set E_C. In fact d is independent of the initial small set chosen. Therefore, the following definition makes sense:

Definition 3.11. *Suppose that Φ is a ψ-irreducible Markov chain. If $d = 1$, the chain is called* aperiodic. *If there exists a ν_1-small set C with $\nu_1(C) > 0$, the chain is called* strongly aperiodic.

It is often easy to check strong aperiodicity in the MD setting using some global accessibility results. In particular, Condition (C1) implies aperiodicity (see [240, Theorem 5.4.4]). The following theorem then states the ergodicity of recurrent aperiodic chains.

Theorem 3.16 ([240], Theorem 13.3.4). *If Φ is positive recurrent and aperiodic, then for every initial distribution λ such that $\lambda(N) = 0$ (where N is the π-null set defined in Theorem 3.13),*

$$\left|\left| \int \lambda(dx) P^n(x, \cdot) - \pi \right|\right| \to 0 \quad \text{as } n \to \infty.$$

In particular, the case $\lambda = \delta_x$ can be considered for a.e. point x (*i.e.* for $x \in H$). This result holds as soon as conditions (C1) and (C2) are verified.

The convergence in total variation norm implies convergence of the expectations for bounded observables A. It is therefore not sufficient in practice for non-bounded observables A (see for instance the examples presented in the Introduction). Fortunately, the ergodicity results can be

strengthened in a straightforward way. For a given measurable non-negative function W, let us define the W-total variation norm for a signed Borel measure μ as

$$||\mu||_W = \sup_{h \text{ measurable, } |h| \leq W} |\mu(h)|.$$

Then Theorem 3.16 can be readily extended to integrable functions A.

Theorem 3.17 ([240], Theorem 14.0.1). *Suppose that $A \geq 1$ is measurable and $\pi(|A|) < +\infty$. If Φ is positive recurrent and aperiodic, then for π-a.e. $x \in X$,*

$$\left|\left| \int \lambda(dx) P^n(x, \cdot) - \pi \right|\right|_A \to 0 \quad \text{as } n \to \infty.$$

Rate of convergence for the LLN: a Central Limit Theorem.

Additional conditions are required to get not only a LLN, but a CLT, precising the rate of convergence of a sample path average toward its limit. The drift ΔW is defined, for $x \in X$, as

$$\Delta W(x) = \int_X P(x, dy) W(y) - W(x).$$

We then consider the following

Criterion 3.1. *Assume Φ is ergodic, and there exist a measurable function $L : X \to [1, \infty[$, a petite set $C \in \mathcal{B}(X)$, $b < +\infty$ and a finite-valued measurable function W such that*

$$\Delta W(x) \leq -L(x) + b\mathbf{1}_C(x), \quad \forall x \in X.$$

Denoting by π the invariant measure of the chain, we also assume $\pi(W^2) < \infty$.

Heuristically, this drift condition ensures that ΔW is decreasing outside a petite set C (in practice, a compact set). Therefore, we expect the chain to spend most of its time in the set C. The dynamics of the chain is then almost that of a chain in a compact set. That is why we can expect some stronger recurrence properties and some better convergence results.

For a given measurable function A such that $\pi(|A|) < \infty$, we formally define the function $\hat{A}$ by the following Poisson equation:

$$\hat{A} - P\hat{A} = A - \pi(A).$$

It is not clear in general whether $\hat{A}$ is well-defined. This turns out to be the case when Criterion 3.1 is verified, and allows to state a CLT (see [240, Theorem 17.5.3]):

Theorem 3.18 (CLT). *Assume Criterion 3.1 holds, and let A be a function such that $|A| \leq L$. Then the constant $\gamma_A^2 := \pi(\hat{A}^2 - (P\hat{A})^2)$ is well-defined, non-negative and finite. If $\gamma_A^2 > 0$, then, defining $\bar{A} = A - \pi(A)$, it holds*

$$(n\gamma_A^2)^{-1/2} S_n(\bar{A}) \to \mathcal{N}(0, 1),$$

this convergence being in law.

Notice that we get convergence results only for observables $|A| \leq L$, while the LLN applies for any integrable function. Theorem 3.18 holds true as soon as Conditions (C1), (C2) and (C3) are verified.

Remark 3.4. *In particular, under the assumptions of Theorem 3.15, the whole state space is petite when it is compact. Therefore, Condition (C3) is straightforwardly verified with the choice $C = X$ and W and L arbitrary smooth functions (taking b large enough).*

Geometric ergodicity.

The ergodicity property implies the convergence $\mathbb{E}_x(A(\Phi_n)) \to \int_X A(q)\,d\pi$ for measurable integrable functions A. A convergence rate can be obtained by resorting to the stronger notion of geometric ergodicity, generalizing the notion of ergodicity. The following Criterion, analogous to the drift condition for Criterion 3.1, is of paramount importance.

Criterion 3.2. *There exist a function $W \geq 1$ finite at some $x_0 \in X$, a petite set $C \in \mathcal{B}(X)$, and $b < +\infty$, $c > 0$ such that*

$$\Delta W(x) \leq -cW(x) + b\mathbf{1}_C(x), \quad \forall x \in X. \tag{3.93}$$

This drift criterion can be heuristically interpreted in the same way as Criterion 3.1. We then get the following

Theorem 3.19 ([240], Theorem 15.0.1). *Assume Criterion 3.2 holds. Then there exist $\rho < 1$ and $R < +\infty$ such that, for all $x \in \{y \in X \mid W(y) < +\infty\}$,*

$$||P^n(x,\cdot) - \pi||_W \leq RW(x)\rho^n.$$

In particular, we get

$$\left|\mathbb{E}_x(A(\Phi_n)) - \int_X A(\Phi)\,d\pi\right| \leq RW(x)\rho^n$$

for any starting point $x \in X$ such that $W(x) < +\infty$. This result holds as soon as Conditions (C1), (C2) and (C4) are verified.

Remark 3.5. *When X is compact, Condition (C4) is straightforwardly verified with the choice $C = X$ for any arbitrary smooth function W (taking b large enough). When X is not bounded and the chain is weak Feller (with an irreducibily measure of non-empty interior), Condition (C4) is satisfied when (3.93) holds for a compact set C and for a smooth function W such that $W(x) \to +\infty$ when $|x| \to +\infty$.*

A simple example: The Random-Walk on a (half-)line.

We now present a simple example, taken from [240]. We hope that it illustrates relevantly many of the notions introduced in this section. The setting is the following. Consider a collection of real-valued random variables $\Phi = \{\Phi_0, \Phi_1, \ldots\}$, defined as

$$\Phi_{k+1} = \Phi_k + W_{k+1},$$

where $\{W_k\}$ are independent and identically distributed (i.i.d.) random variables, that we do not precise further for the moment. The distribution of Φ_0 can be chosen arbitrarily. A convenient choice is for example to initialize the chain with a deterministic point $x_0 \in \mathbb{R}$, which amounts to considering the initial measure δ_{x_0}. The so-defined Markov chain is called a "random-walk" (RW).

We can also consider a random-walk on the half-line (RWHL), defined as

$$\Phi_{k+1} = [\Phi_k + W_{k+1}]_+ ,$$

where $[a]_+ = \max(a, 0)$. We examine successively to the questions of irreducibility, recurrence and ergodicity for those two Markov chains.

Irreducibility.

Under reasonable assumptions on the increments $\{W_k\}$, irreducibility is easy to check, and asks only for little comprehension of the behaviour of the system.

Consider first the case of random-walk when the W_k have values in $\mathbb{Q}$ and are such that $\mathbb{P}(W_k = x) > 0$ for all $x \in \mathbb{Q}$. Starting then from $x_0 \in \mathbb{Q}$, it is easily seen that $\mathbb{Q}$ is absorbing. If the chain was irreducible, any irreducibility measure ϕ would be supported by $\mathbb{Q}$. For $x_0 \notin \mathbb{Q}$, the chain has values in $x_0 + \mathbb{Q}$. So, considering the chain starting from x_0, we see that $P^n(x_0, \mathbb{Q}) = 0$ for all $n \in \mathbb{N}$. This shows that ϕ cannot be an irreducibility measure. The chain is not irreducible in this case, and it has an uncountably infinite number of absorbing sets.

In the case when W_k has a smooth positive density γ, the chain is seen to be irreducible with respect to the Lebesgue measure μ^{Leb} (more general conditions could also be considered [240]). Indeed, for any $x \in \mathbb{R}$ and $B \in \mathcal{B}(\mathbb{R})$ such that $\mu^{\text{Leb}}(B) > 0$

$$P(x, B) = P(W_1 \in B - x) = \int_{B-x} \gamma(y)dy > 0.$$

In addition, there exists $\delta, \eta > 0$ such that $\gamma(x) \geq \delta > 0$ for $|x| \leq 2\eta$. Setting $C = \{|x| \leq \eta\}$, and considering $x \in C$ and $B \subset C$, one has

$$P(x, B) = P(W_1 \in B - x) = \int_{B-x} \gamma(y)dy \geq \delta\mu^{\text{Leb}}(B) > 0. \tag{3.94}$$

Setting for example $\phi = (\mu^{\text{Leb}}(C))^{-1}\mathbf{1}_C(\cdot)$, the relation (3.94) shows that C is a ϕ-small set.

For the random-walk on the half-line, we assume that $\mathbb{P}(W_1 < 0) > 0$. It is then straightforward to show that, for all $x \in \mathbb{R}_+$, there exists n such that $P^n(x, \{0\}) > 0$. This shows that δ_0 is an irreducibility measure for RWHL.

Recurrence

In the case of RWHL, it is intuitive that the chain will be recurrent when the mean displacement is negative. In the case when the mean displacement is positive, we expect on the contrary the chain to drift to infinity without coming back (except maybe a finite number of times in average). We now precise these heuristic arguments. Set $m = \int_{\mathbb{R}} x\gamma(x)dx$. When $m > 0$, Proposition 9.5.1 in [240] shows that the chain is transient (the proof uses a comparison with a convenient Markov chain on countable state-space). When $m < 0$, a drift criterion can be stated, ensuring recurrence of the chain (see [240], Section 8.5). Indeed, consider $x_* < 0$ such that $\int_{x_*}^{+\infty} x\gamma(x)dx \leq \frac{m}{2}$, and take $W(x) = x$. Then, for x in $[0, -x_*]$,

$$\Delta W(x) = \int_{\mathbb{R}} P(x, dy)(y - x) = \int_{y \geq 0} P(x, dy)(y - x) = \int_{y \geq 0} (y - x)\gamma(y - x)dy \leq \frac{m}{2} \leq 0.$$

This shows that a drift criterion holds with $C = [0, -x_*]$. Heuristically, this means that the values of W cannot grow too much, which implies that the chain remains in a vicinity of the origin. We resort to Theorem 8.0.2 in [240] to prove that the chain is recurrent. It then has a unique invariant measure (see below for conditions ensuring that this invariant measure is finite).

For the random-walk on the full line, it is still quite clear that non-zero mean increments will lead to a transient behaviour. Conditions for recurrence in the case when the mean increment is zero can be precised when the increments have bounded range. We refer to [240, Section 9.5]. However, the chain can never be positive recurrent since the Lebesgue measure is invariant (see [240, Section

10.5]), and is therefore at best null recurrent. Ergodicity does not make sense for the general RW model.

Ergodicity for the Random-Walk on the half-line

We still assume that the mean drift $m = \int_{\mathbb{R}} x\gamma(x)dx$ is negative in order to ensure recurrence of the chain, and the existence of an invariant measure. We need however a better drift criterion to ensure that the invariant measure is a probability measure (that is, a finite measure) and to get ergodicity. To this end, we assume in addition that $\int_0^{+\infty} \mathrm{e}^{st}\gamma(t)dt < +\infty$ for $0 < s \leq \eta$ for some $\eta > 0$. Notice that this can be interpreted as sufficient fast decrease in the increments. Then, for $0 < s < \eta$, and $L(x) = \mathrm{e}^{sx}$,

$$\frac{1}{s}\frac{\int_{\mathbb{R}} P(x,dy)(L(y)-L(x))}{L(x)} = \int_{\mathbb{R}} \gamma(x)\frac{\mathrm{e}^{sx}-1}{s}dx \to m$$

when $s \to 0$ by dominated convergence. There exists $0 < s_0 < \eta$ such that, setting $W(x) = \exp(s_0 x)$,

$$\Delta W(x) \leq \frac{m}{2}s_0 W(x) + b\mathbf{1}_C(x)$$

for some $b > 0$ and with $C = [0, c]$ for some $c > 0$ large enough (see [240, page 399] for precisions). The chain is therefore W-uniformly ergodic, in the sense that there exists $R > 0$ and $0 < r < 1$ such that

$$\forall x \in \mathbb{R}_+, \quad ||P^n(x,\cdot) - \pi||_W \leq RW(x)r^{-n}.$$

3.6.2 Some convergence results for Markov processes.

We extend here the results of Appendix 3.6.1, stated for Markov chains, to Markov processes. We will focus on diffusion equations of the form

$$d\Phi_t = b(\Phi_t)dt + \Sigma dW_t, \tag{3.95}$$

where Φ_t is a stochastic process with values in X, b is a C^∞ function, Σ is a matrix of dimension $d = \dim(X)$, and W_t is a d-dimensional standard Wiener process.

We assume that trajectorial existence and uniqueness hold true for (3.95). This is classical for globally Lipschitz drifts [152, Theorem III.3.2], namely for functions b satisfying for some positive constant D

$$\forall (x,y) \in X^2, \quad |b(x) - b(y)| \leq D\ |x-y|. \tag{3.96}$$

When this condition is not satisfied, it is possible to conclude to trajectorial existence and uniqueness under the following hypothesis (see [152, Theorem III.4.1]): there exist a C^2 function $W(x)$ that goes to infinity at infinity and a positive constant c such that

$$\mathcal{A}W \leq cW. \tag{3.97}$$

Besides, under assumption (3.96) or (3.97), one can prove that the Markov process (3.95) is Feller. That means that, for each bounded measurable function $g : X \to \mathbb{R}$, the mapping

$$x \mapsto \mathbb{E}_x(g(\Phi_t^x))$$

is continuous, where Φ_t^x is the solution of (3.95) with initial condition $\Phi_0^x = x$. We assume in the sequel that either (3.96) or (3.97) is satisfied. Some extensions for less smooth functions b and $\Sigma \equiv \Sigma(x)$ can be found in [328].

The transition kernel P^t is defined, for $t > 0$ and $B \in \mathcal{B}(X)$, as

$$P^t(x, B) = \mathbb{P}_x(\Phi_t \in B),$$

where $\mathbb{P}_x$ is the probability generated by the process starting at x. The infinitesimal generator $\mathcal{A}$ associated with (3.95) is

$$\mathcal{A}g(x) = b(x) \cdot \nabla g(x) + \frac{1}{2}[\Sigma\Sigma^T]_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}(x) \tag{3.98}$$

for $g \in \mathrm{C}^2(X)$.

Main convergence results.

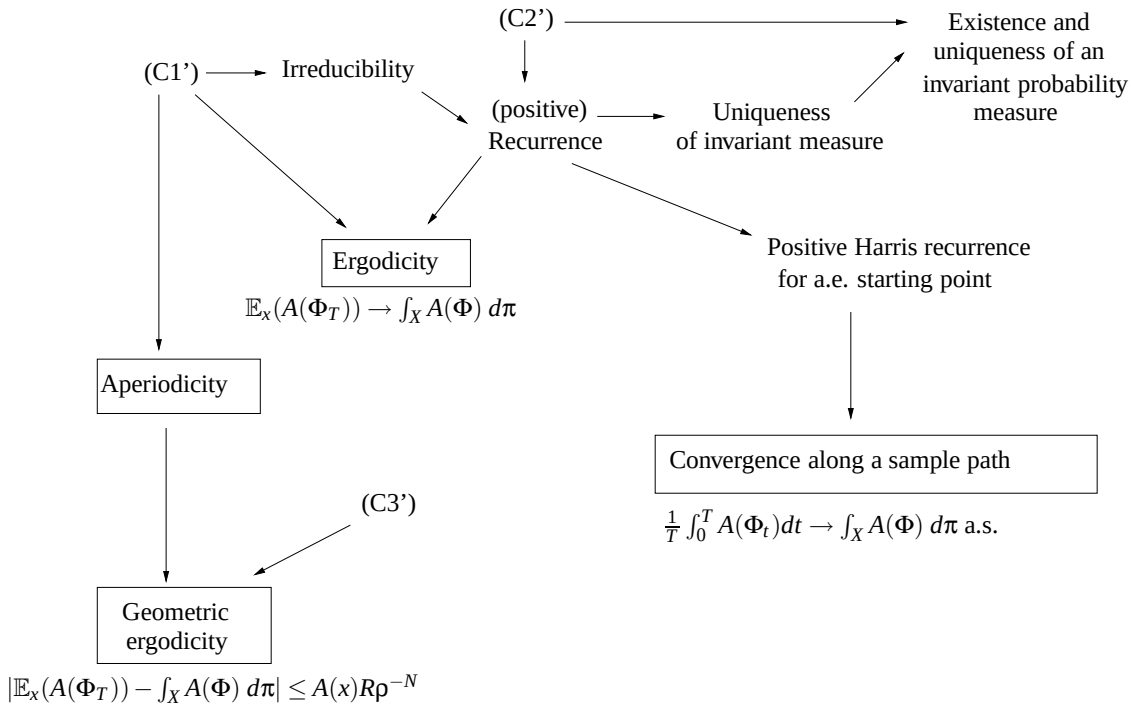


Fig. 3.7. The different levels of stability for Markov processes.

Figure 3.7 summarizes the main results, as in the discrete time case. The definitions of the different concepts and the proofs of the implications can be found in the remainder of this Section. Recall that we made the following general assumption throughout this Section

(C0') Condition (3.96) or (3.97) holds.

The conditions (C1'), (C2'), and (C3') read:

(C1') For all $q \in X$ and open set $\mathcal{O} \in \mathcal{B}(X)$, $P^t(q, \mathcal{O}) > 0$,

(C2') π is an invariant probability measure for the process,

(C3') There exist a measurable functions $W \ge 1$ going to infinity at infinity, real numbers $c > 0$, $b \in \mathbb{R}$ and a compact set C such that $\mathcal{A}W(x) \le -cW(x) + b\mathbf{1}_C$.

Notice that conditions (C1') and (C2') are usually quite easy to show in a MD setting, already giving ergodicity (without convergence rate however). Conditions (C3') can be easily shown when the state space X is compact (when it is a d-dimensional torus for example).

Stability concepts.

We first precise the concepts of irreducibility, Harris recurrence and ergodicity in the continuous time setting, which are quite analogous to the corresponding discrete time concepts [86, 241]. Consider, for $B \in \mathcal{B}(X)$, the random variables

$$\tau_B = \inf\{t \geq 0 \mid \Phi_t \in B\}, \qquad \eta_B = \int_0^{+\infty} \mathbf{1}_{\{\Phi_t \in B\}}\, dt.$$

Definition 3.12. *A Markov process is said to be ϕ-irreducible if for a σ-finite measure ϕ,*

$$\forall x \in X, \forall B \in \mathcal{B}(X), \quad \phi(B) > 0 \Rightarrow \mathbb{E}_x(\eta_B) > 0.$$

A process is Harris recurrent if, for a σ-finite measure ψ,

$$\forall x \in X, \forall B \in \mathcal{B}(X), \quad \psi(B) > 0 \Rightarrow \mathbb{P}_x(\tau_B < +\infty) = 1.$$

When a Harris recurrent process has a finite invariant measure (which can be normalized into a probability measure), it is called positive Harris recurrent.

Note also that a Harris recurrent process is irreducible.

Irreducibility can be checked in two steps. First, one can show *open set* irreducibility, which is usually easy to check using controllability arguments (see *e.g* [231, 336, 337]). We then get irreducibility using the continuity of the transition kernel (resulting from the Feller property).

When an invariant probability measure for the stochastic differential equation (3.95) exists, and when the process is irreducible, it is also recurrent, since there is also a dichotomy between recurrence and transience as in the discrete-time case [348, Theorem 2.3]. When Φ is recurrent, we also have existence of a maximal absorbing Harris set of full measure, and uniqueness of the invariant measure [348]. Therefore, the results of the discrete-time case can be completely transposed.

(Weak) Regularity of the transition kernel.

In contradiction with the Markov chain case, we often need some (weak) regularity properties on the transition kernel in the continuous-time setting. The minimal assumption that has to be made is that the process is a T-process.

Definition 3.13. *The Markov process is a T-process if there exists a probability measure a on $\mathbb{R}_+$ and a kernel T such that $T(\cdot, B)$ is lower semi-continuous for all $B \in \mathcal{B}(X)$ and*

$$K_a = \int_0^{+\infty} a(dt) P^t \geq T.$$

In particular, this property holds whenever the process is Feller since in this case, for all $t_0 > 0$ and all $B \in \mathcal{B}(X)$, $P^{t_0}(\cdot, B)$ is continuous.

Convergence of the average along one sample path.

The concepts introduced above allow us to state a result concerning the asymptotic behaviour of the average

$$S_T(A) = \frac{1}{T}\int_0^T A(\Phi_t)\ dt,$$

for some observable $A \in \mathrm{L}^1(\pi)$. Notice that this average is in fact a random variable.

Theorem 3.20 ([241], Theorem 8.1). *Suppose that Φ is a positive recurrent T-process. Then for any π-a.e. $x \in X$ and $A \in \mathrm{L}^1(\pi)$,*

$$S_T(A) \to \int_X A(q)\ d\pi \qquad \mathbb{P}_x - \text{a.s.}$$

Therefore, as in the discrete time case, we obtain convergence over a single sample path realization. Notice that this result can be extended to all starting points in X, and not only for starting points in the full maximal Harris subset [241]. Some results also exist for non-irreducible Markov process [241], but we restrict here to positive recurrent processes, which is the natural MD setting.

Central Limit Theorems can also be stated for the convergence of $S_T(A)$. However, the setting is not as clear as in the discrete time case. We refer for example to [172].

(Geometric) Ergodicity.

As for the discrete time case, convergence of the expectations $\mathbb{E}_x(A(\Phi_t))$ to the state space average $\int_X A(\Phi)\, d\pi$ can be stated under certain conditions. This is precisely the notion of ergodicity. As in Appendix 3.6.1, $||\cdot||$ denotes the total variation norm, and $||\cdot||_W$ the W-total variation norm.

Definition 3.14. *The Markov process is called ergodic if an invariant probability π exists and*

$$\forall x \in X, \quad ||P^t(x,\cdot) - \pi|| \to 0$$

when $t \to +\infty$.

The fact that the process is Harris recurrent and that some skeleton chain is irreducible is enough to ensure ergodicity. A skeleton chain is a Markov chain obtained by sampling the process at times $\Delta > 0$, and is thus the Markov chain with the associated transition kernel P^{Δ}.

Theorem 3.21 ([241], Theorem 6.1). *Suppose that Φ is positive Harris recurrent. Then Φ is ergodic if and only if some skeleton chain is irreducible.*

Notice that Condition (C1') immediately gives the irreducibility of the skeleton chain. Therefore, ergodicity holds whenever (C1') and (C2') are verified. This gives the convergence $\mathbb{E}_x(A(\Phi_t)) \to \int_X A(\Phi)\ d\pi$ for bounded measurable functions A.

A rate of convergence can also be obtained and extensions to non-bounded functions can be stated, as in the time-discrete case, using drift criteria. These criteria have to be checked on the generator $\mathcal{A}$ given by (3.98). We still need the process to be aperiodic. The definition of this notion for Markov processes is quite analogous to the corresponding discrete-time definition. We therefore refer to [86, 241] for more precisions, and simply note that the Feller property of the chain and (C1') are sufficient to conclude to aperiodicity. The definition of petite sets is also a straightforward extension of the discrete-time case, so we also refer to [86, 241] for example for a more formal definition. The following result shows that it is often enough to consider compact sets in applications.

Theorem 3.22 ([241], Theorem 4.1). *For a Harris recurrent T-process, every compact set is petite.*

We then have the following

Theorem 3.23 ([86], Theorem 5.2). *Consider a ψ-irreducible aperiodic Markov process, and assume there exist a measurable function $W \geq 1$ such that*

$$\mathcal{A}W \leq -cW + b\mathbf{1}_C \tag{3.99}$$

for $c > 0$, $b < +\infty$ and a petite set $C \in \mathcal{B}(X)$. Then the process is W-geometrically ergodic in the sense that there exist $R > 0$ and $0 < \rho < 1$ such that for every $t \geq 0$,

$$||P^t(x, \cdot) - \pi||_W \leq RW(x)\rho^t.$$

Together with conditions (C1') and (C2'), Condition (C3') then gives geometric ergodicity. As in the time-discrete case, Condition (C3') holds whenever the state space is compact. Another common situation is when the drift condition (3.99) is verified for some smooth W going to infinity at infinity and for some compact set C.

4

Computation of free energy differences

4.1 Nonequilibrium computation of free energy differences 120
4.1.1 The Jarzynski equality (The alchemical case) 120
4.1.2 The Jarzynski equality (The reaction coordinate case) 122
4.1.3 Practical computation of free energy differences 131
4.1.4 Numerical results 134
4.2 Equilibration of the nonequilibrium computation of free energy differences 138
4.2.1 The IPS and its statistical properties 139
4.2.2 Consistency through a mean-field limit 141
4.2.3 Numerical implementation 143
4.2.4 Applications of the IPS method 143
4.3 Path sampling techniques 148
4.3.1 The path ensemble with stochastic dynamics 150
4.3.2 Equilibrium sampling of the path ensemble 152
4.3.3 (Non)equilibrium sampling of the path ensemble 163
4.4 Adaptive computation of free energy differences 169
4.4.1 A general framework for adaptive methods 170
4.4.2 Rigorous convergence results for the Adaptive Biasing Force method .. 179

The free energy of a system is a quantity of paramount importance in statistical physics. It is defined as

$$F = -\frac{1}{\beta} \ln Z, \qquad Z = \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H}.$$

The constant Z is the partition function of the system, and the space $T^*\mathcal{M}$ is phase-space (see Section 2.2 for notations). In many applications, the quantity of interest is the free energy *difference* between an initial and a final state. These differences are related to transitions from an initial to a final state, and can be classified in two categories:

(i) the so-called alchemical case considers transitions indexed by an external parameter λ. The system is then governed by a Hamiltonian H_λ (or a potential V_λ), such as $H_\lambda(q,p) = (1-\lambda)H_0(q,p) + \lambda H_1(q,p)$. The corresponding free energy difference is

$$\Delta F = -\beta^{-1} \ln \left(\frac{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_1(q,p)}\, dq\, dp}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_0(q,p)}\, dq\, dp} \right),$$

(ii) in the reaction coordinate case, the transition is indexed through some level set function $\xi(q)$ indexing disjoint submanifolds of the configuration space, and

$$\Delta F = -\beta^{-1} \ln \left(\frac{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z_1}\, dq\, dp}{\displaystyle\int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H(q,p)}\, \delta_{\xi(q)-z_0}\, dq\, dp} \right).$$

Therefore, free energies can be expressed in both cases as

$$F = -\beta^{-1} \ln Z, \qquad Z = \int_{\Sigma} \exp(-\beta V)\, d\nu \tag{4.1}$$

where $\beta = 1/(k_{\mathrm{B}}T)$ (T denotes the temperature and k_{B} the Boltzmann constant). The Boltzmann-Gibbs measure $\exp(-\beta V)d\nu$ is defined for a reference positive measure $d\nu$, which has support Σ. We will consider here that Σ is a submanifold of $\mathbb{R}^{3N}$, but all the results extend to the case when Σ is a submanifold of $\mathbb{T}^{3N}$ (the $3N$-dimensional torus, which arises when using periodic boundary conditions). The statistics of the system are completely defined by (V, ν). We consider here that (V, ν) is labeled using a d-dimensional parameter z (with $d \ll 3N$) which characterizes the system at some coarser level. Examples of such parameters are $\xi(q)$ or λ with the above notations. In the alchemical case, the parameter $z = \lambda$ is independent of the current configuration of the system.

This chapter is organized as follows. In Section 4.1, we recall the usual Jarzynski equality when computing free-energy differences using nonequilibrium dynamics (stated for alchemical transitions), and present an extension to the reaction coordinate case. We then present, in Section 4.2, an equilibration of the nonequilibrium dynamics, which ensures that the sample is always canonically distributed even for fast switchings. In Section 4.3, we present a new algorithm for sampling paths governed by stochastic dynamics. Sampling paths can be useful to compute free energy differences, and in any cases, uses techniques reminiscent from free energy computation schemes. Finally, we present adaptive dynamics in Section 4.4, proposing a unified framework, new parallel implementations and a proof of convergence using entropy estimates in a specific case.

4.1 Nonequilibrium computation of free energy differences

4.1.1 The Jarzynski equality (The alchemical case)

Markovian nonequilibrium simulations

The usual way to achieve a nonequilibrium switching is to perform a time inhomogeneous irreducible Markovian dynamics

$$t \mapsto X_t, \quad X_0 \sim \mu_0, \tag{4.2}$$

for $t \in [0, T]$, and a smooth schedule $t \mapsto \lambda(t)$ verifying $\lambda(1) = 0$ and $\lambda(T) = 1$. The variable x can represent the whole degrees of freedom (q, p) of the system, or only the configuration part q. Depending on the context, the invariant measure μ will therefore be the canonical measure

$$d\mu_\lambda(q,p) = \frac{1}{Z_\lambda} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp, \tag{4.3}$$

with $Z_\lambda = \int_{T^*\mathcal{M}} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp$ or its marginal with respect to the momenta, which reads

$$d\tilde{\mu}_\lambda(q) = \frac{1}{\tilde{Z}_\lambda} \mathrm{e}^{-\beta V_\lambda(q)}\, dq,$$

with $\tilde{Z}_\lambda = \int_{\mathcal{M}} \mathrm{e}^{-\beta V_\lambda(q)}\, dq$. When we do not wish to precise further the dynamics, we simply call x the configuration of the system, $H_\lambda(x)$ its energy and $d\mu_\lambda(x)$ the invariant measure. The actual invariant measure should be clear from the context.

The dynamics is such that for a *fixed* $\lambda \in [0,1]$, the Boltzmann distribution $d\mu_\lambda$ is invariant. For example, the Langevin dynamics (3.47) or its overdamped limit (3.38) can be considered. In this last case, $X_t = q_t$ and the evolution of the system is given by

$$dq_t = -\nabla V(q_t)\, dt + \sigma\, dW_t,$$

with $\sigma^2 = 2/\beta$ and W_t a standard Wiener process.

Denoting by $p_{s,t}(x,y)dy = \mathbb{E}\left(X_t \in dy | X_s = x\right)$ the density kernel of the process, the evolution of the process law is characterized by the backward Kolmogorov equation (t and y being given):

$$\partial_s p_{s,t}(.,y) = -L_{\lambda(s)}\left(p_{s,t}(.,y)\right),$$

or its forward version (s and x being given):

$$\partial_t p_{s,t}(x,.) = L^*_{\lambda(t)}\left(p_{s,t}(x,.)\right).$$

The operator $L_{\lambda(t)}$ is called the infinitesimal generator of the dynamics, and $L^*_{\lambda(t)}$ is its dual. The invariance of $\mu_{\lambda(t)}$ under the instantaneous dynamic can then be expressed through the balance condition:

$$\forall \varphi, \quad \int L_{\lambda(t)}(\varphi) d\mu_{\lambda(t)} = 0. \tag{4.4}$$

When the schedule is sufficiently slow, the dynamics is said quasi-static, and the law of the process X_t is assumed to stay close to its local steady state throughout the transformation. This is out of reach at low temperature (more precisely, large deviation results [112] ensure that the typical escape time from metastable states grows exponentially fast with β, which compells quasi-static transformations to being exponentially slow with β). It is therefore interesting to consider approaches built on switched Markovian dynamics, but able to deal with reasonably fast transition schemes.

Importance weights of non equilibrium simulations.

For a given nonequilibrium run X_t we denote by

$$\mathcal{W}_t = \int_0^t \frac{\partial H_{\lambda(s)}}{\partial \lambda}(X_s)\lambda'(s)\, ds$$

the out of equilibrium virtual work induced on the system during the time interval $[0,t]$. The quantity $\mathcal{W}_t$ gives the importance weights of nonequilibrium simulations with respect to the target equilibrium distribution. Indeed, it was shown in [187] that

$$\boxed{\mathbb{E}(\mathrm{e}^{-\beta \mathcal{W}_t}) = \mathrm{e}^{-\beta(F(\lambda(t)) - F(0))}.} \tag{4.5}$$

This fluctuation equality is known as the Jarzynski's equality, and can be derived through a Feynman-Kac formula [177], as follows: consider the Feynman-Kac density kernel defined by

$$\int \varphi(y) p^w_{s,t}(x,y) dy = \mathbb{E}\left(\varphi(X_t)\mathrm{e}^{-\beta(\mathcal{W}_t - \mathcal{W}_s)} | X_s = x\right), \tag{4.6}$$

and characterized by the following extended backward Komogorov evolution:

$$\partial_s p^w_{s,t}(.,y) = -L_{\lambda(s)}\left(p^w_{s,t}(.,y)\right) + \beta\frac{\partial H_{\lambda(s)}}{\partial\lambda}\lambda'(s)p^w_{s,t}(.,y).$$

Using this identity and the balance equation (4.4) gives:

$$\partial_s \int p^w_{s,t}(x,y)\mathrm{e}^{-\beta H_{\lambda(s)}(x)}dx = 0$$

and thus after integration on $[0,t]$, we get the fundamental Feynamn-Kac fluctuation equality:

$$\frac{Z_t}{Z_0}\int \varphi\, d\mu_{\lambda(t)} = \mathbb{E}\left(\varphi(X_t)\mathrm{e}^{-\beta\mathcal{W}_t}\right). \tag{4.7}$$

Therefore, taking $\varphi = 1$, it follows

$$\mathbb{E}(\mathrm{e}^{-\beta\mathcal{W}_t}) = \mathrm{e}^{-\beta(F(\lambda(t))-F(0))},$$

and Jensen's inequality then gives

$$\mathbb{E}(\mathcal{W}_t) \geq F(\lambda(t)) - F(0).$$

This inequality is an equality if and only if the transformation is quasi-static on $[0,t]$; in this case the random variable $\mathcal{W}_t$ is actually constant and equal to ΔF. When the evolution is reversible, this means that equilibrium is maintained at all times.

As an improvement, we will see how to avoid the exponential importance weights of the nonequilibrium paths by a selection rule between replicas (see Section 4.3.3).

4.1.2 The Jarzynski equality (The reaction coordinate case)

Nonequilibrium computations of free energy differences in the reaction coordinate setting using stochastic dynamics could be performed using soft constraints to switch between the initial state centered on the submanifold $\{\xi(q) = z_0\}$ and the final state centered on $\{\xi(q) = z_1\}$. Steered molecular dynamics techniques use for example a penalty term $K(\xi(q) - z)^2$ in the energy of the system [267] (with K large) to 'softly' constraint the system to remain close to the submanifold $\{\xi(q) - z = 0\}$, and varying the value z from 0 to 1 in a finite time T. It is shown in [177] how to use such a biasing potential to exactly compute free energy differences (even for a finite K), which is of particular interest for experimental studies. From a computational viewpoint however, it is expected that large values of K require small integration time steps. Moreover, it is observed in practice that the statistical fluctuations increase with larger K (see [267]). Instead, we propose to replace the stiff constraining potential $K(\xi(q) - z)^2$ by a projection onto the submanifold $\{\xi(q) - z = 0\}$. This situation is reminiscent of the case of molecular constraints, that can be enforced using a stiff penalty term, or more elegantly and often more efficiently, using some projection of the dynamics involving Lagrange multipliers. This is the spirit of the well known SHAKE algorithm [295].

We present here a nonequilibrium stochastic dynamics and an equality that allow to compute free energy differences between states defined by different values of a reaction coordinate. The dynamics relies on a projection onto the current submanifold at each time step, and we use the Lagrange multipliers associated with this projection to estimate the free energy difference. More precisely, we use the difference between these Lagrange multipliers and the external forcing term required for the finite time switching (see for example the discretization (4.43)). The main results of this section are the Feynman-Kac equality of Theorem 4.1 (which extends the proof of [177] to hard constraints), as well as the associated discretizations (4.45) and (4.46).

We first present the equilibrium computation of free energy differences using projected stochastic differential equations, before turning to the extension to the non-equilibrium case.

Equilibrium computation of free energy differences in the reaction coordinate case

The aim of this section is to introduce the definitions of the free energy and the mean force in the reaction coordinate setting, and to recall how thermodynamic integration is used to compute free energy differences. The computation of the mean force is based on projected stochastic differential equations (SDE). The presentation is done for a one-dimensional reaction coordinate (the extension to the multi-dimensional case being postponed until the end of this section) and the dynamics used is an extension of the overdamped Langevin dynamics.

Free energy and mean force

The state of the system is characterized by the value of a reaction coordinate $\xi : \mathcal{M} \to [0,1]$. The function ξ is supposed to be smooth and such that $\nabla\xi(q) \neq 0$ for all $q \in \mathcal{M}$. For a given value $z \in [0,1]$, we denote by Σ_z the submanifold

$$\Sigma_z = \{ \, q \in \mathcal{M}, \, \xi(q) = z \, \} \tag{4.8}$$

and we assume that $\bigcup_{z\in[0,1]} \Sigma_z \subset \mathcal{M}$. For each point $q \in \Sigma_z$, we also introduce the orthogonal projection operator $P(q)$ onto the tangent space to Σ_z at point q defined by:

$$P(q) = \mathrm{Id} - \frac{\nabla\xi \otimes \nabla\xi}{|\nabla\xi|^2}(q), \tag{4.9}$$

where $\otimes$ denotes the tensor product. The orthogonal projection operator on the normal space to Σ_z at point q is defined by $P^\perp(q) = \mathrm{Id} - P(q)$.

The free energy is then defined as

$$F(z) = -\beta^{-1} \ln\left(Z_z\right), \tag{4.10}$$

with

$$Z_z = \int_{\Sigma_z} \exp(-\beta V) \, d\sigma_{\Sigma_z}, \tag{4.11}$$

where for any submanifold Σ of $\mathbb{R}^{3N}$, σ_Σ denotes the Lebesgue measure induced on Σ as a submanifold of $\mathbb{R}^{3N}$. The associated Boltzmann probability measure is

$$d\mu_{\Sigma_z} = Z_z^{-1} \exp(-\beta V) \, d\sigma_{\Sigma_z}. \tag{4.12}$$

Remark 4.1 (On the definition of the free energy). *Two comments are in order about formula (4.10). First, this formula is valid up to an additive constant, which is not important when considering free energy differences. Second, the potential V in (4.11) may be a potential different from the actual potential seen by the particles. More precisely, if the particles evolve in a potential V, the standard definition of the free energy in the physics and chemistry literature is (4.10) with*

$$Z_z = \int \exp(-\beta V) \, \delta_{\xi(q)-z},$$

where $\delta_{\xi(q)-z}$ is a measure supported by Σ_z and defined by: for all test functions ϕ,

$$\int \phi(q)\delta_{\xi(q)-z} = \int_{\Sigma_z} \phi |\nabla\xi|^{-1} \, d\sigma_{\Sigma_z}.$$

This amounts to considering (4.10)–(4.11) with V replaced by an effective potential $V + \beta^{-1} \ln|\nabla\xi|$ (see Remark 4.2 for the case of a multi-dimensional constraint). With this definition,

$$\int_{\mathcal{M}} A(\xi(q)) \mathrm{e}^{-\beta V(q)} \, dq = \int_{\mathcal{M}} A(z) \mathrm{e}^{-\beta F(z)} \, dz,$$

but the free energy differences $F(z_1) - F(z_2)$ depend on the choice of the reaction coordinate (and not only on the level sets Σ_z).

Since the results we present here hold irrespective of the physical signification of the potential V, we may assume without loss of mathematical generality that the free energy is indeed given by (4.10)–(4.11), and the choice of the definition of the free-energy is left to the user. Let us emphasize that, in practice, the cumbersome computation of the gradient of the additional term $\beta^{-1} \ln |\nabla \xi|$ in the modified potential (which intervenes in the projected SDEs we use, see (4.39)–(4.40) or (4.41)–(4.42)) can be avoided resorting to some finite differences, as explained in [66].

Using the co-area formula (see (4.33) and Proposition 4.3 for a proof in the multi-dimensional case), it is possible to derive the following expression of the derivative of the free energy F with respect to z (the so-called *mean force*) (see [83, 320]):

$$F'(z) = Z_z^{-1} \int_{\Sigma_z} \frac{\nabla \xi}{|\nabla \xi|^2} \cdot (\nabla V + \beta^{-1} H) \exp(-\beta V) d\sigma_{\Sigma_z}, \tag{4.13}$$

where

$$H = -\nabla \cdot \left(\frac{\nabla \xi}{|\nabla \xi|} \right) \frac{\nabla \xi}{|\nabla \xi|} \tag{4.14}$$

is *the mean curvature vector field* of the surface Σ_z. The free energy can thus be expressed as an average with respect to μ_{Σ_z}:

$$F'(z) = \int_{\Sigma_z} f(q) d\mu_{\Sigma_z}(q), \tag{4.15}$$

where f is the local mean force defined by:

$$f = \frac{\nabla \xi}{|\nabla \xi|^2} \cdot (\nabla V + \beta^{-1} H). \tag{4.16}$$

We explain next how it is possible to compute this average with respect to μ_{Σ_z}, without explicitly computing f, by using projected SDEs. This avoids in particular the computation of the mean curvature vector H which involves second-order derivatives of ξ.

The principle of thermodynamic integration is to recast the free energy difference

$$\Delta F(z) = F(z) - F(0) \tag{4.17}$$

between two reaction coordinates 0 and z as an integral over the mean force:

$$\Delta F(z) = \int_0^z F'(y) \, dy. \tag{4.18}$$

Therefore, in practice, thermodynamic integration computation of free-energy is as follows. First, the free energy difference $\Delta F(z)$ is estimated using quadrature formulae for the integral in (4.18), such as for example a Gauss-Lobatto scheme:

$$\Delta F(z) \simeq \sum_{i=0}^{K} \omega_i F'(y_i)$$

where the points $\{y_0, y_1, \ldots, y_K\}$ are in $[0, z]$ and $\{\omega_0, \omega_1, \ldots, \omega_K\}$ are their associated weights. Second, the derivatives $F'(y_i)$ are computed as canonical averages over the submanifolds Σ_{y_i}, using projected SDEs (see next section).

To obtain a free-energy profile (and not only a free-energy difference for a fixed final state), it is possible to approximate the function $\Delta F(z)$ on the interval $[0, 1]$ by a polynomial. This can

be done for example by interpolating the derivative F' by splines, and integrating the resulting function (consistently with the normalization $\Delta F(0) = 0$).

Projected stochastic differential equations

We now explain how to compute the mean force $F'(z)$ defined by (4.13) using projected SDEs, for a fixed parameter z. We consider the solution Q_t to the following SDE:

$$\begin{cases} Q_0 \in \Sigma_z, \\ dQ_t = -P(Q_t)\nabla V(Q_t)\,dt + \sqrt{2\beta^{-1}}P(Q_t) \circ dB_t, \end{cases} \tag{4.19}$$

where B_t is the standard $3N$-dimensional Brownian motion and $\circ$ denotes the Stratonovich product. It is possible (see [66]) to check that μ_{Σ_z} is an invariant probability measure associated with the SDE (4.19). Under suitable assumptions, which we assume in the rest of the section, on the potential V and the surface Σ_z, the process Q_t is ergodic with respect to μ_{Σ_z}. Moreover, the SDE (4.19) can be rewritten in the following way:

$$dQ_t = -\nabla V(Q_t)\,dt + \sqrt{2\beta^{-1}}dB_t + \nabla\xi(Q_t)d\Lambda_t, \tag{4.20}$$

where Λ_t is a real valued process, which can be interpreted as the Lagrange multiplier associated with the constraint $\xi(Q_t) = z$ (see the discretization in Section 4.1.3). This process can be decomposed into two parts:

$$d\Lambda_t = d\Lambda_t^{\mathrm{m}} + d\Lambda_t^{\mathrm{f}}. \tag{4.21}$$

The so-called martingale part Λ_t^{m} (whose fluctuation is of order $\sqrt{\Delta t}$ over a timestep Δt) is

$$d\Lambda_t^{\mathrm{m}} = -\sqrt{2\beta^{-1}}\frac{\nabla\xi}{|\nabla\xi|^2}(Q_t)\cdot dB_t, \tag{4.22}$$

where $\cdot$ implicitly denotes the Itô product. The so-called bounded variation part Λ_t^{f} (whose fluctuation is of order Δt over a timestep Δt) is

$$d\Lambda_t^{\mathrm{f}} = \frac{\nabla\xi}{|\nabla\xi|^2}(Q_t)\cdot\nabla V(Q_t)\,dt + \beta^{-1}\frac{\nabla\xi}{|\nabla\xi|^2}(Q_t)\cdot H(Q_t)\,dt = f(Q_t)\,dt, \tag{4.23}$$

f being the local mean force defined above by (4.16). Thus, since Q_t is ergodic with respect to μ_{Σ_z} the mean force can be obtained as a mean over the Lagrange multiplier Λ_t:

Proposition 4.1. *The mean force is given by:*

$$F'(z) = \lim_{T\to\infty}\frac{1}{T}\int_0^T d\Lambda_t = \lim_{T\to\infty}\frac{1}{T}\int_0^T d\Lambda_t^{\mathrm{f}}. \tag{4.24}$$

Notice that the martingale part $d\Lambda_t^{\mathrm{m}}$, which has the largest fluctuations, has zero mean. In order to reduce the variance, it is thus numerically convenient to perform the mean over the bounded variation part $d\Lambda_t^{\mathrm{f}}$ rather than over the whole Lagrange multiplier $d\Lambda_t$ (see Section 4.1.3).

We refer to [66] for a proof of Proposition 4.1, as well as for formulae involving higher dimensional reaction coordinates. Such ideas have been used for a long time in the framework of Hamiltonian dynamics (see [83, 320]).

The interest of Equation (4.24) is that the SDE (4.20) can be very naturally discretized as explained in Section 4.1.3 below. Then, the average over a discretized trajectory of the process Λ_t converges to $F'(z)$. This is particularly convenient for numerical purposes since it does not ask for explicitly computing the local force f. For further details, we refer to [66] and to Section 4.1.3. In the next section, we use these ideas for the computation of the free energy difference given through the Jarzynski equality.

Nonequilibrium stochastic methods in the reaction coordinate case

We wish here to extend the Feynman-Kac formula derived in [177] (see Section 4.1.1) for a parameter z which appears only in the potential V, to the reaction coordinate case, where z labels submanifolds Σ_z (defined by Equation (4.8)) of the state space. To this end, we need to make precise the evolution of the constraints.

We consider a $\mathcal{C}^1$ path $z : [0,T] \to [0,1]$ of values of the reaction coordinate ξ, with $z(0) = 0$, and $z(T) = 1$. Recall that the associated family of submanifolds of admissible configurations is denoted by

$$\Sigma_{z(t)} = \{q \in \mathcal{M},\, \xi(q) = z(t)\}\,,$$

and that the associated Boltzmann probability measures are

$$d\mu_{\Sigma_{z(t)}} = Z_{z(t)}^{-1} \exp(-\beta V) d\sigma_{\Sigma_{z(t)}}.$$

We construct a diffusion $(Q_t)_{t\in[0,T]}$ so that $Q_t \in \Sigma_{z(t)}$ for all $t \in [0,T]$ and $(Q_t)_{t\in[0,T]}$ satisfies the following properties:

- $Q_0 \sim \mu_{\Sigma_{z(0)}}$,
- For all $t \in [0,T]$, Q_{t+dt} is the orthogonal projection on $\Sigma_{z(t+dt)}$ of the position obtained by the unconstrained displacement: $Q_t - \nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t$.

More precisely, the considered diffusion reads, in the Stratonovich setting:

$$\begin{cases} Q_0 & \sim \mu_{\Sigma_{z(0)}}, \\ dQ_t & = -P(Q_t)\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}P(Q_t) \circ dB_t + \nabla\xi(Q_t)\, d\Lambda_t^{\mathrm{ext}}, \\ d\Lambda_t^{\mathrm{ext}} & = \dfrac{z'(t)}{|\nabla\xi(Q_t)|^2}dt. \end{cases} \tag{4.25}$$

With a view to the discretization of Q_t, let us notice that Q_t can be characterized by the following property:

Proposition 4.2. *The process Q_t solution to (4.25) is the only Itô process satisfying for some real-valued adapted Itô process $(\Lambda_t)_{t\in[0,T]}$:*

$$\begin{cases} Q_0 & \sim \mu_{\Sigma_{z(0)}}, \\ dQ_t & = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t + \nabla\xi(Q_t)\, d\Lambda_t, \\ \xi(Q_t) & = z(t). \end{cases}$$

Moreover, the process $(\Lambda_t)_{t\in[0,T]}$ can be decomposed as

$$\Lambda_t = \Lambda_t^{\mathrm{m}} + \Lambda_t^{\mathrm{f}} + \Lambda_t^{\mathrm{ext}}, \tag{4.26}$$

with the martingale part

$$d\Lambda_t^{\mathrm{m}} = -\sqrt{2\beta^{-1}}\frac{\nabla\xi}{|\nabla\xi|^2}(Q_t)\cdot dB_t,$$

the local force part (see (4.16) for the definition of f)

$$d\Lambda_t^{\mathrm{f}} = \frac{\nabla\xi}{|\nabla\xi|^2}(Q_t)\cdot\left(\nabla V(Q_t)\, dt + \beta^{-1}H(Q_t)\right) dt = f(Q_t)\, dt, \tag{4.27}$$

and the external forcing (or switching) term

$$d\Lambda_t^{\mathrm{ext}} = \frac{z'(t)}{|\nabla\xi(Q_t)|^2}\, dt.$$

The proof of Proposition 4.2 is easy and consists in computing $d\xi(Q_t)$ by Itô's calculus and identifying the bounded variation and the martingale parts of the stochastic processes.

The difference with the projected stochastic differential equation (4.19) considered in the thermodynamic integration setting is that the out-of-equilibrium evolution of the constraints $z(t)$ creates a drift $\nabla\xi(Q_t)\,d\Lambda_t^{\mathrm{ext}}$ along the reaction coordinate. This drift can be interpreted as an external forcing required for the switching to take place at a finite rate, and must be subtracted from the Lagrange multiplier Λ_t in order to obtain a correct expression for the work $\mathcal{W}(t)$ involved in the Feynman-Kac fluctuation equality (see Equations (4.43) and (4.45) below). This correction is quantitatively important when the switching is not slow.

The Feynman-Kac fluctuation equality

Let us define the nonequilibrium work exerted on the diffusion (4.25) by:

$$\mathcal{W}(t) = \int_0^t f(Q_s)\, z'(s)\, ds, \tag{4.28}$$

where f is the local mean force defined above by (4.16). Notice that, at least formally, in the limit of an infinitely slow switching from $z(0) = 0$ to $z(T) = 1$, Formula (4.30) corresponds to the thermodynamic integration formula (4.18). Formula (4.30) enables the computation of free energy differences at arbitrary rates, through a correction consisting in a reweighting of the nonequilibrium paths.

In practice, the nonequilibrium work $\mathcal{W}(t)$ can be computed by using the local force part $d\Lambda_t^{\mathrm{f}}$ (see (4.27)), as in the thermodynamic integration method (see (4.24)). Thus, the formula we use to compute $\mathcal{W}(t)$ is rather:

$$\mathcal{W}(t) = \int_0^t z'(s)\, d\Lambda_s^{\mathrm{f}}, \tag{4.29}$$

since Λ_t^{f} can be obtained by a natural numerical scheme (see Section 4.1.3), avoiding the cumbersome computations of the mean curvature vector H in the expression of f (as already explained above).

We can now state the generalization of the Jarzynski nonequilibrium equality to the case when the switching is parameterized by a reaction coordinate.

Theorem 4.1 (Feynman-Kac fluctuation equality). *For any test function φ and $\forall t \in [0,T]$, it holds*

$$\frac{Z_{z(t)}}{Z_{z(0)}} \int_{\Sigma_{z(t)}} \varphi\, d\mu_{\Sigma_{z(t)}} = \mathbb{E}\left(\varphi(Q_t)\mathrm{e}^{-\beta\mathcal{W}(t)}\right).$$

In particular, we have the work fluctuation identity: $\forall t \in [0,T]$,

$$\Delta F(z(t)) = F(z(t)) - F(z(0)) = -\beta^{-1}\ln\left(\mathbb{E}\left(\mathrm{e}^{-\beta\mathcal{W}(t)}\right)\right). \tag{4.30}$$

As in the alchemical case [177], the proof follows from a Feynman-Kac formula (see Theorem 4.2 for a proof in the general multi-dimensional case).

Extension to the general multi-dimensional case and proofs

In this section, we generalize the previous results for nonequilibrium computation of free energy differences presented for a one-dimensional reaction coordinate to the case of multi-dimensional reaction coordinates.

Geometric setting and basic notation and formulae.

We consider a d-dimensional system of smooth reaction coordinates $\xi = (\xi_1, \ldots, \xi_d) : \mathbb{R}^{3N} \to \mathbb{R}^d$, non-singular on an open domain $\mathcal{M} \subset \mathbb{R}^{3N}$

$$\forall q \in \mathcal{M}, \quad \text{range}(\nabla\xi_1(q), \ldots, \nabla\xi_d(q)) = d,$$

and a smooth path of associated coordinates

$$z = (z_1, \ldots, z_d) : [0, T] \to \mathbb{R}^d.$$

Accordingly, we define for all $t \in [0, T]$ a smooth submanifold of codimension d contained in $\mathcal{M}$:

$$\Sigma_{z(t)} = \left\{ q \in \mathbb{R}^{3N},\, \xi(q) = z(t) \right\} \subset \mathcal{M}.$$

In the constraints space $\mathbb{R}^d$, coordinates are labeled by Greek letters and we use the summation convention on repeated indices. In the configuration space $\mathbb{R}^{3N}$, coordinates are labeled by Latin letters and we also use the summation convention on repeated indices. We denote by $X \cdot Y = X_i Y_i$ the scalar product of two vector fields of $\mathbb{R}^{3N}$, by $M : N = M_{i,j} N_{i,j}$ the contraction of two tensor fields of $\mathbb{R}^{3N}$, and by $(X \otimes Y)_{i,j} = X_i Y_j$ the tensor product of two vector fields of $\mathbb{R}^{3N}$.

The $d \times d$ matrix

$$G_{\alpha,\gamma} = \nabla\xi_\alpha \cdot \nabla\xi_\gamma$$

is the Gram matrix of the constraints. It is symmetric and strictly positive on $\mathcal{M}$. We denote by $G^{-1}_{\alpha,\gamma}$ the (α, γ) component of G^{-1}, the inverse matrix of G. At each point $q \in \mathcal{M}$, we define the orthogonal projection operator

$$P^\perp = G^{-1}_{\alpha,\gamma} \nabla\xi_\alpha \otimes \nabla\xi_\gamma$$

onto the normal space to $\Sigma_{\xi(q)}$ and the orthogonal projection operator

$$P = \text{Id} - P^\perp$$

onto the tangent space to $\Sigma_{\xi(q)}$. The mean curvature vector field of the submanifold is defined by:

$$H = -\nabla \cdot \left((\det G)^{1/2} G^{-1}_{\alpha,\gamma} \nabla\xi_\gamma \right) (\det G)^{-1/2} \nabla\xi_\alpha \tag{4.31}$$

and satisfies:

$$H_i = P_{j,k} \nabla_j P_{i,k}.$$

We recall the divergence theorem on submanifolds: for any smooth function $\phi : \mathbb{R}^{3N} \to \mathbb{R}^{3N}$ with compact support,

$$\int_{\Sigma_z} \text{div}_\Sigma(\phi)\, d\sigma_{\Sigma_z} = -\int_{\Sigma_z} H \cdot \phi\, d\sigma_{\Sigma_z} \tag{4.32}$$

where $\text{div}_\Sigma(\phi) = P_{i,j} \nabla_i \phi_j$ denotes the surface divergence, and σ_{Σ_z} is the induced Lebesgue measure on the submanifold Σ_z of $\mathbb{R}^{3N}$. We will also use the co-area formula: for any smooth function $\phi : \mathbb{R}^{3N} \to \mathbb{R}$,

$$\int_{\mathbb{R}^{3N}} \phi(q) (\det G(q))^{1/2} dq = \int_{\mathbb{R}^d} \int_{\Sigma_z} \phi\, d\sigma_{\Sigma_z}\, dz. \tag{4.33}$$

These definitions and formulae are provided with more details in [66].

Free energy and constrained diffusions for multi-dimensional reaction coordinates

As in the one-dimensional case, the Boltzmann-Gibbs distribution restricted on the submanifold Σ_z is defined by:

$$d\mu_{\Sigma_z} = Z_z^{-1} \exp(-\beta V) d\sigma_{\Sigma_z},$$

with

$$Z_z = \int_{\Sigma_z} \exp(-\beta V) d\sigma_{\Sigma_z}.$$

The associated free energy is:

$$F(z) = -\beta^{-1} \ln (Z_z).$$

Remark 4.2 (On the definition of the free energy: the multi-dimensional case). *As in the one-dimensional case (see Remark 4.1), if the particles initially evolve in a potential V, the classical definition of the free energy is as above, but with V replaced by an effective potential $V + \beta^{-1} \ln \left((\det G)^{1/2} \right)$. The computation of the gradient of this potential in the dynamics then involves second-order derivatives of ξ, which can be approximated in practice by finite differences (see [66]).*

For any $1 \leq \alpha \leq d$, we now introduce the local mean force along $\nabla \xi_\alpha$ (which generalizes (4.16)):

$$f_\alpha = G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \cdot \left(\nabla V + \beta^{-1} H \right). \tag{4.34}$$

As in the one-dimensional case (see Equation (4.15)), we obtain the derivative of the mean force by averaging the local mean force:

Proposition 4.3. *The derivative of the free energy F with respect to z_α is given by:*

$$\nabla_\alpha F(z) = \int_{\Sigma_z} f_\alpha \, d\mu_{\Sigma_z}.$$

Proposition 4.3 is a corollary of

Lemma 4.1. *For any test function φ with compact support in $\mathcal{M}$, we have:*

$$\nabla_\alpha \left(\int_{\Sigma_z} \varphi \exp(-\beta V) d\sigma_{\Sigma_z} \right) = \int_{\Sigma_z} \left(G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \cdot \nabla \varphi - \beta f_\alpha \varphi \right) \exp(-\beta V) d\sigma_{\Sigma_z}.$$

Proof. It is enough to prove the formula in the case $V = 0$, up to a modification of the test function φ. For any test function $g : \mathbb{R} \to \mathbb{R}$ with compact support, we have (using successively an integration by parts on $\mathbb{R}$, the co-area formula (4.33), an integration by parts on $\mathbb{R}^{3N}$, and finally again (4.33)):

$$\begin{aligned}
\int_{\mathbb{R}^d} g(z_\alpha) \nabla_\alpha \left(\int_{\Sigma_z} \varphi d\sigma_{\Sigma_z} \right) dz &= - \int_{\mathbb{R}^d} \int_{\Sigma_z} g'(z_\alpha) \varphi \, d\sigma_{\Sigma_z} dz, \\
&= - \int_{\mathbb{R}^{3N}} g' \circ \xi_\alpha \, \varphi \, (\det G)^{1/2} \, dq, \\
&= - \int_{\mathbb{R}^{3N}} G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \cdot \nabla (g \circ \xi_\alpha) \, \varphi \, (\det G)^{1/2} \, dq, \\
&= \int_{\mathbb{R}^{3N}} g \circ \xi_\alpha \nabla \cdot \left(G^{-1}_{\alpha,\gamma} \, \nabla \xi_\gamma \, \varphi \, (\det G)^{1/2} \right) dq, \\
&= \int_{\mathbb{R}^d} g(z_\alpha) \int_{\Sigma_z} \nabla \cdot \left(G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \, \varphi \, (\det G)^{1/2} \right) (\det G)^{-1/2} \, d\sigma_{\Sigma_z} \, dz,
\end{aligned}$$

which gives the result using the expression (4.31) of the mean curvature vector H.

We now define the constrained diffusion (which generalizes (4.25)):

$$\begin{cases}
Q_0 & \sim \mu_{\Sigma_{z(0)}}, \\
dQ_t & = -P(Q_t) \nabla V(Q_t) dt + \sqrt{2\beta^{-1}} P(Q_t) \circ dB_t + \nabla \xi_\alpha(Q_t) d\Lambda^{\text{ext}}_{\alpha,t}, \\
d\Lambda^{\text{ext}}_{\alpha,t} & = G^{-1}_{\alpha,\gamma}(Q_t) z'_\gamma(t) dt, \qquad \forall 1 \leq \alpha \leq d.
\end{cases} \tag{4.35}$$

The stochastic process Q_t can be characterized by the following property:

Proposition 4.4. *The process Q_t solution to (4.35) is the only Itô process satisfying for some adapted Itô processes $(\Lambda_{1,t}, \ldots, \Lambda_{d,t})_{t\in[0,T]}$ with values in $\mathbb{R}^d$:*

$$\begin{cases} Q_0 & \sim \mu_{\Sigma_{z(0)}}, \\ dQ_t & = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t + \nabla\xi_\alpha(Q_t)d\Lambda_{\alpha,t}, \\ \xi(Q_t) & = z(t). \end{cases}$$

Moreover, the process $(\Lambda_{\alpha,t})_{t\in[0,T]}$ can be decomposed as

$$\Lambda_{\alpha,t} = \Lambda^{\mathrm{m}}_{\alpha,t} + \Lambda^{\mathrm{f}}_{\alpha,t} + \Lambda^{\mathrm{ext}}_{\alpha,t},$$

with the martingale part

$$d\Lambda^{\mathrm{m}}_{\alpha,t} = -\sqrt{2\beta^{-1}}G^{-1}_{\alpha,\gamma}\nabla\xi_\gamma(Q_t)\cdot dB_t,$$

the local force part (see (4.34) for the definition of f_α)

$$d\Lambda^{\mathrm{f}}_{\alpha,t} = f_\alpha(Q_t)dt,$$

and the external forcing (or switching) term

$$d\Lambda^{\mathrm{ext}}_{\alpha,t} = G^{-1}_{\alpha,\gamma}(Q_t)z'_\gamma(t)dt.$$

The proof consists in computing $d\xi(Q_t)$ by Itô's calculus and identifying the bounded variation and the martingale parts of the stochastic processes.

The Feynman-Kac fluctuation equality

Theorem 4.1 is generalized as:

Theorem 4.2 (Feynman-Kac fluctuation equality). *Let us define the nonequilibrium work exerted on the diffusion Q_t solution to (4.35) by:*

$$\mathcal{W}(t) = \int_0^t f_\alpha(Q_s)z'_\alpha(s)\,ds = \int_0^t z'_\alpha(s)d\Lambda^{\mathrm{f}}_{\alpha,s}.$$

Then, we have the following fluctuation equality: for any test function φ, and $\forall t \in [0,T]$,

$$\frac{Z_{z(t)}}{Z_{z(0)}}\int_{\Sigma_{z(t)}} \varphi\, d\mu_{\Sigma_{z(t)}} = \mathbb{E}\left(\varphi(Q_t)\mathrm{e}^{-\beta\mathcal{W}(t)}\right). \tag{4.36}$$

In particular, we have the work fluctuation identity: $\forall t \in [0,T]$,

$$\Delta F(z(t)) = F(z(t)) - F(z(0)) = -\beta^{-1}\ln\left(\mathbb{E}\left(\mathrm{e}^{-\beta\mathcal{W}(t)}\right)\right). \tag{4.37}$$

Proof. For any $s \in [0,T]$ and $x \in \mathcal{M}$, let us introduce $(Q^{s,x}_t)_{t\in[s,T]}$, the stochastic process satisfying the SDE (4.35), starting from x at time s:

$$\begin{cases} Q^{s,x}_s & = x, \\ dQ^{s,x}_t & = -P(Q^{s,x}_t)\nabla V(Q^{s,x}_t)dt + \sqrt{2\beta^{-1}}P(Q^{s,x}_t)\circ dB_t + \nabla\xi_\alpha(Q^{s,x}_t)d\Lambda^{\mathrm{ext}}_{\alpha,t}, \\ d\Lambda^{\mathrm{ext}}_{\alpha,t} & = G^{-1}_{\alpha,\gamma}(Q^{s,x}_t)z'_\gamma(t)dt, \qquad \forall 1 \le \alpha \le d. \end{cases} \tag{4.38}$$

Notice that for any $s \in [0,T]$, there is an open neighborhood $(s^-, s^+)\times\mathcal{M}_s$ of $(s, \Sigma_{z(s)})$ in $\mathbb{R}\times\mathcal{M}$ such that the diffusion $(Q^{s,x}_t)_{t\in[s,T]}$ remains in $\mathcal{M}$ almost surely. This holds since this process

satisfies $d\xi(Q_t^{s,x}) = z'(t)\,dt$ and therefore $\xi(Q_t^{s,x}) = \xi(x) + z(t) - z(s)$. This gives usual regularity assumptions sufficient to get a backward semi-group (t being from now on fixed in $(0,T)$ and s varying in $[0,t]$):

$$u(s,x) = \mathbb{E}\left(\varphi(Q_t^{s,x})\exp\left(-\beta\int_s^t f_\alpha(Q_r^{s,x})z'_\alpha(r)\,dr\right)\right),$$

satisfying the following partial differential equation (PDE) on $(s^-,s^+)\times\mathcal{M}_s$:

$$\partial_s u = -L_s(u(s,.)) + \beta z'_\alpha(s) f_\alpha u,$$

where L_s is the generator of the diffusion Q_t solution to (4.35):

$$L_s = \beta^{-1}P : \nabla^2 - P\nabla V\cdot\nabla + \beta^{-1}H\cdot\nabla + z'_\gamma(s)G^{-1}_{\alpha,\gamma}\nabla\xi_\alpha\cdot\nabla.$$

Now, using Lemma 4.1, we have:

$$\begin{aligned}
&\frac{d}{ds}\int_{\Sigma_{z(s)}} u(s,.)\exp(-\beta V)d\sigma_{\Sigma_{z(s)}}\\
&= \int_{\Sigma_{z(s)}}\left(-L_s(u(s,.)) + z'_\alpha(s)G^{-1}_{\alpha,\gamma}\nabla\xi_\gamma\cdot\nabla u(s,.)\right)\exp(-\beta V)d\sigma_{\Sigma_{z(s)}},\\
&= -\int_{\Sigma_{z(s)}}\left(\beta^{-1}P:\nabla^2 u(s,.) - P\nabla V\cdot\nabla u(s,.) + \beta^{-1}H\cdot\nabla u(s,.)\right)\exp(-\beta V)d\sigma_{\Sigma_{z(s)}},\\
&= -\beta^{-1}\int_{\Sigma_{z(s)}}\Big(\mathrm{div}_\Sigma\left(\nabla u(s,.)\exp(-\beta V)\right) + H\cdot\nabla u(s,.)\exp(-\beta V)\Big)d\sigma_{\Sigma_{z(s)}},\\
&= 0,
\end{aligned}$$

by the divergence theorem (4.32). Therefore

$$\int_{\Sigma_{z(t)}} u(t,.)\exp(-\beta V)d\sigma_{\Sigma_{z(t)}} = \int_{\Sigma_{z(0)}} u(0,.)\exp(-\beta V)d\sigma_{\Sigma_{z(0)}},$$

which yields

$$\int_{\Sigma_{z(t)}}\varphi\exp(-\beta V)d\sigma_{\Sigma_{z(t)}} = Z_{z(0)}\mathbb{E}\left(\varphi(Q_t)\exp\left(-\beta\int_0^t f_\alpha(Q_r)z'_\alpha(r)\,dr\right)\right),$$

where Q_t satisfies (4.35). This proves (4.36), and (4.37) is obtained by taking $\varphi = 1$. □

4.1.3 Practical computation of free energy differences

We present in this section numerical strategies suited for the reaction coordinate case, the numerical discretization of the alchemical case being trivial.

Discretization of the projected dynamics

The main interest of the above formulae (4.18)–(4.24) and (4.29)–(4.30) is that they admit natural time discretizations. The principle is to use a predictor-corrector scheme for the associated dynamics (4.19) and (4.25), and to use the Lagrange multiplier Λ_t to compute the local mean force f.

Discretization of the projected diffusion (equilibrium case)

For the projected SDE (4.20) onto a submanifold $\Sigma_z = \{\xi(q) - z = 0\}$, two discretizations of the dynamics, extending the usual Euler-Maruyama scheme, are proposed in [66]. These numerical

schemes for constrained Brownian dynamics are in the spirit of the so-called RATTLE [8] and SHAKE [295] algorithms classical used for constrained Hamiltonian dynamics, and also related with the algorithms proposed in [6, 262, 358].

The first one is:

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{n+1}\,\nabla\xi(Q_{n+1}), \\ \text{where } \Delta\Lambda_{n+1} \text{ is such that } \xi(Q_{n+1}) = z, \end{cases} \tag{4.39}$$

where Δt is the time step and U_n is a $3N$-dimensional standard Gaussian random vector. Notice that (4.39) admits a natural variational interpretation, since Q_{n+1} can be seen as the closest point on the submanifold Σ_z to the predicted position $Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\beta^{-1}}\,U_n$. The real $\Delta\Lambda_{n+1}$ is then the Lagrange multiplier associated with the constraint $\xi(Q_{n+1}) = z$.

Another possible discretization of (4.20) is

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{n+1}\,\nabla\xi(Q_n), \\ \text{where } \Delta\Lambda_{n+1} \text{ is such that } \xi(Q_{n+1}) = z. \end{cases} \tag{4.40}$$

Although this scheme is not naturally associated with a variational principle, it may be more practical since its formulation is more explicit. Notice also that we use the same notation $\Delta\Lambda_n$ for the Lagrange multipliers for both (4.39) and (4.40) (and later for (4.41) and (4.42)), since all the formulas we state in terms of $\Delta\Lambda_n$ are verified whatever the constrained dynamics.

To solve Equation (4.39), classical methods for optimization problems with constraints can be used. We refer to [135] for a presentation of the classical Uzawa algorithm, and to [36] for more advanced methods. Problem (4.40) can be solved using classical methods for nonlinear problems, such as the Newton method (see [36]). We also refer to Chapter 7 of [205] where similar problems are discussed, for the classical RATTLE and SHAKE schemes used for Hamiltonian dynamics with constraints.

Both schemes are consistent (the discretization error goes to 0 when the time step Δt goes to 0) with the projected diffusion (4.20) (see [66]). Accordingly, $\Delta\Lambda_{n+1}$ is a consistent discretization of $\int_{t_n}^{t_{n+1}} d\Lambda_t$ and therefore, it can be proven [66]:

$$\lim_{T\to\infty}\lim_{\Delta t\to 0}\frac{1}{T}\sum_{n=1}^{T/\Delta t}\Delta\Lambda_n = F'(z)$$

which is the discrete counterpart of the trajectory average (4.24). In [66], a variance reduction technique is proposed, which consists in extracting the bounded variation part $\Delta\Lambda_n^{\mathrm{f}}$ of $\Delta\Lambda_n$ (resorting locally to reversed Brownian increments). We give some details of an adaptation of this method for evolving constraints in next section.

Discretization with evolving constraints

When nonequilibrium dynamics are considered, the constraint is stated as $\xi(Q_t) = z(t)$. The reaction coordinate path is first discretized as $\{z(0), \ldots, z(t_{N_T})\}$ where N_T is the number of timesteps. For example, equal time increments can be used, in which case $\Delta t = \frac{T}{N_T}$ and $t_n = n\Delta t$ (we refer to Remark 4.3 below for some refinements). The initial conditions Q_0 are sampled according to μ_{Σ_0}. A way to do that is to subsample a long trajectory of the projected SDE on Σ_0 (using the schemes (4.39) or (4.40)).

The projected SDE on evolving constraints (4.25) is then discretized with the scheme (4.39) or (4.40), taking into account the evolution of the constraint:

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{n+1}\,\nabla\xi(Q_{n+1}), \\ \text{where } \Delta\Lambda_{n+1} \text{ is such that } \xi(Q_{n+1}) = z(t_{n+1}), \end{cases} \tag{4.41}$$

or

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{n+1}\,\nabla\xi(Q_n), \\ \text{where } \Delta\Lambda_{n+1} \text{ is such that } \xi(Q_{n+1}) = z(t_{n+1}). \end{cases} \tag{4.42}$$

It remains to extract the force part $\Delta\Lambda^{\mathrm{f}}_{n+1}$ from the discretized Lagrange multiplier $\Delta\Lambda_{n+1}$ (consistently with (4.26)). We propose two methods. First, this can be done by simply subtracting the drift and the martingale part

$$\Delta\Lambda^{\mathrm{f}}_{n+1} = \Delta\Lambda_{n+1} - \frac{z(t_{n+1}) - z(t_n)}{|\nabla\xi(Q_n)|^2} + \sqrt{2\Delta t\beta^{-1}}\,\frac{\nabla\xi(Q_n)}{|\nabla\xi(Q_n)|^2}\cdot U_n. \tag{4.43}$$

Another possibility in the spirit of the variance reduction techniques used in [66] can also be used. Consider the following coupled dynamic with locally time-reversed constraint evolution (written here for the scheme (4.41)):

$$Q^{\mathrm{R}}_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t - \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda^{\mathrm{R}}_{n+1}\,\nabla\xi(Q^{\mathrm{R}}_{n+1}),$$

with $\Delta\Lambda^{\mathrm{R}}_{n+1}$ such that:

$$\frac{1}{2}(\xi(Q^{\mathrm{R}}_{n+1}) + \xi(Q_{n+1})) = \xi(Q_n).$$

The position Q^{R}_{n+1} is computed as Q_{n+1} in (4.41), but with a projection on $\Sigma_{2\xi(Q_n)-\xi(Q_{n+1})}$ instead of $\Sigma_{z(t_{n+1})}$, and using the Brownian increment $-\sqrt{\Delta t}\,U_n$ instead of $\sqrt{\Delta t}\,U_n$. Notice that in case of a constant increment for the constraints, we have $\xi(Q^{\mathrm{R}}_{n+1}) = 2\xi(Q_n) - \xi(Q_{n+1}) = z(t_{n-1})$. The force part $\Delta\Lambda^{\mathrm{f}}_{n+1}$ is then obtained through

$$\Delta\Lambda^{\mathrm{f}}_{n+1} = \frac{1}{2}(\Delta\Lambda_{n+1} + \Delta\Lambda^{\mathrm{R}}_{n+1}) \tag{4.44}$$

which can be shown to be a consistent time discretization of $\int_{t_n}^{t_{n+1}} d\Lambda^{\mathrm{f}}_t$.

Computation of free energy using a Feynman-Kac equality

The consistent discretization of Q_t, and more precisely of $\int_{t_n}^{t_{n+1}} d\Lambda^{\mathrm{f}}_t$, we have obtained in the previous section can now be used to approximate the work $\mathcal{W}(t)$ defined by (4.29) by

$$\begin{cases} \mathcal{W}_0 = 0, \\ \mathcal{W}_{n+1} = \mathcal{W}_n + \dfrac{z(t_{n+1}) - z(t_n)}{t_{n+1} - t_n}\,\Delta\Lambda^{\mathrm{f}}_{n+1}, \end{cases} \tag{4.45}$$

using either the dynamics (4.41) or (4.42), and the local force part of the Lagrange multiplier computed by (4.43) or (4.44). Averaging over M independent realizations (the corresponding works being labeled by an upper index $1 \le m \le M$), an estimator of the free energy difference $\Delta F(z(T))$ is, using Theorem 4.1,

$$\widehat{\Delta F}(z(T)) = -\beta^{-1}\ln\left(\frac{1}{M}\sum_{m=1}^{M} \mathrm{e}^{-\beta\mathcal{W}^m_{N_T}}\right). \tag{4.46}$$

The estimator $\widehat{\Delta F}(z(T))$ converges to $\Delta F(z(T))$ as $\Delta t \to 0$ and $M \to +\infty$. It is clear that the estimation of $\Delta F(z(T))$ by (4.46) is straightforward to parallelize since the $(\mathcal{W}^m_{N_T})_{1\le m\le M}$ are independent.

For a fixed $M < +\infty$, notice that, even in the limit $\Delta t \to 0$, $\widehat{\Delta F}(z(T))$ is a biased estimator. Indeed,
$\exp(-\beta\widehat{\Delta F}(z(T)))$ is an unbiased estimator of $\exp(-\beta\Delta F(z(T)))$, and therefore, using the concavity of ln, $\mathbb{E}(\widehat{\Delta F}(z(T))) \ge \Delta F(z(T))$. Recent works propose corrections to this systematic bias using asymptotic expansions in the limit $M \to +\infty$ (see for instance [286, 378]).

Remark 4.3 (On practical implementation). *Notice that it may be useful to adaptively refine the time step over each stochastic trajectories, using for example the work evolution rate* $(\mathcal{W}_n - \mathcal{W}_{n-1})_{n\geq 1}$ *as a refinement criterion. As noticed in [286], it is also possible to optimize the evolution of the constraint* $z(t)$*, for example by minimizing the variance of the results obtained for a priori schedules for the evolving constraint on a small set of preliminary runs.*

The numerical scheme in the multi-dimensional case

The adaptation of the algorithm we propose for the one-dimensional case to the multi-dimensional case is straightforward. Indeed, the generalizations of schemes (4.41) and (4.42) to the multi-dimensional case are, respectively:

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{\alpha,n+1}\,\nabla\xi_\alpha(Q_{n+1}), \\ \text{where } (\Delta\Lambda_{\alpha,n+1})_{1\leq\alpha\leq d} \text{ is such that } \xi(Q_{n+1}) = z(t_{n+1}), \end{cases}$$

$$\begin{cases} Q_{n+1} = Q_n - \nabla V(Q_n)\,\Delta t + \sqrt{2\Delta t\,\beta^{-1}}\,U_n + \Delta\Lambda_{\alpha,n+1}\,\nabla\xi_\alpha(Q_n), \\ \text{where } (\Delta\Lambda_{\alpha,n+1})_{1\leq\alpha\leq d} \text{ is such that } \xi(Q_{n+1}) = z(t_{n+1}). \end{cases}$$

The force part $\Delta\Lambda^{\mathrm{f}}_{\alpha,n}$ of $\Delta\Lambda_{\alpha,n}$ is obtained by similar procedures as those described in Section 4.1.3. For example, the generalization of (4.43) is:

$$\Delta\Lambda^{\mathrm{f}}_{\alpha,n+1} = \Delta\Lambda_{\alpha,n+1} - G^{-1}_{\alpha,\gamma}(Q_n)\,(z_\gamma(t_{n+1}) - z_\gamma(t_n)) + \sqrt{2\Delta t\beta^{-1}}G^{-1}_{\alpha,\gamma}\nabla\xi_\gamma(Q_n)\cdot U_n.$$

The generalization of (4.44) is also straightforward.

Now, the estimator $\widehat{\Delta F}(z(T))$ of the free energy difference $\Delta F(z(T))$ is given by (4.46), with the following approximation of the work $\mathcal{W}(t)$:

$$\begin{cases} \mathcal{W}_0 = 0, \\ \mathcal{W}_{n+1} = \mathcal{W}_n + \dfrac{z_\alpha(t_{n+1}) - z_\alpha(t_n)}{t_{n+1} - t_n}\,\Delta\Lambda^{\mathrm{f}}_{\alpha,n+1}, \end{cases}$$

which generalizes (4.45). Notice that Remark 4.3 also holds for a multi-dimensional reaction coordinate.

4.1.4 Numerical results

We present in this section some illustrations of the algorithm we have described above to compute free energy differences through nonequilibrium paths. In Section 4.1.4, a two-dimensional toy potential V is used, for which we can compare the results with analytical profiles. A more realistic test case in Section 4.1.4 demonstrates the ability of the method to compute free energy profiles in presence of a free energy barrier.

Our aim in this section is not to compare the numerical efficiency of the thermodynamic integration method presented (or any other method) with nonequilibrium computations, since it is difficult to draw *general* conclusions about such comparisons. However, we compare on a simple example in Section 4.1.4, the numerical efficiency of out-of-equilibrium computations using a few long trajectories or many short trajectories, at a fixed computational cost.

A two-dimensional toy problem

We consider the two-dimensional potential introduced in [365]:

$$V(x,y) = \cos(2\pi x)(1 + d_1 y) + d_2 y^2, \tag{4.47}$$

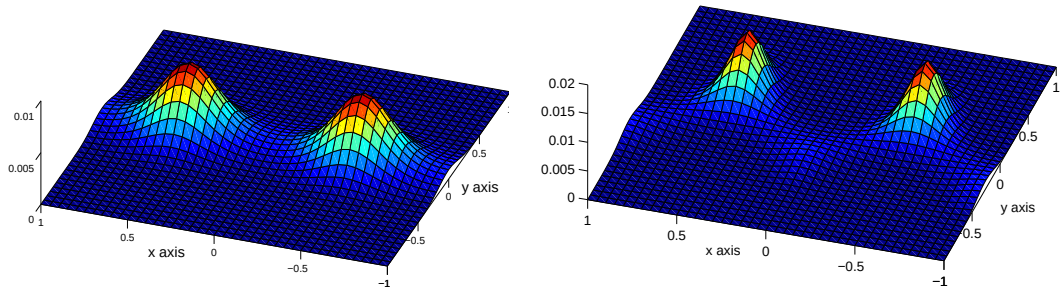


Fig. 4.1. Plot of some probability densities corresponding to the potential (4.47) for $\beta = 1$, $d_2 = 2\pi^2$, and $d_1 = 0$ on the left or $d_1 = 10$ on the right.

where d_1 and d_2 are two positive constants. Some corresponding Boltzmann-Gibbs probability densities are depicted in Figure 4.1.

We want to compute the free energy difference profile between the initial state $x = x_0 = -0.5$ and the transition state $x = x_1 = 0$. Notice that the saddle point is $(x_1, y_1) = (0, 0)$ for $d_1 = 0$, but is increasingly shifted toward lower values of y_1 as d_1 increases. We parameterize the transition along the x-axis, either with the reaction coordinate

$$\xi(x, y) = \frac{x - x_0}{x_1 - x_0}, \tag{4.48}$$

or with the reaction coordinate ($n \geq 2$)

$$\eta_n(x, y) = \frac{1}{2^n - 1}\left[\left(1 + \frac{x - x_0}{x_1 - x_0}\right)^n - 1\right]. \tag{4.49}$$

For these reaction coordinates, the initial state (resp. the transition state) corresponds to a value of the reaction coordinate $z = 0$ (resp. $z = 1$). The analytical expression of the free energy difference that we consider here is, for a reaction coordinate $\nu(x, y)$ (such as ξ or η_n defined above)

$$\Delta F_\nu(z) = -\beta^{-1} \ln\left(\frac{\int_{\mathbb{R}^2} \mathrm{e}^{-\beta V(x,y)} \delta_{\nu(x,y)-z}}{\int_{\mathbb{R}^2} \mathrm{e}^{-\beta V(x,y)} \delta_{\nu(x,y)}}\right),$$

where the distribution $\delta_{\nu(x,y)-z}$ is defined in Remark 4.1 above. Notice that even though the initial state $\Sigma_0 = \{x = -0.5\}$ and the final state $\Sigma_1 = \{x = 0\}$ are the same for the reaction coordinates ξ and η_n, the associated free energy differences differ. This is due to the fact that $\nabla\xi \neq \nabla\eta_n$, and therefore $\delta_{\xi(x,y)-z} \neq \delta_{\eta_n(x,y)-z}$. More precisely,

$$\Delta F_\xi(z) = -\cos(2\pi x_0) + \cos(2\pi x_\xi(z)) + \frac{(d_1)^2}{4d_2}(\cos^2(2\pi x_0) - \cos^2(2\pi x_\xi(z))),$$

with

$$x_\xi(z) = x_0 + z(x_1 - x_0),$$

and

$$\Delta F_{\eta_n}(z) = -\cos(2\pi x_0) + \cos(2\pi x_{\eta_n}(z)) + \frac{(d_1)^2}{4d_2}(\cos^2(2\pi x_0) - \cos^2(2\pi x_{\eta_n}(z))) + \frac{n-1}{\beta}\ln\left(1 + \frac{x_{\eta_n}(z) - x_0}{x_1 - x_0}\right),$$

with

$$x_{\eta_n}(z) = x_0 + ((2^n - 1)z + 1)^{1/n} - 1)(x_1 - x_0).$$

Free energy profiles for the two reaction coordinates considered here can then be computed using the discretization proposed in Section 4.1.3. Averaging over several realizations, error estimates can be proposed: in particular, the standard deviation can be computed for all intermediate points $z \in [0, 1]$, so that, for all values z, a confidence interval around the empirical mean can be proposed. We represent on Figure 4.2 the analytical profiles, and the lower and upper bounds of the 95 % confidence interval for $M = 10^3$ and $M = 10^4$, using here and henceforth a linear schedule: $z(t) = t/T$. The initial conditions are created by subsampling a trajectory constrained to remain on the initial submanifold Σ_0. As announced above, the profiles obtained with η_n and ξ are not exactly the same, though the general shape is preserved. These figures also show that the variance increases with z. Therefore, to further test the convergence of the method, it is enough here to characterize the convergence of the value for the end point at $z = 1$.

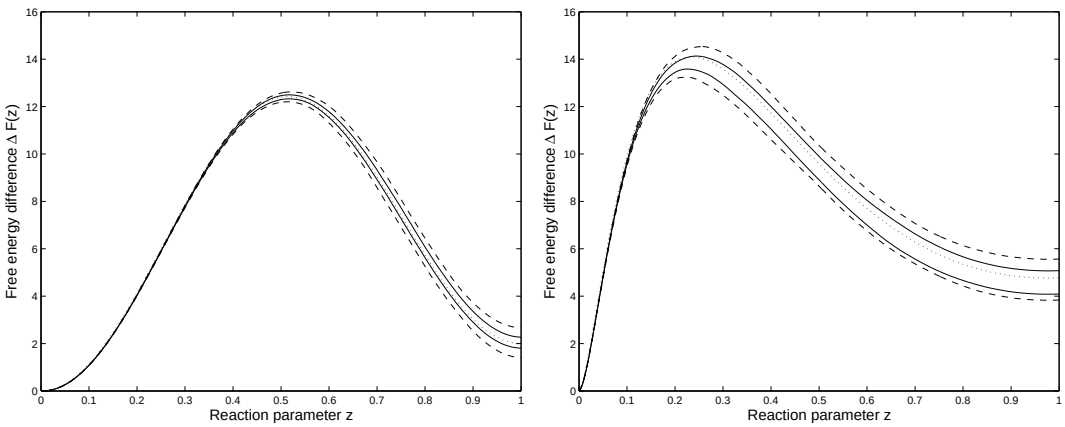


Fig. 4.2. Free energy profiles using the potential (4.47) with $\beta = 1$, $d_1 = 30$ and $d_2 = 2\pi^2$, and the reaction coordinate (4.48) on the left, or the reaction coordinate (4.49) with $n = 5$ on the right. Analytical reference profiles are in dotted lines. The dashed lines (resp. the solid lines) represent the upper and lower bound of the 95 % confidence interval (obtained over 100 independent realizations) for nonequilibrium computations with $M = 10^3$ replicas (resp. with $M = 10^4$ replicas). The switching time is $T = 1$ and the time step is $\Delta t = 0.005$ on the left and $\Delta t = 0.0025$ on the right.

We study the convergence of the end value $\Delta F(1)$ computed with the out-of-equilibrium dynamics with respect to the number of replicas M and the time step Δt, using the reaction coordinate (4.48) as an example. The results are presented in Table 4.1. The time step Δt does not seem to have any noticeable influence on the final result, as long as it remains in a reasonable range. As expected, the error gets smaller as M increases.

In Table 4.1, we also show that, in this particular case, for a fixed computational cost and provided that the switching time is large enough[1], computing many short trajectories is as efficient as computing a few longer ones (the mean and the variance are essentially unchanged). This conclusion also holds for the more realistic test case presented in next section. The computation of many trajectories can be straightforwardly and very efficiently parallelized.

[1] Of course, this threshold time depends on the system under study.

We finally mention that we are able to exhibit the bias of the Jarzynski estimator in this particular case (see Section 4.1.3 and [378]). We observe that the estimator $\widehat{\Delta F}(z(T))$ is generally greater than $\Delta F(z(T))$. More precisely, averaging over 10^4 realizations, with the parameters $T = 1$ and $\Delta t = 0.005$, we obtain the following 95 % confidence intervals for $\widehat{\Delta F}(z(T))$, for various values of M: $\widehat{\Delta F}(z(T)) = 2.0576 \pm 0.0059$ for $M = 10^3$, $\widehat{\Delta F}(z(T)) = 2.0095 \pm 0.0026$ for $M = 10^4$, and $\widehat{\Delta F}(z(T)) = 2.00075 \pm 0.0010$ for $M = 10^5$. As expected, the bias goes to zero when $M \to \infty$.

Table 4.1. Free energy differences $\Delta F(1)$ obtained by nonequilibrium computations for the reaction coordinate (4.48) with $\beta = 1$, $d_1 = 1$ and $d_2 = 30$. The results are presented as follows: $\mathbb{E}\left(\widehat{\Delta F}(z(T))\right)$ $\left(\sqrt{\mathrm{Var}\left(\widehat{\Delta F}(z(T))\right)}\right)$ (the estimates of these quantities are obtained by averages over 100 independent runs). The exact value is $\Delta F(1) = 2$.

Δt	T	M	$\widehat{\Delta F}(z(T))$	
0.001	1	10^3	2.056	(0.274)
0.0025	1	10^3	2.033	(0.259)
0.005	1	10^3	2.076	(0.286)
0.01	1	10^3	2.073	(0.278)
0.005	1	10^3	2.076	(0.286)
0.005	1	10^4	2.014	(0.116)
0.005	1	10^5	2.001	(0.045)

Δt	T	M	$\widehat{\Delta F}(z(T))$	
0.005	1	10^4	2.014	(0.116)
0.005	10	10^3	1.999	(0.029)
0.005	100	10^2	2.001	(0.025)
0.005	1000	10^1	1.997	(0.022)

Model system for conformational changes influenced by solvation

We consider a system composed of N particles in a periodic box of side length l, interacting through the purely repulsive WCA pair potential [79, 329]:

$$V_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{if } r \le r_0, \\ 0 & \text{if } r > r_0, \end{cases}$$

where r denotes the distance between two particles, ϵ and σ are two positive parameters and $r_0 = 2^{1/6}\sigma$. Among these particles, two (numbered 1 and 2 in the following) are designated to form a dimer while the others are solvent particles. Instead of the above WCA potential, the interaction potential between the two particles of the dimer is a double-well potential

$$V_{\mathrm{S}}(r) = h\left[1 - \frac{(r - r_0 - w)^2}{w^2}\right]^2, \tag{4.50}$$

where h and w are two positive parameters. The potential V_{S} exhibits two energy minima, one corresponding to the compact state where the length of the dimer is $r = r_0$, and one corresponding to the stretched state where this length is $r = r_0 + 2w$. The energy barrier separating both states is h. Figure 4.3 presents a schematic view of the system.

The reaction coordinate used is

$$\xi(q) = \frac{|q_1 - q_2| - r_0}{2w}, \tag{4.51}$$

where q_1 and q_2 are the positions of the particles forming the dimer. The compact state (resp. the stretched state) corresponds to a value of the reaction coordinate $z = 0$ (resp. $z = 1$).

The parameters used for the simulations are: $\beta = 1$, $\epsilon = 1$, $\sigma = 1$, $h = 1$, $w = 0.5$ and $N = 16$. We still use a linear schedule: $z(t) = t/T$. The side length l of the simulation box

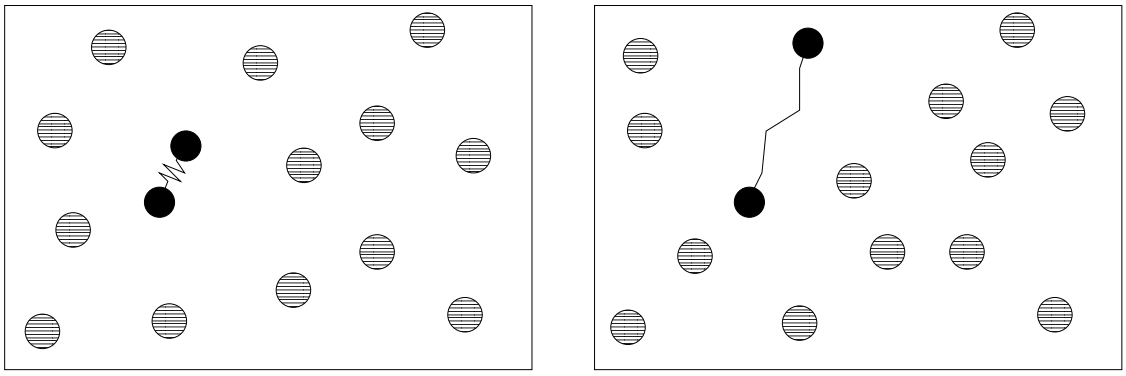


Fig. 4.3. Schematic views of the system, when the dimer is in the compact state (Left), and in the stretched state (Right). The interaction of the particles forming the dimer is described by a double well potential. All the other interactions are of WCA form.

takes two values: $l = 1.3$ (high density state) and $l = 3$ (low density state). Figure 4.4 presents some plots of the free energy difference profiles computed using nonequilibrium dynamics, as well as thermodynamic integration reference profiles. The results show that nonequilibrium estimates are consistent with thermodynamic integration. Our experience on this particular example also shows that it is computationally as efficient to simulate several short nonequilibrium trajectories (provided the switching time is not too small, say, $T \sim 1$ in the units used here, so that the diffusion process can take place), or one single long trajectory where the switching is done slowly (as already observed in the previous example).

The free energy profiles highlight the relative stabilities of the two conformations of the dimer: at low densities (Figure 4.4, Left) the stretched conformation has a lower free energy and is thus expected to be more stable (this can indeed be verified by running long molecular dynamics trajectories and monitoring the time spent in each conformation). When the density increases, the compact conformation becomes more and more likely. At the density considered in Figure 4.4 (Right), the compact state already has a free energy slightly smaller than the stretched state. Notice also that the free energy barrier increases as the density increases, so that spontaneous transitions are less and less frequent. But since we know here a reaction coordinate, we can enforce the transition. This prevents us from running and monitoring long trajectories to get sufficient statistics to compare relative occurrences of both states.

4.2 Equilibration of the nonequilibrium computation of free energy differences

We present in this section a complementary approach to the above nonequilibrium strategies in the Jarzynski way, to prevent the degeneracy of weights. It is similar to the method of [174], known as "population Monte-Carlo", in which multiple replicas are used to represent the distribution under study. A weight is associated to each replica, and resamplings are performed at discrete fixed times to avoid degeneracy of the weights. This methodology is widely used in the fields of Quantum Monte Carlo [13, 289] or Bayesian Statistics, where it is referred to as Sequential Monte Carlo [84, 85]. Note that in the probability and statistics fields, each simulation is called a 'walker' or 'particle'; we use here the name 'replica', which is more apppropriate to the Molecular Dynamics context.

The method used here extends the population Monte-Carlo method to the time-continuous case. It consists in running M replicas of the system in parallel, resorting typically to a stochastic dynamic, and considering exchanges between them, according to a certain probabilitic rule depending on the work done on each system. This procedure can be seen as automatic time continuous

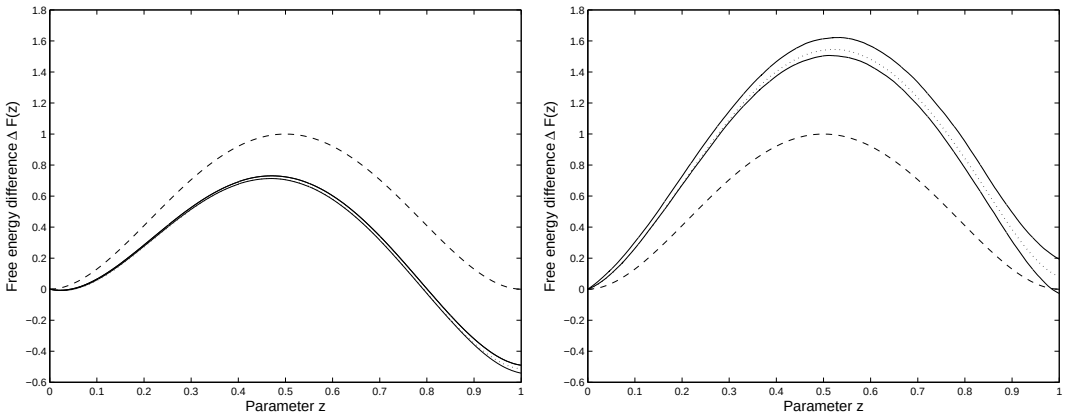


Fig. 4.4. Comparison of free energy difference profiles using the reaction coordinate (4.51), at low densities ($l = 3$) on the left, and high densities ($l = 1.3$) on the right. The double well potential V_S is represented in dashed line. The reference free energy difference profile computed with a very precise thermodynamic integration is represented in dotted line. We used $N_{\mathrm{TI}} = 101$ thermodynamic integration points (uniformly distributed over $(0, 1)$) and averaged the mean force over $M_{\mathrm{TI}} = 10^7$ configurations for each fixed value of z. The upper and lower bounds of the 95 % confidence interval (obtained over 50 independent realizations) for out-of-equilibrium computations are represented with solid lines. We used $M = 1000$ nonequilibrium trajectories, a switching time $T = 1$, and a timestep $\Delta t = 0.0005$ (left) or $\Delta t = 0.00025$ (right).

resampling, and all replicas have the same weight at any time of the simulation. This method drastically increases the number of significative transitions paths in nonequilibrium simulations. The set of all replicas (or walkers) is called an 'Interacting Particle System' (IPS) [248], and can be seen as a genetic algorithm where the mutation step is the stochastic dynamics considered.

This method also allows to end up the simulation with a well distributed sample of configurations. It is therefore a way to perform *simulated annealing* [193] rigorously: the idea is to switch slowly from an initial simple sampling problem, to the target sampling problem, through a well chosen interpolation. This allows to attain deeper local minima, but, due to its nonequilibrium nature, is not efficient as such to sample accurately the target measure. We mention that variations have been proposed, especially *tempering* methods (see [180] for a review), the most famous being *parallel tempering* [225]. These methods consider an additional parameter describing the configuration system (e.g. the temperature), and sample those extended configurations according to some stochastic rules. However, these methods asks for a prior distribution of the additional parameters (for example a temperature ladder in parallel tempering method), which are usually estimated through some preliminary runs [180].

We first present the IPS approximation (in the alchemical case for simplicity, though the results can easily be extended to the reaction coordinate case), as well as convergence results of the discretized measure to the target measure. A justification through a mean-field interpretation is then proposed in Section 4.2.2. The numerical implementation of the IPS method is eventually discussed.

4.2.1 The IPS and its statistical properties

We use here the notations and definitions of Section 4.1.1. Recall that the potential of mean force defined in the alchemical case by

$$\mathcal{F}_{\lambda(t)} = \int \frac{\partial H_\lambda}{\partial \lambda}(x)\, d\mu_{\lambda(t)}(x)$$

is the average force applied to the system during an infinitely slow transformation. The first step is to rewrite the Feynman-Kac formula (4.7) by introducing a dichotomy when a replica is receiving

either excess or deficit work compared to the potential of mean force. To this end, we define respectively the excess and deficit force, and the excess and deficit work as

$$f_t^{\mathrm{ex}}(x) = \left(\frac{\partial H_{\lambda(t)}}{\partial \lambda} - \mathcal{F}_{\lambda(t)}\right)^+(x), \quad f_t^{\mathrm{de}}(x) = \left(\frac{\partial H_{\lambda(t)}}{\partial \lambda} - \mathcal{F}_{\lambda(t)}\right)^-(x)$$

$$\mathcal{W}_t^{\mathrm{ex}} = \int_0^t f_s^{\mathrm{ex}}(X_s)\lambda'(s)\, ds, \quad \mathcal{W}_t^{\mathrm{de}} = \int_0^t f_s^{\mathrm{de}}(X_s)\lambda'(s)\, ds, \tag{4.52}$$

where $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$ (so that $x = x^+ - x^-$). We then rewrite

$$\mu_{\lambda(t)}(\varphi) = \frac{\mathbb{E}\left(\varphi(X_t)\mathrm{e}^{-\beta(\mathcal{W}_t^{\mathrm{ex}} - \mathcal{W}_t^{\mathrm{de}})}\right)}{\mathbb{E}\left(\mathrm{e}^{-\beta(\mathcal{W}_t^{\mathrm{ex}} - \mathcal{W}_t^{\mathrm{de}})}\right)}. \tag{4.53}$$

We now present the particle interpretation of (4.53) enabling a numerical computation through the use of empirical distributions. Consider M Markovian systems described by variables X_t^k $(1 \le k \le M)$. We approximate the virtual force and the Boltzmann distribution by their empirical counterparts, which read respectively

$$\mathcal{F}_{\lambda(t)}^M = \frac{1}{M}\sum_{k=1}^M \frac{\partial H_{\lambda(t)}}{\partial \lambda}(X_t^k), \qquad d\mu_{\lambda(t)}^M(x) = \frac{1}{M}\sum_{k=1}^M \delta_{X_t^k}(dx).$$

This naturally gives from definitions (4.52) empirical approximations of excess/deficit forces $f_t^{M,\mathrm{ex/de}}$ and works $\mathcal{W}_t^{k,\mathrm{ex/de}}$. The replicas evolve according to a branching process with the following stochastic rules (see [289, 290] for further details):

Interacting particle system process

Process 4.1. Consider an initial distribution $(X_0^1, \ldots, X_0^M)$ generated from $d\mu_0(x)$. Generate independent times $\tau_1^{k,b}, \tau_1^{k,d}$ from an exponential law of mean β^{-1} (the upperscripts b and d refer to 'birth' and 'death' respectively), and initialize the jump times $T^{b/d}$ as $T_0^{k,d} = 0, T_0^{k,b} = 0$.

For $0 \le t \le T$,

(1) Between each jump time, evolve independently the replicas X_t^k according to the dynamics (4.2);

(2) At random times $T_{n+1}^{k,d}$ defined by

$$\mathcal{W}_{T_{n+1}^{k,d}}^{k,\mathrm{ex}} - \mathcal{W}_{T_n^{k,d}}^{k,\mathrm{ex}} = \tau_{n+1}^{k,d},$$

an index $l \in \{1, \ldots, M\}$ is picked at random, and the configuration of the k-th replica is replaced by the configuration of the l-th replica. A time $\tau_{n+2}^{k,d}$ is generated from an exponential law of mean β^{-1};

(3) At random times $T_{n+1}^{k,b}$ defined by

$$\mathcal{W}_{T_{n+1}^{k,d}}^{k,\mathrm{de}} - \mathcal{W}_{T_n^{k,d}}^{k,\mathrm{de}} = \tau_{n+1}^{k,b},$$

an index $l \in \{1, \ldots, M\}$ is picked at random, and the configuration of the l-th replica is replaced by the configuration of the k-th replica. A time $\tau_{n+2}^{k,b}$ is generated from an exponential law of mean β^{-1}.

The selection mechanism therefore favors replicas which are sampling values of the virtual work $\mathcal{W}_t$ lower than the empirical average. The system of replicas is 'self-organizing' to keep closer to a quasi-static transformation.

In [248, 289], several convergence results and statistical properties of the replicas distribution are proven. They are summarized in the following

Proposition 4.5. *Assume that* $(t,x) \mapsto \frac{\partial H_{\lambda(t)}}{\partial \lambda}(x)$ *is a continuous bounded function on* $[0,T] \times T^*\mathcal{M}$ *(or* $[0,T] \times \mathcal{M}$ *in the case of overdamped Langevin dynamics), and that the dynamics* (4.2) *is ergodic. Then for any* $t \in [0,T]$,

(i) The estimator

$$\exp\left(-\beta \int_0^t \mathcal{F}^M_{\lambda(s)} \lambda'(s)\, ds\right) \tag{4.54}$$

is an unbiased estimator of $\mathrm{e}^{-\beta(F(\lambda(t))-F(0))}$;

(ii) For all test function φ, *the estimator* $\int \varphi \, d\mu^M_{\lambda(t)}$ *is an asymptotically normal estimator of* $\int \varphi \, d\mu_{\lambda(t)}$, *with bias and variance of order* M^{-1}.

The proof follows from Lemma 3.20, Proposition 3.25 and Theorem 3.28 of [248] (see also [289, 290] for further details). The unbiased estimation of un-normalized quantities is a very usual property in particle system methods. It comes from the fundamental property that at each "time step", each replica may branch with a number of offsprings equal in average to its relative importance weight.

Let us emphasize that the sample $(X_t^k)_{1\le k\le M}$ is in particular an empirical approximation of the canonical measure $d\mu_{\lambda(t)}$ for all t, and that no exponential reweighting of the works needs to be done at the end of the simulation to obtain the free energy differences. In the case of interacting replicas, the exponential reweighting of the Jarzynski equality (4.5) is replaced by the simple average

$$\Delta \hat{F}_{\mathrm{IPS}} = \int_0^T \mathcal{F}^M_{\lambda(t)} \lambda'(t)\, dt = \frac{1}{M} \sum_{k=1}^M \int_0^T \frac{\partial H_{\lambda(t)}}{\partial \lambda}(X_t^k) \lambda'(t)\, dt,$$

which, by Proposition 4.5, is asymptotically normal with bias and variance of order M^{-1}, and the estimator $\mathrm{e}^{-\beta \Delta \hat{F}_{\mathrm{IPS}}}$ is unbiased estimator of $\mathrm{e}^{-\beta \Delta F}$. Defining the work along one trajectory as

$$\mathcal{W}_t = \int_0^T \frac{\partial H_{\lambda(t)}}{\partial \lambda}(X_t) \lambda'(t)\, dt,$$

it therefore holds in the limit $M \to +\infty$,

$$\boxed{\mathbb{E}(\mathcal{W}_t) = F(\lambda(t)) - F(0),} \tag{4.55}$$

which should be compared to (4.5). Notice however that the notion of a single trajectory is only formal and has no meaning since all trajectories interact continuously. The above equality has only a pedagogical purpose.

4.2.2 Consistency through a mean-field limit

In order to prove the consistency of the IPS approximation, we consider the ideal setting where the number of replicas goes to infinity ($M \to +\infty$). This point of view is equivalent to a mean-field or Mc Kean interpretation of the IPS (denoted by the superscript 'mf'). In this limit, the behavior of any single replica, denoted by X_t^{mf}, is then independent from any finite number of other ones. We shall consider the mean field distribution

$$\mathrm{Law}(X_t^{\mathrm{mf}}) = d\mu_t^{mf} = \mu_t^{\mathrm{mf}}(x)dx,$$

and the mean-field force

$$\mathcal{F}_t^{\mathrm{mf}} = \int \frac{\partial H_{\lambda(t)}}{\partial \lambda} d\mu_t^{\mathrm{mf}}.$$

The associated mean field excess/deficit force $f_t^{\mathrm{mf,ex/de}}$ and works $\mathcal{W}_t^{\mathrm{mf,ex/de}}$ are defined as in (4.52). In view of Process 4.1, the stochastic process X_t^{mf} is a jump-diffusion process which evolves according to the following stochastic rules:

MEAN-FIELD JUMP-DIFFUSION PROCESS

Process 4.2. Generate X_0^{mf} from $d\mu_0(x)$. Generate idependent clocks $(\tau_n^b, \tau_n^d)_{n\geq 1}$ from an exponential law of mean β^{-1}, and initialize the jump times $T^{b/d}$ as $T_0^d = 0, T_0^b = 0$.
For $0 \leq t \leq T$,

(1) Between each jump time, $t \mapsto X_t^{\mathrm{mf}}$ evolves according to the dynamics (4.2);

(2) At random times T_{n+1}^d defined by

$$\mathcal{W}_{T_{n+1}^d}^{\mathrm{mf,ex}} - \mathcal{W}_{T_n^d}^{\mathrm{mf,ex}} = \tau_{n+1}^d,$$

the process jumps to a configuration x, chosen according to the probability measure $d\mu_{T_{n+1}^d}^{\mathrm{mf}}(x)$;

(3) At random times T_{n+1}^b defined by

$$\mathbb{E}(\mathcal{W}_t^{\mathrm{mf,de}})|_{t=T_{n+1}^b} - \mathbb{E}(\mathcal{W}_t^{\mathrm{mf,de}})|_{t=T_n^b} = \tau_{n+1}^b,$$

the process jumps to a configuration x, chosen according to the probability measure proportional to $f_{T_{n+1}^b}^{\mathrm{mf,de}}(x)d\mu_{\lambda(T_{n+1}^b)}(x)$.

Remark 4.4. *Note that, in the treatment of the deficit work, we take in Process 4.2 the point of view of the jumping replica; whereas in Process 4.1, we take the point of view of the attracting replica which induces a branching.*

From the above probabilistic description, we can derive the Markov generator of the mean-field process, given by the sum of a diffusion and a jump generator:

$$L_t^{\mathrm{mf}} = L_{\lambda(t)} + J_{t,\mu_t^{\mathrm{mf}}},$$

where the jump generator $J_{t,\mu_t^{\mathrm{mf}}}$ is defined as

$$J_{t,\mu_t^{\mathrm{mf}}}(\varphi)(x) = \beta\lambda'(t)\int(\varphi(y)-\varphi(x))(f_t^{\mathrm{mf,ex}}(x) + f_t^{\mathrm{mf,de}}(y))d\mu_t^{\mathrm{mf}}(y).$$

A straightforward integration gives the fundamental balance identity of the jump generator:

$$J_{t,\mu_t^{\mathrm{mf}}}^*(\mu_t^{mf}) = \beta\left(\mathcal{F}_t^{\mathrm{mf}} - \frac{\partial H_{\lambda(t)}}{\partial\lambda}\right)\lambda'(t)\mu_t^{mf}$$

which implies, by forward Kolmogorov,

$$\partial_t\mu_t^{\mathrm{mf}} = L_{\lambda(t)}^*(\mu_t^{\mathrm{mf}}) + \beta\left(\mathcal{F}_t^{\mathrm{mf}} - \frac{\partial H_{\lambda(t)}}{\partial\lambda}\right)\lambda'(t)\mu_t^{\mathrm{mf}}$$

so that finally

$$\partial_t \left(\mu_t^{\mathrm{mf}} \mathrm{e}^{-\beta \int_0^t \mathcal{F}_s^{\mathrm{mf}}\, ds} \right) = L^*_{\lambda(t)} \left(\mu_t^{\mathrm{mf}} \mathrm{e}^{-\beta \int_0^t \mathcal{F}_s^{\mathrm{mf}}\, ds} \right) - \beta \frac{\partial H_{\lambda(t)}}{\partial \lambda} \lambda'(t) \mu_t^{\mathrm{mf}} \mathrm{e}^{-\beta \int_0^t \mathcal{F}_s^{\mathrm{mf}}\, ds}.$$

The latter is exactly the forward evolution equation of the Feynamn-Kac kernel $p^w_{0,t}$ defined in (4.6), and thus $\int p^w_{0,t}(x,.) d\mu_0(x) = \mu_t^{mf} \mathrm{e}^{-\beta \int_0^t \mathcal{F}_s^{\mathrm{mf}}\, ds}$. Using (4.7), this gives the identities:

$$\mu_t^{\mathrm{mf}} = \mu_{\lambda(t)}, \quad \mathcal{F}_t^{\mathrm{mf}} = \mathcal{F}_{\lambda(t)}, \quad f_t^{\mathrm{mf,ex/de}} = f^{\mathrm{ex/de}}_{\lambda(t)}.$$

and proves the consistency of the IPS approximation scheme.

4.2.3 Numerical implementation

In the previous section, we discretized the measure by considering an empirical approximation. For a numerical implementation to be tractable, it remains to discretize the time evolution. Notice already that the IPS method induces no extra computation of the forces, and is therefore unexpensive to implement. However, although the IPS can be parallelized, the processors have to exchange informations at the end of each time step, which can slow down the simulation.

For the discretization of the dynamics, we refer to the corresponding sections in Chapter 3. It only remains to precise the discretization of the selection operation. We consider for example the following discretization of the force exerted on the k-th replica on the time interval $[i\Delta t, (i+1)\Delta t]$:

$$\frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda} = \frac{1}{2} \left(\frac{\partial H_{\lambda(i\Delta t)}}{\partial \lambda}(x^{i,k}) + \frac{\partial H_{\lambda((i+1)\Delta t)}}{\partial \lambda}(x^{i+1,k}) \right).$$

The mean force is then approximated by

$$\mathcal{F}^{M,\Delta t}_{\lambda_{i+1/2}} = \frac{1}{M} \sum_{k=1}^{M} \frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda}.$$

To get a time dicretization of the IPS, Process 4.1 is mimicked using the following rules:
- the time integrals are changed into sums;
- the selection times are defined as the first discrete times *exceeding* the exponential clocks $\tau^{b/d}$.

Further details about the numerical implementation can be found in [291]. Note that one can find more elaborate methods of discretization of the IPS (see [290]), but this one seems to be sufficient in view of the intrinsic errors introduced by the discretization of the dynamics.

4.2.4 Applications of the IPS method

Computation of canonical averages

The most obvious application of the IPS method is the computation of phase-space integrals, since an unweighted sample of all Boltzmann distributions $(\mu_{\lambda(t)})_{t\in[0,T]}$ is generated. The sample obtained can of course be improved by some additional sampling process (according to a dynamics leaving the target canonical measure invariant). This will decorrelate the replicas and may increase the quality of the sample.

We consider for example a pentane molecule, and a cooling process from $\beta = 1$ to $\beta = 2$, in the case when the Lennard-Jones interactions involve only extremal atoms in the chain, so that $\epsilon_{\mathrm{CH_3\text{-}CH_3}} = 0.29$ and $\epsilon_{\mathrm{CH_3\text{-}CH_2}} = 0$ (see Section 3.4.1 for more precisions on the model). The simulations are done as follows. We first generate an initial distribution of configurations from

the canonical measure at inverse temperature $\beta = 1$ using a classical rejection method so that no initial bias is introduced. We then first perform a bare simulated annealing from $\beta = 1$ to $\beta = 2$, using a Langevin dynamics. We then compare the resulting empirical distribution for the dihedral angles with the one arising from an IPS simulation. Figure 4.5 presents the results for $M = 10,000$, $\Delta t = 0.01$ and $T = 1$, with a linear scheme $\lambda(t) = t/T$.

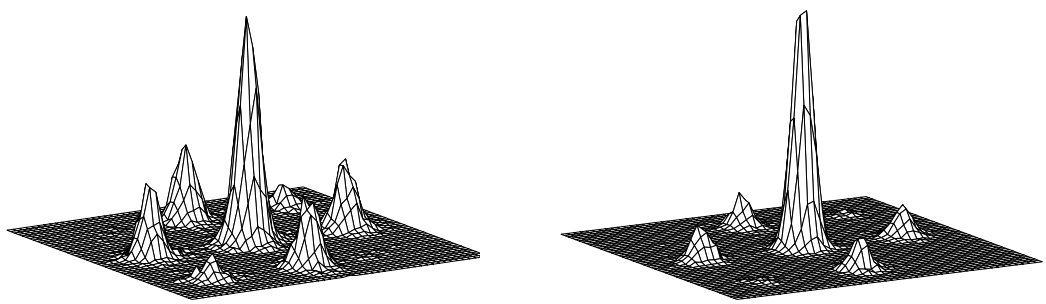


Fig. 4.5. Empirical probability distribution of the dihedral angles (ϕ_1, ϕ_2) at $\beta = 2$ of the pentane molecule generated from a sample at $\beta = 1$, using simulated annealing (Left), and IPS (Right), with sample size $M = 10,000$. The reference distribution is drawn in Figure 3.1 (Right).

As can be seen in Figure 4.5, the distribution generated with IPS is much closer to the reference distribution than the distribution generated with simulated annealing. Of course, as the time T is increased, the difference between both methods is reduced. However, this simple application shows the interest of IPS for computing distributions at low temperature starting from distributions at a higher temperature, even if the driving scheme is quite fast. This is indeed almost always the case in practice when there are several important metastable states.

Initial guesses for path sampling

The problem of free energy estimation is deeply linked with the problem of sampling meaningful transition paths (see also Section 4.3). In the IPS method, one can associate to each replica X_t^k a *genealogical continuous* path $(X_s^{k,\mathrm{gen}})_{s\in[0,t]}$. The latter is constructed recursively as follows for a replica k (for $0 \le t \le T$):

- at each time t, set $X_t^{k,\mathrm{gen}} = X_t^k$;
- at each random time T_n when the replica jumps and adopts a new configuration (say of replica l), set $(X_s^{k,\mathrm{gen}})_{[0,T_n]} = (X_s^{l,\mathrm{gen}})_{[0,T_n]}$.

This path represents the ancestor line of the replica, and is composed of the past paths selected for their low work values. For the study of the set of genealogical paths, see [247] for a discussion in the discrete time case. However, let us mention that for a given $t \in [0, T]$, the set of genealogical paths is sampled, in the limit $M \to \infty$, according to the law of the non-equilibrium paths $(X_s)_{s\in[0,t]}$ weighted by the factor $e^{-\beta \mathcal{W}_t}$ (with statistical properties analogous to those of proposition 4.5). These paths are thus typical among non-equilibrium dynamics of those with non-degenerate work. Therefore, they might be fruitfully used as non-trivial initial conditions for more specialized path sampling techniques (as e.g. [374]).

A toy example of exploration abilities

Consider the following family of Hamiltonians $(H_\lambda)_{\lambda\in[0,1]}$:

$$H_\lambda(x) = \frac{x^2}{2} + \lambda Q_1(x) + \frac{\lambda^2}{2}Q_2(x) + \frac{\lambda^3}{6}Q_3(x) + \frac{\lambda^4}{24}Q_4(x) \tag{4.56}$$

with

$$Q_1(x) = \frac{-1}{8x^2+1}, \quad Q_2(x) = \frac{-4}{8(x-1)^2+1},$$
$$Q_3(x) = \frac{-18}{32(x-3/2)^2+1}, \quad Q_4(x) = \frac{-84}{64(x-7/4)^2+1}.$$

Some of those functions are plotted in Figure 4.6. This toy one-dimensional model is reminiscent of the typical difficulties encountered when μ_0 is very different from μ_1. Notice indeed that several transitional metastable states (denoted by A and B in Figure 4.6) occur in the canonical distribution when going from $\lambda = 0$ to $\lambda = 1$. The probability of presence in the basins of attraction of the main stable states of H_1 (C and D in Figure 4.6) is only effective when λ is close to 1.

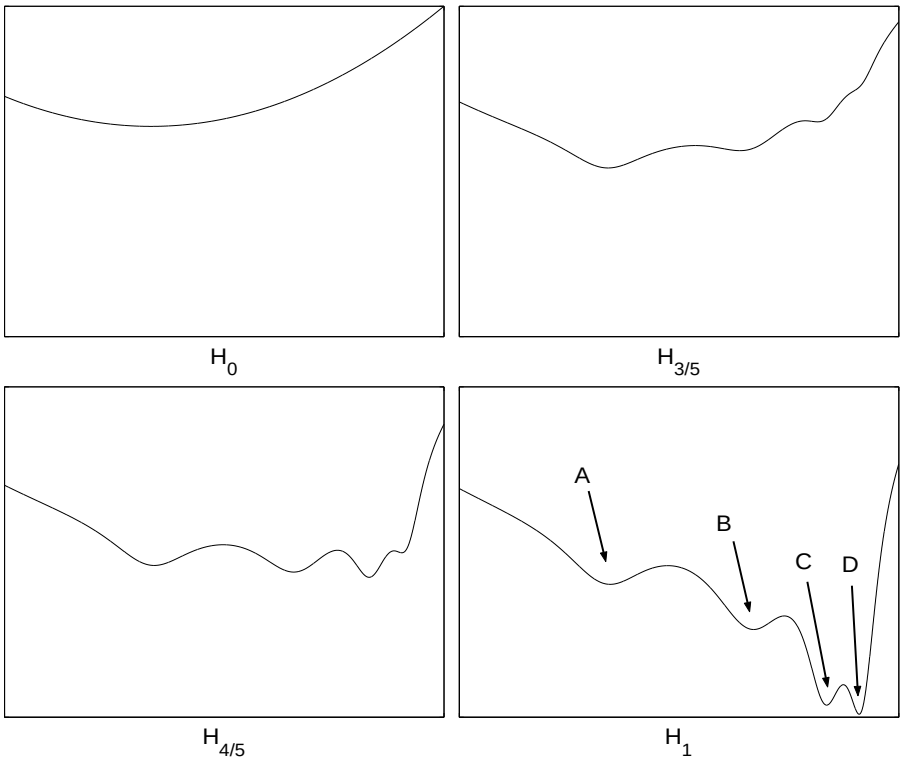


Fig. 4.6. Plot of some Hamiltonian functions, as defined by (4.56).

Simulations were performed at $\beta = 13$ with the overdamped Langevin dynamics, and the above Hamiltonian family (4.56). The number of replicas was $M = 1000$, the time step $\Delta t = 0.003$, and λ is linear: $\lambda(t) = t/T$. Figure 4.7 presents the distribution of replicas during a slow out of equilibrium plain dynamic: $T = 30$. Figure 4.8 presents the distribution of replicas during a faster dynamics with interaction: $T = 15$.

When performing a plain out of equilibrium dynamics (even 'slow') from $\lambda = 0$ to $\lambda = 1$, almost all replicas are trapped by the energy barrier of these transitional metastable states (see Figure 4.7). In the end, a very small (almost null) proportion of replicas have performed interesting

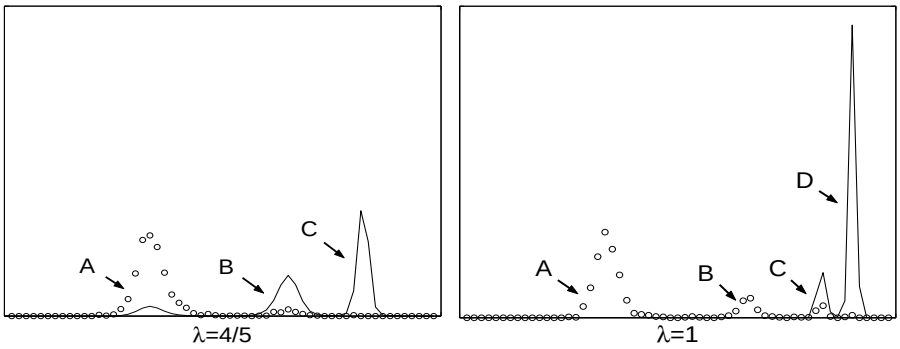


Fig. 4.7. Empirical densities (in dots) obtained using independant replicas.

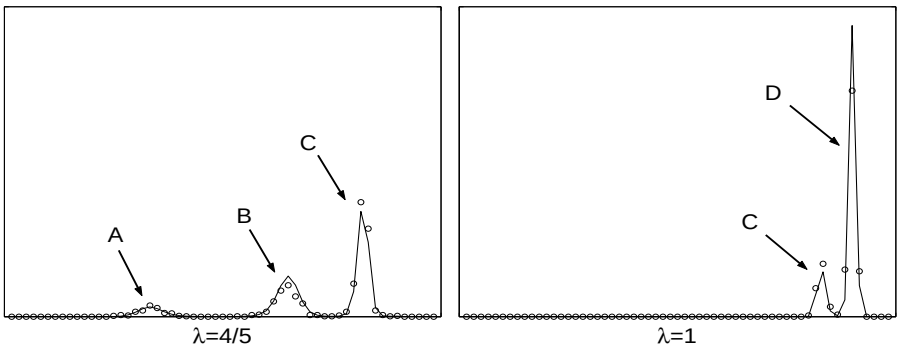


Fig. 4.8. Empirical densities (in dots) obtained using interacting replicas.

paths associated with low values of virtual work $\mathcal{W}$. When using (4.7) to compute thermodynamical quantities, these replicas bear almost all the weight of the degenerate sample, in view of the exponential weighting. The quality of the result therefore depends crucially on these rare values.

On the contrary, in the interacting version, the replicas can perform jumps in the configuration space thanks to the selection mechanism, and go from one metastable basin to another. In our example, as new transition states appear, only few clever replicas are necessary to attract the others in good areas (see Figure 4.8). In the end, all replicas have the same weight, and the sample is not degenerate. Notice also that the final empirical distribution is fairly close to the theoretical one.

We have also made a numerical estimation of the error of the free energy estimation, with 40 realizations of the above simulation. The results are presented in Table 4.2, and show an important reduction of standard deviation and bias up to a factor 2 when using the IPS method.

Table 4.2. Error in free energy estimation.

Method	Bias	Variance
Plain	+0.25	0.19
Interacting	+0.15	0.10

Application to the computation of free energy differences

Our numerical comparisons using (4.55) often turned out to give similar free energy estimations for the IPS method and the standard Jarzynski method. However, we have mostly considered the

issue of pure energetic barriers, where the difficulty of sampling comes from overcoming a *single high* barrier. The observed numerical equivalence may be explained by the fact that the selection mechanism in the IPS method does not really help to *explore* those regions of high potential energy.

When the sampling difficulties also come from barriers of more *entropic* nature (*e.g.* a succession of very many transition states separated by low energy barriers), the IPS may improve the estimation. Indeed, the selection mechanism helps keeping a statistical amount of replica in the areas of high probability with respect to the local Boltzmann distribution μ_λ throughout the switching process (see the numerical example in testing the explotation ability). This relaxation property may be crucial to ensure at each time a meaningful exploration ability.

Gradual Widom insertion

We present here an application to the computation of the chemical potential of a soft sphere fluid. This example was considered in [156, 261] for example. We consider a two-dimensional (2D) fluid of volume $|\Omega|$, simulated with periodic boundary conditions, and formed of N particles interacting via a pairwise potential V. The chemical potential is defined, in the NVT ensemble, as

$$\mu = \frac{\partial F}{\partial N},$$

where F is the free-energy of the system. Actually, the kinetic part of the partition function Z can be straightforwardly computed, and accounts for the ideal gas contribution μ_{id}. In the large N limit, the chemical potential can be rewritten as [113]

$$\mu = \mu_{\mathrm{id}} + \mu_{\mathrm{ex}},$$

with

$$\mu_{\mathrm{id}} = -\beta^{-1} \ln\left(\frac{|\Omega|}{(N+1)\Lambda^3}\right),$$

where Λ is the "thermal de Broglie wavelength" $\Lambda = h(2\pi m\beta^{-1})^{-1/2}$ (with h Planck's constant). The excess part μ_{ex} is

$$\mu_{\mathrm{ex}} = -\beta^{-1} \ln\left(\frac{\int_{\Omega^{N+1}} \exp(-\beta V(q^{N+1}))\, dq^{N+1}}{|\Omega| \int_{\Omega} \exp(-\beta V(q^N))\, dq^N}\right),$$

where $V(q^N)$ is the potential energy of a fluid composed of N particles. We restrict ourselves to pairwise interactions, with an interaction potential Φ. Then, $V(q^N) = \sum_{1\le i<j\le N} \Phi(|q_i - q_j|)$. Setting $\pi(q^N) = Z^{-1} \exp(-\beta V(q^N))$ (with $Z = \int_{\Omega^N} \exp[-\beta V(q^N)]\, dq^N$) and $\Delta V(q^N, q) = V(q^{N+1}) - V(q^N)$ with $q^{N+1} = (q^N, q)$, it follows

$$\mu_{\mathrm{ex}} = -\beta^{-1} \ln\left(\frac{1}{|\Omega|} \int_{\Omega} \mathrm{e}^{-\beta \Delta V(q, q^N)} d\pi(q^N)\, dq\right). \tag{4.57}$$

The formula (4.57) can be used to compute the value of chemical potential using stochastic methods such as the free energy perturbation (FEP) method [380]. In this case, we first generate a sample of configurations of the system according to π, and then evaluate the integration in the remaining q variable by drawing positions q of the remaining variable uniformly in Ω.

Another possibility is to use fast growth methods, resorting to the following parametrization

$$H_\lambda(q^{N+1}, p^{N+1}) = \sum_{i=1}^{N+1} \frac{p_i^2}{2m} + V_\lambda(q^{N+1}) = \sum_{i=1}^{N+1} \frac{p_i^2}{2m} + V(q^N) + \lambda \Delta V(q^N, q).$$

In this case, the interactions of the remaining particle with the N first ones are progressively turned on.

As in [156,261], we use a smoothed Lennard-Jones potential in order to avoid the singularity at the origin (Let us however note that, once the particle is inserted, it is still possible to change all the potentials to Lennard-Jones potentials, and compute the correponding free-energy difference). The Lennard Jones potential reads here

$$\Phi_{\mathrm{LJ}}(r) = 2\epsilon \left(\frac{1}{2} \left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right),$$

and the modified potential is

$$\Phi(r) = \begin{cases} a - br^2, & 0 \leq r \leq 0.8\,\sigma, \\ \Phi_{\mathrm{LJ}}(r) + c(r - r_{\mathrm{c}}) - d, & 0.8\,\sigma \leq r \leq r_{\mathrm{c}}, \\ 0, & r \geq r_{\mathrm{c}}. \end{cases}$$

The values a, b, c are chosen so that the potential is C^1. The distance r_{c} is a prescribed cut-off radius. We consider the insertion of a particle in a 2D fluid of 25 particles, at a density $\rho\sigma^3 = 0.8$, with $r_{\mathrm{c}} = 2.5\,\sigma$, $\beta\epsilon = 1$, $\Delta t = 0.0005$, and a schedule $\lambda(t) = t/T$ where T is the transition time. The results are presented in Table 4.3, for different transitions times, but at a fixed computational cost, since MT is constant. Some work distributions are also depicted in Figure 4.9. A reference value was computed using FEP, with 10^8 insertions, done by running $M = 10^3$ independent Langevins dynamics for the system composed of N particles, for a time $t_{\mathrm{FEP}} = 50$ (after an initial thermalization time to decorrelate the systems), and inserting one particle at random after each time-step. The reference value obtained is $\mu_{\mathrm{ex}} = 1.32\ k_{\mathrm{B}}T$ ($\pm 0.01\ k_{\mathrm{B}}T$).

Table 4.3. Free energy estimation for one realization of each method, depending on the switching time T and the number of replicas M used, keeping MT constant. The results are averaged over 10 realizations, and are presented under the form $< \mu >$ $(\sqrt{\mathrm{Var}(\mu)})$. The reference value obtained through FEP is $\mu_{\mathrm{ex}} = 1.32\ k_{\mathrm{B}}T$ ($\pm 0.01\ k_{\mathrm{B}}T$). Notice that the results are quite comparable.

Method	$M = 10^5$ $T = 1$		$M = 5 \times 10^4$ $T = 2$		$M = 2 \times 10^4$ $T = 5$		$M = 10^4$ $T = 10$	
Jarzynski	1.31	(0.015)	1.33	(0.017)	1.32	(0.023)	1.32	(0.038)
IPS	1.37	(0.025)	1.35	(0.040)	1.33	(0.033)	1.32	(0.037)

As can be seen from the results in Table 4.3, the IPS algorithm has a comparable accuracy to Jarzynski's estimates provided the switching time is long enough. However, the work distribution is very different, and has a stable gaussian shape for all switching rates considered, whereas the work distribution obtained through the fast growth method are much wider (see in particular Figure 4.9, Left), so that the relevant part of the work distribution (the lower tail) is only of small relative importance.

4.3 Path sampling techniques

The Transition Path Sampling (TPS) formalism, first proposed in [272] and further developped in [80] (see also [34,81] for extensive reviews), is a strategy to sample only those paths that lead to a transition between metastable states. It also gives some information on the transition kinetics, such as the rate constant as a function of time or the activation energies [78]. Recent practical and theoretical developments (such as Transition Interface Sampling [355,356]) are still aiming at

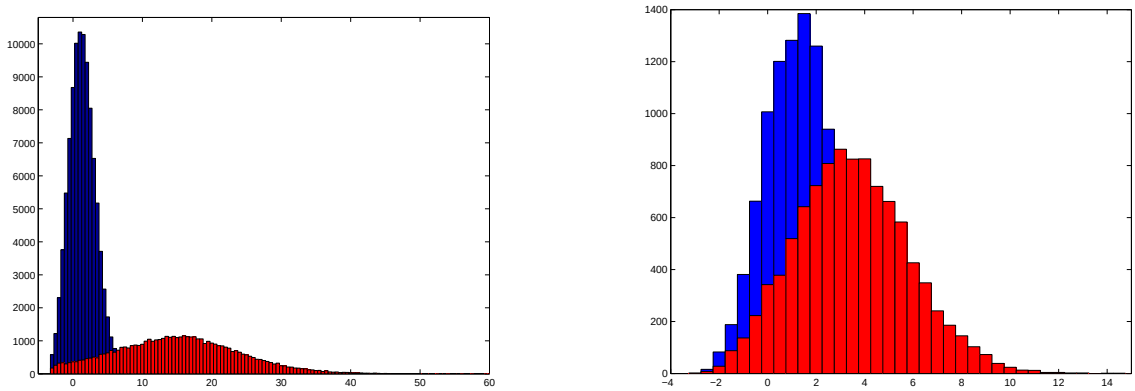


Fig. 4.9. Left: Comparison of the work distribution for $T = 1$. Right: Comparison of the work distributions for $T = 10$. The IPS results appear in darker colors. The target value is 1.32 $k_{\mathrm{B}}T$. Notice that the IPS work distribution is Gaussian with low variance even for the fast switching simulation.

increasing the power of the method. State of the art applications of path sampling, such as [189], now involve as much as 3,000 atoms with paths about 3 ns long.

Recently, relying on the Jarzynski formula [186, 187] (see also Section 4.1), path sampling techniques have also been used to compute free energy differences more efficiently [261, 331, 374] by precisely enhancing the paths that have the larger weights (which correspond to the unlikely lower work values). More precisely,

$$\mathrm{e}^{-\beta \Delta F} = \frac{\displaystyle\int \mathrm{e}^{-\beta \mathcal{W}(x)}\, d\pi_L(x)}{\displaystyle\int d\pi_L(x)},$$

where $d\pi_L$ is a measure on a discrete path of length L, and $\mathcal{W}(x)$ is the work along a given path x. In the case of the overdamped Langevin dynamics (3.38) with $\lambda(t) = t/(L\Delta t)$, the probability to observe the path $x = (q_0, q_{\Delta t}, \ldots, q_{L\Delta t})$ is

$$d\pi_L(x) = Z_L^{-1}\, \mathrm{e}^{-\beta V_0(q_0)} \prod_{i=1}^{L} \exp\left(-\frac{\beta}{4\Delta t} \left| q_{(i+1)\Delta t} - q_{i\Delta t} - \Delta t \nabla V_{i/L}(q_{i\Delta t}) \right|^2 \right) dx,$$

and the work is approximated by

$$\mathcal{W}_{\Delta t}(x) = \frac{1}{L} \sum_{i=1}^{L} \left. \frac{\partial V_\lambda}{\partial \lambda} \right|_{\lambda = i/L} (q_{i\Delta t}).$$

Importance sampling techniques can then be used, such as rewriting

$$\mathrm{e}^{-\beta \Delta F} = \frac{\displaystyle\int \mathrm{e}^{-\beta \mathcal{W}(x)/2}\, d\Pi_L(x)}{\displaystyle\int \mathrm{e}^{\beta \mathcal{W}(x)/2}\, d\Pi_L(x)},$$

where the paths are sampled according to the modified measure $d\Pi_L(x) = \mathrm{e}^{-\beta \mathcal{W}(x)/2} d\pi_L(x)$, which enhances the paths with lower work values. Methods to sample paths can be found in [34, 81, 325].

Many path sampling studies (especially TPS studies) have used deterministic dynamics (Path sampling in the NVE ensemble has already been thoroughly studied, see [81] for a review). However, path sampling with stochastic dynamics is of great interest for nonequilibrium simulations [74].

Besides, some models are stochastic by nature (see *e.g* [5] where the authors consider a model system of protein pulling in implicit solvent, and a chemical reaction simulated with kinetic Monte Carlo). Finally, we believe that there is room for improvement in the path sampling techniques for stochastic dynamics. We therefore restrict ourselves to the stochastic setting in this section.

To this date, the usual equilibrium sampling of paths with stochastic dynamics is done either with the usual shooting dynamics inspired from the corresponding algorithm for deterministic paths [81]; or with the so-called "noise history" algorithm introduced in [74], which relies on the description of paths as a starting point and the sequence of random numbers used to generate the trajectory. It is one of our aims here to relate both strategies and generalize them by introducing a new way to propose paths: namely by generating random numbers correlated with the ones used to generate the previous path. When the correlation is zero, the usual shooting dynamics is recovered. When the correlation is one everywhere except for some index along the path where it is zero, the noise-history algorithm is recovered. This generalization may be useful for example when the dynamics are too diffusive (Langevin dynamics in the high friction limit) since the shooting dynamics are inefficient in this limit; or to enhance the decorrelation of the paths generated using the noise history algorithm.

We also consider nonequilibrium sampling of paths, using some switching dynamics on paths [122], inspired from the Jarzynski out-of-equilibrium switching in phase-space [186, 187]. This switching can be performed whatever the underlying dynamics on paths. It can be used to transform a sample of unconstrained paths to reactive paths (ending up in some given region). This approach was already followed in [122], and allows to compute rate constants. However, the final sample of paths is very degenerate, and cannot be used as a reliable equilibrium sample of reactive paths. In the same vein, one could imagine doing simulated annealing on paths (simulated tempering on paths has already been investigated in [363]), in order to obtain typical transition paths at temperatures where direct sampling is not feasible. However, unless the annealing process is very slow, the final sample is usually not correctly distributed. We therefore also present the application to path sampling of the IPS birth/death process of Section 4.2. The corresponding reequilibration is of paramount importance for the end sample to be distributed according to the canonical measure on paths. Besides, since the sample of paths follows the canonical distribution at all times, the properties of interest can be computed in a single simulation for a whole range of values. For example, the rate constant could be obtained for a whole range of temperatures, which allows to compute the activation energy following the method presented in [78].

This section is organized as follows. We first present the path ensemble in Section 4.3.1, and turn to equilibrium sampling of paths in Section 4.3.2. We introduce in particular in Section 4.3.2 the "brownian tube" proposal function which generalizes the previous algorithms for path sampling with stochastic dynamics, and compare this new proposal functions to the previous ones using some two-level sampling indicators. Finally, we present in Section 4.3.3 the switching dynamics on paths, with the IPS extension enabling a reequilibration of the paths distribution at all times, even when the switching is done at a finite rate.

4.3.1 The path ensemble with stochastic dynamics

The canonical measure on discretized paths

We consider a system of N particles, with mass matrix $M = \mathrm{Diag}(m_1, \ldots, m_N)$, described by a configuration variable $q = (q_1, \ldots, q_N)$, and a momentum variable $p = (p_1, \ldots, p_N)$. The dimension of the space is denoted by d, so that $q_i, p_i \in \mathbb{R}^d$ for all $1 \leq i \leq N$. We consider stochastic dynamics of the form

$$dX_t = b(X_t)\, dt + \Sigma dW_t, \tag{4.58}$$

where the variable X_t represents either the configurational part q_t, or the full phase space variables (q_t, p_t). The function b is the force field, the matrix Σ is the magnitude of the random forcing, and W_t is a standard Brownian motion (the dimension of W_t depending on the dynamics used).

We restrict ourselves in this study to the most famous stochastic dynamics used in practice, namely the Langevin dynamics

$$\begin{cases} dq_t = M^{-1}\, p_t\, dt, \\ dp_t = -\nabla V(q_t)\, dt - \gamma M^{-1} p_t\, dt + \sigma\, dW_t, \end{cases} \tag{4.59}$$

where W_t denotes a standard dN-dimensional Brownian motion, and with the fluctuation-dissipation relation $\sigma^2 = 2\gamma/\beta$. In this case, the variable $x = (q, p)$ describes the system and the energy is given by the Hamiltonian $E(x) = H(q, p) = V(q) + \frac{1}{2} p^T M^{-1} p$. Some studies (see *e.g.* [374]) however resort to the overdamped Langevin dynamics

$$dq_t = -\nabla V(q_t)\, dt + \sqrt{\frac{2}{\beta}}\, dW_t,$$

in which case $x = q$ and $E(x) = V(q)$. The ideas presented in the sequel can of course be straightforwardly extended to this case.

In practice, the dynamics have to be discretized. Considering a time step Δt and a trajectory length $T = L\Delta t$, a discrete trajectory is then defined through the sequence

$$x = (x_0, \ldots, x_L).$$

Its weight is

$$\pi(x) = Z_L^{-1} \rho(x_0) \prod_{i=0}^{L-1} \mathrm{p}(x_i, x_{i+1}), \tag{4.60}$$

where $\rho(x_0) = Z_0^{-1} \mathrm{e}^{-\beta E(x_0)}$ is the Boltzmann weight of the initial configuration, $\mathrm{p}(x_i, x_{i+1})$ is the probability that the system is in the state x_{i+1} conditionally that it starts from x_i, and Z_L is a normalization constant. This conditional probability depends on the discretization of the dynamics used.

Denoting by $\mathbf{1}_{\mathrm{A}}(x), \mathbf{1}_{\mathrm{B}}(x)$ the indicator functions of some sets A, B defining respectively the initial and the final states, the probability of a given *reactive* path between the sets A and B is then

$$\pi_{AB}(x) = Z_{AB}^{-1} \mathbf{1}_{\mathrm{A}}(x_0) \rho(x_0) \prod_{i=0}^{L-1} \mathrm{p}(x_i, x_{i+1}) \mathbf{1}_{\mathrm{B}}(x_L). \tag{4.61}$$

Transition Path Sampling [80, 81] aims at sampling the measure[2] π_{AB}, using in particular Monte-Carlo moves of Metropolis-Hastings type.

Discretization of the dynamics

We present here a possible discretization of the Langevin dynamics, and the corresponding transition probability $\mathrm{p}(x_i, x_{i+1})$. This discretization, called "Langevin Impulse" [310], relies on an operator splitting technique, and is more appealing from a theoretical viewpoint than previous discretizations (such as the BBK algorithm [45], or schemes proposed in [4]). For particles of equal masses (up to a rescaling of time, $M = \mathrm{Id}$; the extension to the general case is straightforward), the numerical scheme we use here reads [310]:

[2] Notice that the measure $\pi_{AB} \equiv \pi_{AB}^{L,\Delta t}$ depends in fact explicitely on the length of the paths, and of the time steps used in practice. See [147] for a continuous formulation using SPDEs. In this case, the measure on paths is formulated at a continuous level.

$$\begin{cases} p_{i+1/2} &= p_i - \dfrac{\Delta t}{2}\nabla V(q_i), \\ q_{i+1} &= q_i + c_1\, p_{i+1/2} + U_{1,i}, \\ p_{i+1} &= c_0\, p_{i+1/2} - \dfrac{\Delta t}{2}\nabla V(q_{i+1}) + U_{2,i}, \end{cases} \tag{4.62}$$

with

$$c_0 = \exp(-\gamma\Delta t), \quad c_1 = \frac{1-\exp(-\gamma\Delta t)}{\gamma}.$$

The centered gaussian random variables $(U_{1,i}, U_{2,i})$ with $U_{k,i} = (u^1_{k,i}, \ldots, u^{dN}_{k,i})$ are such that

$$\mathbb{E}\left[(u^l_{1,i})^2\right] = \sigma_1^2, \quad \mathbb{E}\left[(u^l_{2,i})^2\right] = \sigma_2^2, \quad \mathbb{E}\left[u^l_{1,i}\cdot u^l_{2,i}\right] = c_{12}\sigma_1\sigma_2,$$

with

$$\sigma_1^2 = \frac{\Delta t}{\beta\gamma}\left(2 - \frac{3 - 4\mathrm{e}^{-\gamma\Delta t} + \mathrm{e}^{-2\gamma\Delta t}}{\gamma\Delta t}\right), \quad \sigma_2^2 = \frac{1}{\beta}\left(1 - \mathrm{e}^{-2\gamma\Delta t}\right), \quad c_{12}\sigma_1\sigma_2 = \frac{1}{\beta\gamma}\left(1 - \mathrm{e}^{-\gamma\Delta t}\right)^2.$$

In practice, the random vectors $(U_{1,i}, U_{2,i})$ are computed from standard gaussian random vectors $(G_{1,i}, G_{2,i})$ with $G_{k,i} = (g^1_{k,i}, \ldots, g^{dN}_{k,i})$:

$$u^l_{1,i} = \sigma_1\, g^l_{1,i}, \quad u^l_{2,i} = \sigma_2\left(c_{12}\, g^l_{1,i} + \sqrt{1-c_{12}^2}\, g^l_{2,i}\right). \tag{4.63}$$

We will always denote by G standard gaussian random vectors in the sequel, whereas the notation U refers to non-standard gaussian random vectors.

Denoting by

$$\mathrm{d}_1 \equiv \mathrm{d}_1((q_{i+1}, p_{i+1}), (q_i, p_i)) = \left| q_{i+1} - q_i - c_1\, p_i + c_1\frac{\Delta t}{2}\nabla V(q_i)\right|,$$

$$\mathrm{d}_2 \equiv \mathrm{d}_2((q_{i+1}, p_{i+1}), (q_i, p_i)) = \left| p_{i+1} - c_0\, p_i + \frac{\Delta t}{2}\left(c_0\nabla V(q_i) + V(q_{i+1})\right)\right|,$$

the conditional probability $\mathrm{p}((q_{i+1}, p_{i+1}), (q_i, p_i))$ to be in the state $x_{i+1} = (q_{i+1}, p_{i+1})$ starting from $x_i = (q_i, p_i)$ reads

$$\mathrm{p}(x_{i+1}, x_i) = Z^{-1}\exp\left[-\frac{1}{2(1-c_{12}^2)}\left(\left(\frac{\mathrm{d}_1}{\sigma_1}\right)^2 + \left(\frac{\mathrm{d}_2}{\sigma_2}\right)^2 - 2c_{12}\left(\frac{\mathrm{d}_1}{\sigma_1}\right)\left(\frac{\mathrm{d}_2}{\sigma_2}\right)\right)\right] \tag{4.64}$$

where the normalization constant is $Z = \left(2\pi\sigma_1\sigma_2\sqrt{1-c_{12}^2}\right)^{-dN}$.

4.3.2 Equilibrium sampling of the path ensemble

The most popular way to sample paths is to resort to a Metropolis-Hastings scheme [153, 238]. Other approaches may be considered in some cases , see [81] for a review of alternative approaches. Those approaches however require some force evaluation (see *e.g.* [80] for a Langevin dynamics in phase space in the case of a toy two-dimensional problem). But the force exerted on a path is proportional to $\nabla(\ln\pi)$, and is difficult to compute in general since it requires the evaluation of second derivatives of the potential in conventional phase space.

We first precise some specifities of the Metropolis-Hastings algorithm, especially when sampling reactive paths. We then recall a usual technique to propose paths in Section 4.3.2, and generalize it in Section 4.3.2. We finally propose some benchmarks to compare the efficiencies of all these proposal functions.

Metropolis-Hastings sampling techniques for path sampling

For a general introduction to the Metropolis-Hastings scheme, we refer to Section 3.1.3. In the case of reactive paths, a study of the acceptance rate asks to decompose the acceptance/rejection procedure in two successive steps: (i) the proposition of a path starting from A and going to B; (ii) the acceptance or rejection of such a path according to the Metropolis-Hastings scheme. The difficult step is the first one, since paths bridging A and B are only a (small) subset of the whole path space. In particular, diffusive dynamics such as the overdamped Langevin dynamics are often not convenient to propose bridging paths; the situtation is however better for dynamics with some inertia, such as the Langevin dynamics. When the paths are constructed using deterministic dynamics (NVE case), some studies have shown that the optimal acceptance rate is about 40 % for the cases under consideration [81].

For path sampling with stochastic dynamics, the "shooting" proposal function is classically used [81]. However, even for moderate values of the friction coefficient γ in the Langevin dynamics, this proposal function may have low acceptance rates, especially if the dimension of the system is high or/and the barriers to cross are large. An alternative way of proposing paths, relying on the so-called "noise history" of the paths [74] (*i.e.* the sequence of random numbers used to generate the trajectory from a given starting point) is to change only one of the random numbers used and to keep the others. In this case, a high acceptance rate is expected, but the paths generated may be very correlated.

A natural generalization of both approaches is to rely on the continuity of the dynamics with respect to the random noise forcing, and to propose a new trajectory by generating new random numbers correlated with the previous one. We call this approach the "brownian tube" proposal. In this case, an arbitrary acceptance rate can be reached, and there is room for optimizing the parameters in order to really tune the efficiency of the sampling.

The shooting proposal function

The acceptance rate of the Metropolis-Hastings algorithm is

$$r(x,y) = \min\left(1, \frac{\pi(y)\mathcal{P}(y,x)}{\pi(x)\mathcal{P}(x,y)}\right).$$

The shooting technique described in [81, Section 3.1.5] consists in the three following steps, starting from a path x^n:

Shooting algorithm for path sampling

Algorithm 4.1. Starting from some initial path x^0, and for $n \geq 0$,

(1) select an index $0 \leq k \leq L$ according to discrete probabilities $(w_i)_{0\leq i\leq L}$ (for example a uniform probability distribution can be considered, unless one wants to increase trial moves starting from certain regions, for example the assumed transition region);

(2) generate a new path $(y_{k+1}, \ldots, y_L)$ forward in time, using the stochastic dynamics (4.59), with a new set of independently and identically distributed (i.i.d.) gaussian random vectors $(U_i^{n+1})_{k+1\leq j\leq L-1}$;

(3) generate a new path $(y_{k-1}, \ldots, y_0)$ backward in time, using a discretized "backward" stochastic dynamics correponding to (4.59), with a new set of i.i.d. gaussian random vectors $(\overline{U}_i^{n+1})_{0\leq j\leq k-1}$;

(4) set $x^{n+1} = y$ with probability $r(x^n, y)$, otherwise set $x^{n+1} = x^n$.

It remains however to precise how the "backward" part of the trajectory is computed in Step (3), which determines the conditional probability $\bar{\mathrm{p}}(y_{j+1}, y_j)$ to go to y_j from y_{j+1} in a backward manner. The proposition density $\mathcal{P}(x, \cdot)$ is then also determined. Indeed, The probability of generating a path $y = (y_0, \ldots, y_L)$ from x, shooting forward and backward from the k-th index, is

$$\mathcal{P}(x, y) = w_k \prod_{j=0}^{k-1} \bar{\mathrm{p}}(y_{j+1}, y_j) \prod_{j=k+1}^{L} \mathrm{p}(y_{j-1}, y_j). \tag{4.65}$$

Notice that the previous path x is present only through the term $y_k = x_k$. It then follows

$$r(x, y) = \min\left(1, \mathbf{1}_{\mathrm{A}}(y_0)\mathbf{1}_{\mathrm{B}}(y_L)c_{\mathrm{exact}}(x, y)\right),$$

with

$$c_{\mathrm{exact}}(x, y) = \frac{\rho(y_0)}{\rho(x_0)} \prod_{j=0}^{k-1} \frac{\mathrm{p}(y_j, y_{j+1})}{\bar{\mathrm{p}}(y_{j+1}, y_j)} \frac{\bar{\mathrm{p}}(x_{j+1}, x_j)}{\mathrm{p}(x_j, x_{j+1})}. \tag{4.66}$$

It is clear that, for reasonable discretizations, $P^2(x, y) > 0$ for all paths x, y of positive probability (under mild assumptions on the potential) so that the correponding Markov chain is irreducible. Since the measure (4.61) is left invariant by the dynamics (this is a classical property of Metropolis-Hastings scheme), the corresponding Markov chain is ergodic [240]. Notice also that it is enough to consider only the forward or the backward integration steps for the ergodicity to hold, as long as both have a positive probability to occur (and that the possible asymmetry in the corresponding probabilities is accounted for).

Backward integration of the trajectory

There are two ways to generate proposal paths backward in time (which are precised in specific cases in the remainder of this section), using either

(i) a *time reversal* (linked to some detailed balance property): The forward dynamics are used to generate the points y_i from y_{i+1} in a time-reversed manner. This means that variables odd with respect to time reversal (such as momenta) are inverted, and variables even with respect to time reversal (such as positions) are kept constant. Denoting by $\mathcal{S}$ the reversal operator, $\mathcal{S}y_i = y_i = q_i$ for overdamped Langevin dynamics, and $\mathcal{S}y_i = (q_i, -p_i)$ when $y_i = (q_i, p_i)$ for Langevin dynamics. The usual one-step integrator $\Phi_{\Delta t}$ is then considered to integrate the corresponding trajectory, using $\mathcal{S}^2 = \mathrm{Id}$:

$$y_i = (\mathcal{S} \circ \Phi_{\Delta t} \circ \mathcal{S})y_{i+1}$$

The time-reversed conditional probability $\bar{\mathrm{p}}_{\mathrm{TR}}(y_{i+1}, y_i)$ to go from y_i to y_{i+1} is then

$$\bar{\mathrm{p}}_{\mathrm{TR}}(y_{i+1}, y_i) = \mathrm{p}(\mathcal{S}y_{i+1}, \mathcal{S}y_i).$$

The detailed balance assumption reads

$$\rho(y_i)\,\mathrm{p}(y_i, y_{i+1}) = \rho(y_{i+1})\,\mathrm{p}(\mathcal{S}y_{i+1}, \mathcal{S}y_i).$$

When this condition is met with a good precision, some cancellations occur in the expression (4.66) of the acceptance rate [81]. In this case, the acceptance rate

$$c_{\mathrm{exact}}(x, y) \simeq c_{\mathrm{TR}}(x, y) = \frac{\rho(y_i)}{\rho(x_i)}. \tag{4.67}$$

In the case when $y_i = x_i$ (which is often the case in practice for path sampling on stochastic paths), $c_{\mathrm{TR}}(x, y) = 1$. However, as will be precised later in this section, numerical

tests suggest that the detailed balance is not always met with a good precision when the dynamics are discretized with large time steps (which is useful in order to avoid too long paths), even if it is usually the case in some mean sense for usual regimes. However, even in those cases, it may be the case that detailed balance is not fulfilled along a whole path (especially since unlikely regions of high gradients are somewhat enhanced), so that the cancellations mentionned above are not always strictly valid.

(ii) a backward integration: in this case, the change of variables $t \mapsto -t$ is done directly in the numerical scheme, so that

$$y_i = \Phi_{-\Delta t}(y_{i+1}).$$

The corresponding backward probability will be denoted by $\bar{\mathrm{p}}_{\mathrm{bck}}(y_{i+1}, y_i)$. The backward schemes are such that a reversibility condition is approximately met (since $\Phi_{\Delta t} \circ \Phi_{-\Delta t} \simeq \mathrm{Id}$)

$$\mathrm{p}(y_i, y_{i+1}) \simeq \bar{\mathrm{p}}_{\mathrm{bck}}(y_{i+1}, y_i),$$

at least in some conditions that can be precised on a specific example.

Let us emphasize that the above approximations are used in some computations to obtain simpler expression for the acceptance rate, but their validity should be carefully checked in any cases, as we now do.

Backward overdamped Langevin dynamics.

The time reversed version of the overdamped Langevin dynamics is still the usual overdamped Langevin dynamics for the Euler-Maruyama discretization

$$q_{i+1} = q_i - \Delta t \nabla V(q_i) + \sqrt{\frac{2\Delta t}{\beta}}\, R_i, \tag{4.68}$$

R_i being i.i.d. dN-dimensional random vectors. It holds

$$\mathrm{p}(q_i, q_{i+1}) = \left(\frac{\beta}{4\pi\Delta t}\right)^{dN/2} \exp\left(-\frac{\beta}{4\Delta t}|q_{i+1} - q_i + \Delta t \nabla V(q_i)|^2\right), \tag{4.69}$$

and

$$\bar{\mathrm{p}}_{\mathrm{TR}}(q_2, q_1) = \mathrm{p}(q_2, q_1). \tag{4.70}$$

Therefore, time reversed paths are generated using the discretization (4.68), and a correction has to be accounted according to (4.66). The validity of the reduced acceptance rate (4.67) can be checked by monitoring

$$R_{\mathrm{TR}} = \max\left\{\frac{c_{\mathrm{TR}}}{c_{\mathrm{exact}}}, \frac{c_{\mathrm{exact}}}{c_{\mathrm{TR}}}\right\}$$

for the reactive paths generated. Notice that the ratio $c_{\mathrm{TR}}/c_{\mathrm{exact}}$ is exactly 1 when the detailed balance assumption is strictly fulfilled, so that $R_{\mathrm{TR}} = 1$ in this case. Therefore, the validity of this assumption along the whole path is related to the magnitude of the values of $R_{\mathrm{TR}} > 1$ (since $R_{\mathrm{TR}} \geq 1$ in all cases).

The discretized backward stochastic dynamics are, for the overdamped Langevin dynamics

$$q_{i-1} = q_i + \Delta t\, \nabla V(q_i) + \sigma\, R_i, \tag{4.71}$$

with $\sigma^2 = 2\Delta t/\beta$, and where the random variables (R_i) are i.i.d. dN-dimensional standard Gaussian random vectors. Note already that the scheme (4.71) is unstable in general (except near saddle points of the energy landscape) since the sign of the force has to be changed in a backward integration, so that only small time steps must be considered. The resulting backward conditional

probability to be in q_{i-1} starting from q_i is therefore

$$\bar{p}_{\rm bck}(q_i, q_{i-1}) = \left(\frac{\beta}{4\pi\Delta t}\right)^{dN/2} \exp\left(-\frac{\beta}{4\Delta t}|q_i - q_{i-1} + \Delta t \nabla V(q_i)|^2\right). \tag{4.72}$$

The reversibility assumption, made for example in [374], can also be checked here by computing

$$R_{\rm bck} = \max\left\{\frac{c_{\rm bck}}{c_{\rm exact}}, \frac{c_{\rm exact}}{c_{\rm bck}}\right\}$$

for the reactive paths generated. The behavior of $R_{\rm bck}$ should be close to the behavior of $R_{\rm TR}$.

To test the above assumptions, we consider the following one-dimensional double well potential:

$$V(x) = 0.5h(x-1)^2(x+1)^2,$$

where h is a factor allowing to modify the barrier height at the transition state $x = 0$.

We first test the detailed balance and reversibility assumptions, for a certain range of time steps and barrier height (the inverse temperature is set to $\beta = 1$). To this end, we sample n initial configurations $(q^i)_{1\le i\le n}$ of the system according to the canonical measure (using a rejection algorithm, so that no additional bias is added to the intrinsic statistical bias arising from the finite size of the sample) and perform a realization of the one step moves using the integration scheme (4.68). We denote by $\tilde{q}^j$ the outcome for a given initial configuration q^j. We then compute the quantities

$$\langle r_{\rm DB}\rangle = \frac{1}{n}\sum_{j=1}^{n} r_{\rm DB}(q^j, \tilde{q}^j), \quad \langle r_{\rm rev}\rangle = \frac{1}{n}\sum_{j=1}^{n} r_{\rm rev}(q^j, \tilde{q}^j),$$

with

$$r_{\rm DB}(q_1, q_2) = \frac{\rho(q_1)\,{\rm p}(q_1, q_2)}{\rho(q_2)\,\bar{\rm p}_{\rm TR}(q_2, q_1)}, \quad r_{\rm rev}(q_1, q_2) = \frac{{\rm p}(q_1, q_2)}{\bar{\rm p}_{\rm bck}(q_2, q_1)},$$

where ${\rm p}, \bar{\rm p}_{\rm TR}$ and $\bar{\rm p}_{\rm bck}$ are given by (4.69), (4.70) and (4.72) respectively. We also compute the associated variances. We then turn to the path sampling algorithm, using the above mentioned shooting algorithm with a forward and a backward shooting (the dynamics being either the time reversed or the backward dynamics). The acceptance/rejection step is done using the exact rate (4.66), and the values $R_{\rm TR}$ and $R_{\rm bck}$ are computed over reactive paths of size $L = 200\,\Delta t$, with the sets $A = [-1-\delta, -1+\delta], B = [1-\delta, 1+\delta]$ with $\delta = 0.2$, and performing $n = 10^5$ iterations of the path sampling algorithm. The canonical averages $r_{\rm DB}$ and $r_{\rm rev}$ are computed using $n = 10^6$ points. The results are presented in Table 4.4.

The reversibility assumption is verified for time steps and barrier heights small enough (which is usually not the interesting range of study for path sampling). Moreover, we studied here this property from an average point of view, and it is expected that the situation will get worse when unlikely regions will be enhanced through the path sampling algorithm. Besides, even if the detailed balance is almost verified for one integration step, it is likely that the precision will deteriorate when considering successive integrations.

As can be seen from the results, the reversibility assumption along the whole path is hardly valid, except for low barriers and small time steps. Besides, it may be the case that the reversibility assumption can be considered to hold as a canonical average (*i.e.* $r_{\rm rev}$ is indeed close to 1 with a small variance), but not along a path[3]. The errors are somewhat magnified by the length of the path, and the enhancement of the high gradient regions. However, the detailed balance assumption is more easily verified in practice than the reversibility assumption. The acceptance results shows that few paths bridging initial and final states are proposed. The overdamped Langevin dynamics is too erratic to provide efficient proposals (the overall acceptance rates are 1-2% at most).

[3] See for example the case $\Delta t = 2.5 \times 10^{-3}$ with $h = 20$.

Table 4.4. Results for the reversibility and detailed balance study for the discretization (4.68) of the overdamped Langevin dynamics. All the results are presented under the form "$\langle A \rangle$ ($\sqrt{\mathrm{Var}(A)}$)".

Parameters		r_{DB}	r_{rev}	R_{TR}	R_{bck}
$\Delta t = 0.001$	$h = 0.5$	1.000 (0.0003)	1.002 (0.0060)	1.001 (0.0007)	1.040 (0.0559)
$\Delta t = 0.001$	$h = 1$	1.000 (0.0005)	1.003 (0.0096)	1.002 (0.0015)	1.096 (0.1177)
$\Delta t = 0.001$	$h = 2$	1.000 (0.0011)	1.006 (0.0163)	1.003 (0.0027)	1.157 (0.1863)
$\Delta t = 0.001$	$h = 10$	1.000 (0.0075)	1.040 (0.0770)	1.017 (0.0149)	5.864 (6.777)
$\Delta t = 0.001$	$h = 20$	1.000 (0.0186)	1.094 (0.1838)	1.044 (0.0362)	-
$\Delta t = 0.0025$	$h = 1$	1.000 (0.0021)	1.009 (0.0255)	1.006 (0.0056)	1.635 (1.640)
$\Delta t = 0.0025$	$h = 10$	1.001 (0.0307)	1.121 (0.3174)	1.084 (0.0786)	1.584×10^5 (6.047×10^5)
$\Delta t = 0.0025$	$h = 20$	1.006 (0.0800)	1.471 (22.09)	1.244 (0.2809)	-
$\Delta t = 0.005$	$h = 1$	1.000 (0.0059)	1.019 (0.0577)	1.021 (0.0230)	13.46 (153.0)
$\Delta t = 0.005$	$h = 10$	1.007 (0.0961)	1.573 (34.04)	1.363 (0.4454)	-
$\Delta t = 0.005$	$h = 20$	1.053 (0.7521)	9431 (2.930×10^6)	2.107 (1.709)	-

Langevin dynamics.

We present first a numerical study similar to the one done for the overdamped Langevin case. We do not consider backward integration using negative time steps (which is even more unstable than in the overdamped case), and limit ourselves to proposal functions for Langevin paths using the time reversed dynamics. More precisely, we use the discretization (4.73), which is a classical integration scheme [4], traditionally used in transition path sampling:

$$\begin{cases} q^{n+1} = q^n + c_1 \Delta t \; p^n - c_2 \Delta t^2 \nabla V(q^n) + W_1^n, \\ p^{n+1} = \mathrm{e}^{-\gamma \Delta t} p^n - (c_1 - c_2) \Delta t \nabla V(q^n) - c_2 \Delta t \nabla V(q^{n+1}) + W_2^n, \end{cases} \tag{4.73}$$

where the random numbers are the same as in (4.62) (only the deterministic part of the dynamics is modified). The time-reversing operation amounts to reverting the momenta, integrating forward in time, and reverting the momenta again. We also test the validity of a detailed balance assumption, both as a static property, and along paths. The computed variables r_{DB} and R_{TR} are defined as for the overdamped case.

We consider as a toy example the two-dimensional (2D) potential

$$V(x,y) = \frac{1}{6}\left[4(1-x^2-y^2)^2 + 2(x^2-2)^2 + \left((x+y)^2-1\right)^2 + \left((x-y)^2-1\right)^2\right], \tag{4.74}$$

which was introduced in [80]. The numerical study is conducted in the same manner as for the overdamped case, and the results are presented in Table 4.5. The detailed balance assumption is indeed satisfied with a very good accuracy for a broad range of parameters regimes. The detailed balance along paths is also satisfied with a good accuracy, though discrepancies of the static detailed balance study are still somewhat magnified, and it could be the case in some more complicated situations (such as higher dimensional dynamics with constraints) that those discrepancies become non negligible. Further numerical studies suggest that the most influential parameter is the time step Δt.

We also tested those assumptions on the model system for conformational changes of Section 4.1.4. The canonical averages r_{DB} are computed using $n = 10^5$ iterations. The values R_{TR} are computed over reactive paths of size $L = 500\,\Delta t$, at $\beta = 1$, using $l_0 = 1.3$, $\sigma = 1$, $\epsilon = 1$, $w = 0.5$, $\Delta t = 0.0025$, with the sets $A = \{r(q) \leq r_0 + 0.6\sigma\}, B = \{r(q) \geq r_0 + 1.4\sigma\}$, and performing $n = 10^4$ iterations of the path sampling algorithm.

Table 4.5. Results for the detailed balance study for the discretization (4.73) of the Langevin dynamics. The canonical averages r_{DB} are computed using $n = 10^6$ points. The values R_{TR} are computed over reactive paths of size $L = 200\,\Delta t$, with the sets $A = \{|x+1|^2 + y^2 \leq \delta\}, B = \{|x-1|^2 + y^2 \leq \delta\}$ with $\delta = 0.6$, and performing $n = 10^5$ iterations of the path sampling algorithm. All the results are presented under the form "$\langle A \rangle$ ($\sqrt{\mathrm{Var}(A)}$)".

Parameters	r_{DB}	R_{TR}
$\Delta t = 0.02$, $\xi = 1$, $\beta = 1$	1.000 (0.0002)	1.002 (0.0024)
$\Delta t = 0.01$, $\xi = 1$, $\beta = 10$	1.000 (0.0000)	1.001 (0.0014)
$\Delta t = 0.025$, $\xi = 5$, $\beta = 5$	1.000 (0.0004)	1.004 (0.0033)
$\Delta t = 0.05$, $\xi = 2$, $\beta = 20$	1.000 (0.0023)	1.022 (0.0180)

Table 4.6. Results for the detailed balance study for the discretization (4.62) of the Langevin dynamics in the WCA case. All the results are still presented under the form "$\langle A \rangle$ ($\sqrt{\mathrm{Var}(A)}$)".

Parameters	r_{DB}	R_{TR}
$h = 1$	1.0000 (0.0031)	1.002 (0.0653)
$h = 2$	1.0000 (0.0031)	1.002 (0.0721)
$h = 5$	1.0000 (0.0032)	1.003 (0.0772)

Once again, as can be seen from the results of Table 4.6, the detailed balance assumption holds in average with a very good accuracy, but there are noticeable deviations from the detailed balance assumption along the paths.

Time-reversal as a backward integration scheme

In conclusion, the previous results show that it is more appropriate to resort to *time reversal*. We will always denote in the sequel the random vectors used in this process by $\bar{U}$. As also shown in the previous computations, the microscopic reversibility ratio

$$R_{\mathrm{rev}}(y_i, y_{i+1}) = \frac{\rho(y_i)\,\mathrm{p}(y_i, y_{i+1})}{\rho(y_{i+1})\,\bar{\mathrm{p}}(y_{i+1}, y_i)}$$

is sometimes close to 1, so that $c_{\mathrm{exact}}(x, y) \simeq 1$ and the acceptance/rejection step is greatly simplified. However, this assumption should always be checked carefully using some preliminary runs since it is sometimes the case that, even if the reversibility ratio r_{DB} is close to 1 pointwise (with a good approximation), it may be false that $c_{\mathrm{exact}}(x, y) \simeq 1$ along the path, especially if the paths are long.

The brownian tube proposal function

A path can also be characterized uniquely by the initial point x_0 and the realization of the brownian process W_t in (4.58). When discretized, the paths are then uniquely determined by the sequence of gaussian random vectors $U = (U_0, \ldots, U_{L-1})$ used to generate the trajectories using (4.62) (or any discretization of another SDE). This was already noted in [74], where a new trajectory was proposed selecting an index at random and changing only the gaussian random number associated with this index.

Since the trajectory is continuous with respect to the realizations of the brownian motion, any convenient small perturbation of the sequence of random vectors is expected to generate a path close to the initial path. Still denoting by $\mathrm{p}(x_i, x_{i+1})$ the probability to generate a point x_{i+1} in phase-space starting from x_i, using the gaussian random vectors U_i and $\bar{U}_i$ obtained from standard gaussian random vectors G_i and $\bar{G}_i$, the transition probabilities for all classical discretizations we consider can be writtten as

$$\mathrm{p}(x_i, x_{i+1}) = Z^{-1} \exp\left(-\frac{1}{2} G_i^T \Gamma G_i\right),$$

and

$$\bar{\mathrm{p}}_{\mathrm{TR}}(x_{i+1}, x_i) = Z^{-1} \exp\left(-\frac{1}{2} \bar{G}_i^T \Gamma \bar{G}_i\right)$$

where Z is a normalization constant. In the case of the discretization (4.62) of the Langevin equation for example, $\Gamma = V^T V$ where the matrix V allows to recast the correlated gaussian random vectors $U_i = (U_{1,i}, U_{2,i})$ (or $\bar{U}_i$) as standard and independent gaussian random vectors G_i (or $\bar{G}_i$) through the transformation $U_i = V G_i$ (or $\bar{U}_i = V \bar{G}_i$) with (see Eq. (4.64))

$$V = \begin{pmatrix} \sigma_1^{-1} \mathrm{Id}_{dN} & 0 \\ \dfrac{c_{12}}{\sigma_1 \sqrt{1 - c_{12}^2}} \mathrm{Id}_{dN} & \dfrac{1}{\sigma_2 \sqrt{1 - c_{12}^2}} \mathrm{Id}_{dN} \end{pmatrix}.$$

The idea is then to modify the standard gaussian vectors G_i by an amount $0 \leq \alpha_i \leq 1$ as

$$\tilde{G}_i = \alpha_i G_i + \sqrt{1 - \alpha_i^2} R_i, \tag{4.75}$$

where R_i is a $2dN$-dimensional standard gaussian random vector. A fraction α_i is associated with each configuration x_i along the path. The usual shooting dynamics is recovered with $\alpha_i = 0$ for all i (all the Brownian increments are uncorrelated with respect to the Brownian increments of the modified path), whereas the so-called 'noise history' algorithm proposed in [74] corresponds to $\alpha_i = 0$ for all i but one i_0 for which $\alpha_{i_0} = 1$ (in this case, all the Brownian increments but one are re-used).

The dynamics we propose looks like the shooting dynamics:

Brownian tube proposal

Algorithm 4.2. Starting from some initial path x^0, and for $n \geq 0$,

(1) select an index $0 \leq k \leq L$ according to discrete probabilities $(w_i)_{0 \leq i \leq L}$ (for example a uniform probability distribution can be considered, unless one wants to increase trial moves starting from certain regions, for example the assumed transition region);

(2) compute a new random gaussian vector starting from the previous one, using (4.75);

(3) generate a new path $(y_{k+1}, \ldots, y_L)$ forward in time, using the stochastic dynamics (4.59), with a new set of independently and identically distributed (i.i.d.) gaussian random vectors $(U_i^{n+1})_{k+1 \leq j \leq L-1}$;

(4) generate a new path $(y_{k-1}, \ldots, y_0)$ backward in time, using a discretized "backward" stochastic dynamics correponding to (4.59), with a new set of i.i.d. gaussian random vectors $(\overline{U}_i^{n+1})_{0 \leq j \leq k-1}$;

(5) set $x^{n+1} = y$ with probability $r(x^n, y)$, otherwise set $x^{n+1} = x^n$.

It remains to precise the proposition function $\mathcal{P}(x, y)$. Denoting by $(\bar{G}_i^x)_{0 \leq i \leq k-1}$, $(G_i^x)_{k \leq i \leq L-1}$ the standard random gaussian vectors associated with the path x (the first ones arise from the time reversed integration, the last ones from a usual foward integration), it follows

$$\mathcal{P}(x, y) = w_k \prod_{0 \leq i \leq k-1} \mathrm{p}_{\alpha_i}(\bar{G}_i^x, \bar{G}_i^y) \prod_{k \leq i \leq L-1} \mathrm{p}_{\alpha_i}(G_i^x, G_i^y),$$

where w_k still denotes the probability to choose k as a shooting index, and

$$\mathrm{p}_\alpha(G,\tilde{G}) = \left(\frac{1}{\sqrt{2\pi(1-\alpha^2)}}\right)^{dN} \exp\left(-\frac{(\tilde{G}-\alpha G)^T(\tilde{G}-\alpha G)}{2(1-\alpha^2)}\right).$$

A tuning of the coefficients α_i can then be performed in order to get the best trade-off between acceptance (which tends to 1 in the limit $\alpha_i = 1$ for all i) and decorrelation (which arises in the limit $\alpha_i \to 0$). An interesting idea could be that α has to be close to 1 in regions where the generating moves have a chaotic behavior (in the sense that even small perturbations to a path lead to large changes to this path), and could be smaller in regions where the generating moves have less impact on the paths (so as to increase the decorrelation). From a more practical point of view, a possible approache to obtain such a trade-off to propose a functional form for the coefficients α_i and to perform short computations to optimize the parameters with respect to some objective function. Some simple choices for the form of the coefficients α_i, involving only one parameter (so that the optimization procedure is easier), are:

(i) constant coefficients $\alpha_i = \alpha$;

(ii) set $\alpha_i = 1$ far from the shooting index, and α_i close to 0 near the shooting index. This can be done by considering $\alpha_i = \min(1, K|i-k|)$ for some $K \geq 0$.

From our experience, the efficiency is robust enough with respect to the choice of the coefficients α_i. Notice also that the second functional form allows to recover both the usual shooting and the noise-history algorithm, respectively in the regimes $K \to 0$ and $K \geq 1$. It is therefore expected that, optimizing the efficiency with respect to $K \in [0,1]$, both the shooting algorithm and the noise-history algorithm should be outperformed.

Intrinsic measure of efficiency

Our aim here is to propose some abstract measure of decorrelation between the paths, so as to measure some diffusion in path space. This approach complements the convergence tests based on some observable of interest for the system. We refer to [81] for some examples of relevant quantities to monitor (and applications to path sampling with deterministic dynamics).

The intrinsic decorrelation is related to the existence of some distance or norm on path space. Given a distance function $\mathrm{d}(x,y)$, the quantity

$$D_p(n) = \left(\int\int [\mathrm{d}(y,x)]^p \; P^n(x,dy)\, d\pi(x)\right)^{1/p}$$

(with $p \geq 1$) precises the *average* amount of decorrelation with respect to the distance d for the measure π on the path ensemble. Notice that two averages are taken: one over the initial paths x, and another over all the realizations of the Monte Carlo iterations starting from x (*i.e.* over all the possible end paths y, weighted by the probability to end up in y starting from x). In practice, assuming ergodicity, $D_p(n)$ is computed as

$$D_p(n) = \lim_{N\to+\infty}\left(\frac{1}{N}\sum_{k=1}^{N} \mathrm{d}^p(x^{k+n}, x^k)\right)^{1/p}.$$

Usual choices for p are $p=1$ or $p=2$. This last case is considered in [59] since a diffusive behavior over the space is expected with stochastic dynamics, the most efficient algorithms having the largest diffusion constants $\lim_{n\to+\infty}\sqrt{D_2(n)/n}$.

It then only remains to precise the distance d, which depends on the system of interest. Some simple choices are to

(i) consider a (weighted) norm $||\cdot||$ on the whole underlying phase-space (for position or position/momenta variables) and set

$$\mathrm{d}(x,y) = \left(\frac{1}{L} \sum_{i=0}^{L} \omega_i ||x_i - y_i||^{p'} \right)^{1/p'}$$

with $p' \geq 1$;

(ii) consider only a projection of the configurations onto some submanifold, such as the level sets of a given (not necessarily completely relevant) reaction coordinate or order parameter ξ:

$$\mathrm{d}(x,y) = \left(\frac{1}{L} \sum_{i=0}^{L} \omega_i |\xi(x_i) - \xi(y_i)|^{p'} \right)^{1/p'},$$

with $p' \geq 1$.

(iii) align the paths projected onto some submanifold around a given value of the reaction coordinate ξ:

$$\mathrm{d}(x,y) = \left(\frac{1}{2K+1} \sum_{i=-K}^{K} \omega_i |\xi(x_{I+i}) - \xi(y_{J+i})|^{p'} \right)^{1/p'}, \tag{4.76}$$

with $p' \geq 1$, and I, J such that $\xi(x_I) = \xi(y_J) = \xi^*$ where ξ^* is fixed in advance (for example, if A is characterized by $\xi = 0$ and B by $\xi = 1$, then ξ^* could be $1/2$). The integer K represents some maximal window frame so that the distance is really restricted to a region around the expected or assumed transition point. In the case when $J-K, I-K < 0$ or $J+K, I+K > L$, the sum is accordingly restricted to less than $2K+1$ points.

The weights ω_i should be non-negative in all cases.

A reasonable choice for non-trivial systems is for example to use (4.76) with $p' = 1$ and $\omega_i = 1$. This approach ensures that the decorrelations arising in the initial and final basins A and B are discarded, and that only the decorrelation arising near the transition region are important. In this sense, we term this decorrelation as 'local decorrelation' since we measure how different the transition mechanisms are. As a measure of 'global decorrelation', we will consider the transition times. A numerical study based on those lines is presented below.

Numerical results

We test the different proposal functions on the model system of conformational changes of Section 4.1.4. We consider the distance (4.76) for reactive paths ($\pi \equiv \pi_{AB}$ in this case), using $p = p' = 1$ and $\omega_i = 1$, $\xi(q) = |q_1 - q_2|$, $\xi^* = r_0 + w$. We use the parameters $L = 500\,\Delta t$, $\beta = 1$, $N = 16$ particles of masses 1, $l_0 = 1.3$, $\sigma = 1$, $\epsilon = 1$, $w = 0.5$, $\Delta t = 0.0025$, with the sets $A = \{\xi(q) \leq r_0 + 0.6w\}, B = \{\xi(q) \geq r_B = r_0 + 1.4w\}$ and averaging over a total of $n = 5 \times 10^4$ Monte Carlo moves. We set $K = 30$ since the typical length of the transitions is about 60 time steps with the parameters used here.

We also consider the correlation in the transition times. We denote by $\tau(x)$ the transition index of some path x. Here, those indexes τ are such that $\xi(q_{\tau\Delta t}) = \xi^*$. The correlation function for this observable is therefore, in the case of reactive paths,

$$C(n) = \frac{\displaystyle\int\int (\tau(y) - \langle\tau\rangle_{\pi_{AB}})(\tau(x) - \langle\tau\rangle_{\pi_{AB}})\, P^n(x,dy)\, d\pi_{AB}(x)}{\displaystyle\int (\tau(x) - \langle\tau\rangle_{\pi_{AB}})^2\, d\pi_{AB}(x)},$$

with $\langle\tau\rangle_{\pi_{AB}} = \int \tau(x) d\pi_{AB}(x)$ This observable is in some sense complementary to the measure of decorrelation in the transition zone defined above since it measures some global spatial decorrelation of the paths. In practice, assuming ergodicity, C is approximated as

$$C(n) = \lim_{N\to+\infty} \frac{\frac{1}{N}\sum_{k=1}^{N}\tau(x^{n+k})\tau(x^k) - \left(\frac{1}{N}\sum_{k=1}^{N}\tau(x^{n+k})\right)\left(\frac{1}{N}\sum_{k=1}^{N}\tau(x^k)\right)}{\frac{1}{N}\sum_{k=1}^{N}\tau(x^k)^2 - \left(\frac{1}{N}\sum_{k=1}^{N}\tau(x^k)\right)^2}.$$

Figures 4.10 to 4.12 present some plots of $D(n)$ and $C(n)$ for $h = 5, 10, 15$, for the usual shooting dynamics, the noise-history algorithm, and the brownian tube proposal (with $\alpha_i = 0.8$ for all i). The average acceptance rates are also presented in Table 4.7. Notice that no shifting moves [81] are used in order to compare the intrinsic efficiencies of the proposal functions. It is likely that these moves would help improving the decorrelation rate of the sampling.

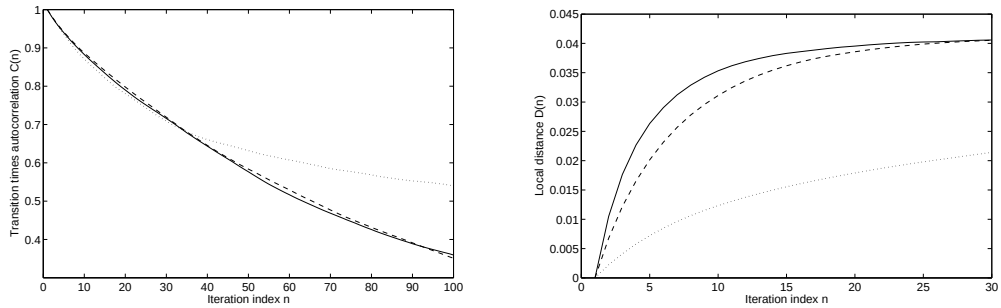


Fig. 4.10. Comparison of efficiencies for different Metropolis-Hastings proposal moves for $h = 5$. Left: Plot of the correlation of the transition times $C(n)$ (related to some global sampling efficiency). Right: Plot of $D(n)$ (local sampling efficiency) for the brownian tube proposal with $\alpha \equiv 0.8$ (solid line), usual shooting dynamics (dashed line), and noise history (dotted line).

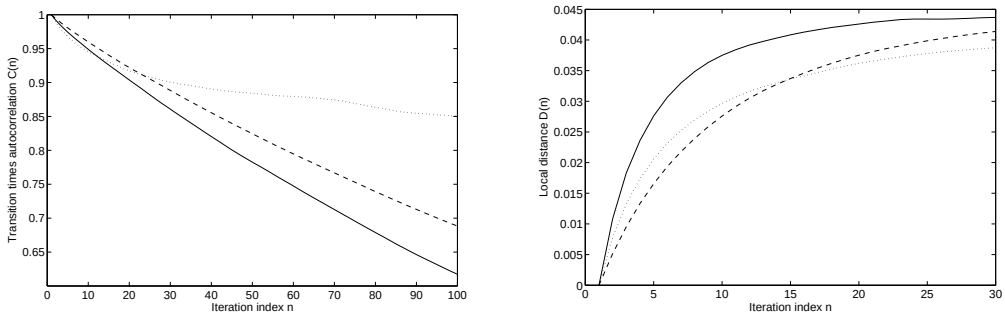


Fig. 4.11. Comparison of efficiencies for different Metropolis-Hastings proposal moves for $h = 10$.

For the shooting algorithm, many paths are rejected so that the local decorrelation (measured by $D(n)$) is rather poor, especially at short algorithmic times and for high barriers (in any cases, lower than for the brownian tube proposal). But when a path is accepted, it is already very decorrelated from the previous one, so that the global decorrelation (measured by $C(n)$) is indeed decreasing rapidly enough. For the noise-history algorithm, the picture is somewhat inverted: since the acceptance rate is very high, even for high barriers, the local decorrelation is quite

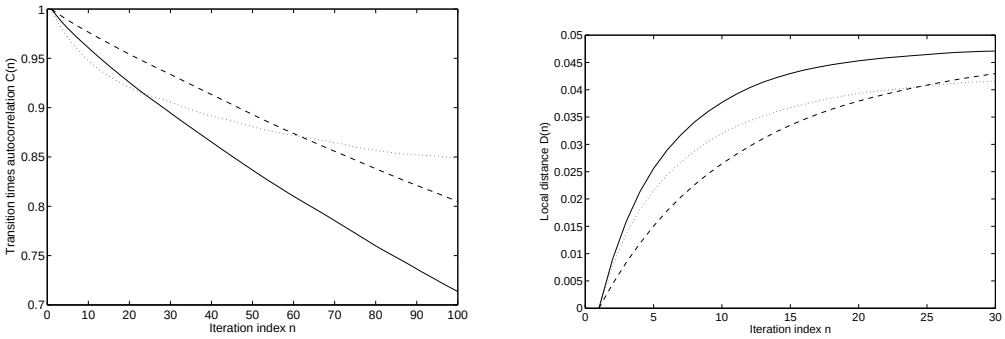


Fig. 4.12. Comparison of efficiencies for different Metropolis-Hastings proposal moves for $h = 15$.

.

Table 4.7. Acceptance rate (%) as a function of h for the three proposal functions considered.

h	5	10	15
Shooting	24.4	18.1	15.2
Noise history	96.7	85.7	81.2
Brownian tube ($\alpha_i = 0.8$)	47.2	48.1	33.0

efficient, but the global decorrelation is not since small local changes make it difficult to change the global features of the paths. The brownian tube approach tries to balance the local and global decorrelations. This is also reflected by a more balanced acceptance/rejection rate.

In conclusion, the brownian tube proposal with the above correlation function is the most efficient sampling scheme in the case considered here. The efficiency could be further increased by a more systematic tuning of the parameters of the correlation factors α_i, possibly depending on the shooting index k. In general, since the usual proposal functions are specific cases of the brownian tube proposal function, it is expected that there is always a parameter range such that this new algorithm outperforms the previous ones.

4.3.3 (Non)equilibrium sampling of the path ensemble

The previous section was dealing with equilibrium sampling of paths. However, when (free) energy barriers in path space are large, direct sampling of paths can be inefficient, since the existence of metastable path sets may considerably slow down the numerical convergence. It is therefore appealing to perform some kind of *simulated annealing* on paths. A regular simulated annealing strategy would be to first sample paths at a higher temperature, and then to cool the sample to the target temperature (see [363] for a simulated tempering version of such an idea). Reactive paths can also be otained by constraining progressively the paths to end up in B. This approach also has the nice feature that it does not ask for an initial guess to start sampling π_{AB}. Finally, a byproduct of such a switching is the ratio of partition functions in path space

$$C(L\Delta t) = \frac{Z_{AB}(L\Delta t)}{Z_A(L\Delta t)}, \tag{4.77}$$

where Z_A, Z_{AB} are such that

$$\pi_A(x) = Z_A(L\Delta t)^{-1}\mathbf{1}_{\mathrm{A}}(x_0)\rho(x_0)\prod_{i=0}^{L-1}\mathrm{p}(x_i, x_{i+1}),$$

and

$$\pi_{AB}(x) = Z_{AB}(L\Delta t)^{-1}\mathbf{1}_{\mathrm{A}}(x_0)\rho(x_0)\prod_{i=0}^{L-1}\mathrm{p}(x_i,x_{i+1})\mathbf{1}_{\mathrm{B}}(x_L)$$

are probability measures. The function C in (4.77) has to be computed at least once to obtain rate constants in practice [81]. The associated free-energy difference in path space is $\Delta F_{A\to AB}(L\Delta t) = -\ln(C(L\Delta t))$.

We start this section by recalling the extension of the classical switching dynamics for nonequilibrium dynamics in phase space to nonequilibrium switching between path ensembles [122]. This method is convenient to compute free energy differences, but the final sample of paths obtained is very degenerate. We therefore present the application to path sampling of a birth/death process introduced in [289, 292] (see also Section 4.2), which allows to keep the sample at equilibrium at all times during the switching. This equilibration may be important in some cases to compute the right free energy values [292], and allows in any cases to end up with a non-degenerate sample of paths and reduce the empirical variance. We will focus in the sequel on switching from constrained to unconstrained paths, but an extension to simulated annealing (cooling process) is straightforward.

Switching between ensembles of paths

We present in this section the approach of [122], where the switching from unconstrained to constrained path ensembles is done by enforcing progressively the constraint on the end point of the path over a time interval $[0,T]$. The constraint is usually parametrized using some order parameter. This order parameter is the same as the one used for usual computations of reaction rates in the TPS framework (and even for more advanced techniques such as Transition Interface Sampling (TIS) [355,356]). The point is that this approximate order parameter needs not to be a "good" reaction coordinate (or a complete one) since the general path sampling approach should help to get rid of some problems arising from a wrong choice of order parameter (see *e.g* [354] for a recent study on this topic).

Assuming an order parameter is given, we can consider a switching schedule $\lambda = (\lambda^0,\ldots,\lambda^n)$ such that $\lambda^0 = 0$ and $\lambda^n = 1$ and a family of functions h_λ such that

$$h_0 = \mathbf{1}, \quad h_1 = \mathbf{1}_B.$$

We also introduce the family of probability measures associated with the functions h_λ:

$$\pi_\lambda(x) = Z_{L,\lambda}^{-1}\mathbf{1}_{\mathrm{A}}(x_0)\rho(x_0)\prod_{i=0}^{L-1}\mathrm{p}(x_i,x_{i+1})h_\lambda(x_L). \tag{4.78}$$

We omit in the sequel the explicit dependence of the partition functions Z on L and Δt. An energy $\mathcal{E}_\lambda(x)$ can then formally be associated to a path x as

$$\pi_\lambda(x) = Z_{L,\lambda}^{-1}\mathrm{e}^{-\mathcal{E}_\lambda(x)}.$$

The aim is to sample from $\pi_1 \equiv \pi_{AB}$, which is usually a difficult task, and sometimes not directly feasible. It may be easier to use a sample of $\pi_0 = \pi_A$ (which is much easier to obtain), and to transform it through some switching dynamics into a (weighted) sample of π_1. Starting from a path $x^{k,0}$, the weight factor for a resulting path $x^{k,n}$ is of the form $\mathrm{e}^{-W^{k,n}}$ where $W^{k,n}$ is the work exerted on an unconstrained path to constrain it to end in B. We now precise the way the work is computed.

Consider an unconstrained initial path $x^0 = (x_0^0,\ldots,x_L^0)$ sampled according to π_0, and a discrete schedule $(\lambda^0,\ldots,\lambda^n)$. The dynamics in path space is as follows:

NONEQUILIBRIUM SWITCHING ON PATHS

Algorithm 4.3 (See Ref. [122]). Consider an initial configuration x^0 generated from π_0. Starting from W^0 and $m = 0$,

(1) Replace λ^m by λ^{m+1};

(2) Update the work as $W^{m+1} = W^m + \mathcal{E}_{\lambda^{m+1}}(x^m) - \mathcal{E}_{\lambda^m}(x^m)$;

(3) Do a Monte Carlo path sampling move using a Metropolis-Hastings scheme with the measure $\pi^{\lambda^{m+1}}$ (using for example the usual shooting moves with a Langevin dynamics, or the Monte Carlo move designed for path switching presented below), so that the current path x^m is transformed into the new path x^{m+1}.

This procedure is repeated for independent initial conditions $x^{k,0}$, so that a sample of M end paths $(x^{1,n}, \ldots, x^{M,n})$ with weights $(\mathrm{e}^{-W^{1,n}}, \ldots, \mathrm{e}^{-W^{M,n}})$ is obtained. Besides, an estimation of the rate constant is given by the exponential average

$$C_M(L\Delta t) = -\ln\left(\frac{1}{M}\sum_{k=1}^{M}\mathrm{e}^{-W^{k,n}}\right),$$

and it can be shown that $C_M \to C$ when $M \to +\infty$.

Since the realizations of the switching procedure are independent provided the initial conditions are independent, the random variables $\{\mathrm{e}^{-W^{k,n}}\}_k$ are i.i.d. A confidence interval can be obtained for C_M as

$$C^-_{M,\sigma_\mathrm{c}} \leq C_M \leq C^+_{M,\sigma_\mathrm{c}},$$

with

$$C^\pm_{M,\sigma_\mathrm{c}} = -\ln\left(\frac{1}{M}\sum_{k=1}^{M}\mathrm{e}^{-W^{k,n}} \pm \sigma_\mathrm{c}\sqrt{\frac{V_M}{M}}\right),$$

where the empirical variance is

$$V_M = \frac{1}{M-1}\sum_{k=1}^{M}\left(\mathrm{e}^{-W^{k,n}} - \frac{1}{M}\sum_{l=1}^{M}\mathrm{e}^{-W^{l,n}}\right)^2.$$

A confidence interval on the free energy difference is then

$$-\ln C^-_{M,\sigma_\mathrm{c}} \leq \Delta F_{A\to AB} \leq -\ln C^+_{M,\sigma_\mathrm{c}}.$$

For example, the 95 % confidence interval corresponds to $\sigma_\mathrm{c} = 1.96$.

Of course, as usual for nonequilibrium switchings, it may the case that the variance of the work distribution is large, so that only very few paths are relevant (and the confidence interval for the rate constant is large), so that an equilibration in the vein of Section 4.2 may be interesting.

Enhancing the number of relevant paths

We present here an extension of the IPS equilibration to the case of path sampling. Then, each path has weight 1 in the end, and the final sample $(x^{1,n}, \ldots, x^{M,n})$ is distributed according to $\pi_1 \equiv \pi_{AB}$ (provided the switching is slow enough and the number of replicas is large enough; therefore, $Mn\Delta t$ should be large enough). More precisely, we consider the

IPS EQUILIBRATION OF THE NONEQUILIBRIUM PATH SWITCHING

Algorithm 4.4. Consider an initial distribution $(x^{1,0}, \ldots, x^{M,0})$ generated from π_0. Generate independent times $\tau^{k,b}, \tau^{k,d}$ from an exponential law of mean 1. Consider two additional variables $\Sigma^{k,b}, \Sigma^{k,d}$ per replica, initialized at 0.

(1) Replace λ^m by λ^{m+1};

(2) Update the works as $W^{k,m+1} = W^{k,m} + \Delta\mathcal{E}^{k,m} = W^{k,m} + \mathcal{E}_{\lambda^{m+1}}(x^{k,m}) - \mathcal{E}_{\lambda^m}(x^{k,m})$, and compute the mean work update $\overline{\Delta\mathcal{E}}^m = M^{-1}\sum_{1\le k\le M}\Delta\mathcal{E}^{k,m}$;

(3) (Diffusion step) Do a Monte Carlo path sampling move using a Metropolis-Hastings scheme with the measure $\pi_{\lambda^{m+1}}$, so that $x^{k,m}$ is transformed into $x^{k,m+1}$.

(4) (Birth/death process) Update the variables $\Sigma^{k,b}$ and $\Sigma^{k,d}$ as

$$\Sigma^{k,b} = \Sigma^{k,b} + \beta(\overline{\Delta\mathcal{E}}^m - \Delta\mathcal{E}^{k,m})^-,$$

and

$$\Sigma^{k,d} = \Sigma^{k,d} + \beta(\overline{\Delta\mathcal{E}}^m - \Delta\mathcal{E}^{k,m})^+.$$

(Death) If $\Sigma^{k,d} \ge \tau^{k,d}$, select an index $m \in \{1, \ldots, M\}$ at random, and replace the k-th path by the m-th path. Generate a new time $\tau^{k,d}$ from an exponential law of mean 1, and set $\Sigma^{k,d} = 0$;

(Birth) If $\Sigma^{k,b} \ge \tau^{k,b}$, select an index $m \in \{1, \ldots, M\}$ at random, and replace the m-th path by the k-th path. Generate a new time $\tau^{k,b}$ from an exponential law of mean 1, and set $\Sigma^{k,b} = 0$;

In this case, an estimation of the rate constant is given by the simple average

$$C_M(L\Delta t) = \frac{1}{M}\sum_{k=1}^{M} W^{k,n},$$

and it can be shown that $C_M \to C$ when $M \to +\infty$. A confidence interval for the free energy difference can be obtained as in Section 4.3.3 as

$$C^{\mathrm{IPS},-}_{M,\sigma_c} \le C^{\mathrm{IPS}}_M \le C^{\mathrm{IPS},+}_{M,\sigma_c},$$

with

$$C^{\mathrm{IPS},\pm}_{M,\sigma_c} = \frac{1}{M}\sum_{k=1}^{M} W^{k,n} \pm \sigma_c\sqrt{\frac{V^{\mathrm{IPS}}_M}{M}},$$

the empirical variance being

$$V^{\mathrm{IPS}}_M = \frac{1}{M-1}\sum_{k=1}^{M}\left(W^{k,n} - \frac{1}{M}\sum_{l=1}^{M} W^{l,n}\right)^2.$$

Specific Monte-Carlo moves for switching from unconstrained to constrained path ensembles

When an interpolating function h_λ appearing in (4.78) (or, equivalently, some order parameter ξ) is known, it is possible to increase the likeliness of the end point of the trajectory by performing a move on the last configuration in the direction opposite to $\nabla h_\lambda(q)$ while keeping the random vectors used for the transitions. These moves should of course be employed with other MC moves,

especially MC moves relying on some trajectory generation, in order to relax the shift toward higher values of h_λ or ξ.

More precisely, using for example an overdamped Langevin dynamics to update the end configuration, the associated Metropolis-Hastings Monte-Carlo elementary step is, starting from a path x for a parameter λ (in the Langevin dynamics setting):

SPECIFIC MONTE-CARLO SWITCHING MOVE

Algorithm 4.5. Starting from a path $x = (x_0, \ldots, x_L)$,

(1) Compute the sequence of $2dN$-dimensional random vectors $(\bar{U}_i)_{0 \le i \le L-1}$ associated with the backward (time-reversed) integration from x_L to x_0;

(2) Compute a final configuration as $q_L^y = q_L^x + \delta_\lambda \nabla \xi(q_L^x) + (2\delta_\lambda/\beta)^{1/2} G$ where G is a d-dimensional random gaussian vector;

(3) Integrate the path backward (time-reversed) starting from y_L, using the noises $(\bar{U}_i)_{0 \le i \le L-1}$ to obtain a path $y = (y_0, \ldots, y_L)$. The probabilty $\mathcal{P}(x, y)$ to obtain y starting from x is therefore the probability to obtain y_L from x_L, so that

$$\mathcal{P}(x, y) = \mathrm{p}_{\mathrm{switch}}(x_L, y_L) = \left(\frac{\beta}{4\pi\delta_\lambda^2}\right)^{d/2} \exp\left(-\frac{\beta}{4\delta_\lambda}|q_L^y - q_L^x - \delta_\lambda \nabla \xi(q_L^x)|^2\right).$$

(4) Accept the new path y with probability

$$r(x, y) = \min\left(1, \frac{\pi(y)\mathcal{P}(y, x)}{\pi(x)\mathcal{P}(x, y)}\right) = \min\left(1, \frac{\mathbf{1}_{\mathrm{A}}(y_0)\rho(y_0)}{\mathbf{1}_{\mathrm{A}}(x_0)\rho(x_0)} \frac{\mathrm{p}_{\mathrm{switch}}(y_L, x_L)}{\mathrm{p}_{\mathrm{switch}}(x_L, y_L)}\right).$$

The magnitude δ_λ can be made to depend a priori on λ. It is then adjusted in pratice on the fly by first computing the values of the gradient for the endpoint of each replica, in order to ensure that the displacement is small enough.

Numerical results

We compute here free energy differences associated with constraining paths for the WCA model system introduced in Section 4.1.4. This is done either with plain nonequilibrium switching, or with the IPS equilibration. Let us notice that the energy is fixed in [122] while we rather have to fix the temperature in the stochastic setting, so that a straightforward comparison of the results is not possible. We set $\beta = 1$ in the sequel. The other parameters are the same as in [122]: $N = 9$ particles, $h = 6$, $\sigma = 1$, $\epsilon = 1$, the particle density $\rho = 0.6\sigma^{-2}$, $w = 0.25$, and the sets $A = \{\xi(q) \le \xi_A = 1.3\sigma\}, B = \{\xi(q) \ge \xi_B = 1.45\sigma\}$. The trajectory length is $L = 320\,\Delta t$ and $\Delta t = 0.0025$, so that $L\Delta t = 0.8(m\sigma^2/\epsilon)^{1/2}$.

We perform a total of n MC moves (using the brownian tube proposal function (with $\alpha_i = \alpha = 0.8$ for all $0 \le i \le L-1$). The function h_λ is the one given in [122]:

$$h_\lambda(q) = \mathrm{e}^{-\lambda K(1-\mathbf{1}_B(q))(\xi_B - \xi(q))}$$

with $K = 100$. The switching schedule is $\lambda^i = (i/n)^2$.

A typical free energy difference profile is presented in Figure 4.13 for $M = 2000$ and $n = 10000$, as well as the associated weights for the plain nonequilibrium switching. These weights are the Jarzynski weights renormalized by the total weight (in order to define a probability distribution):

$$w_k = \frac{\mathrm{e}^{-W^{k,n}}}{\sum_{l=1}^{M} \mathrm{e}^{-W^{l,n}}}. \tag{4.79}$$

Table 4.8. Free energy differences $\Delta F_{A\to AB}$ computed for different switching lengths n, using a sample of $M = 2000$ paths. The results are presented under the form "C_M $(C^-_{M,\sigma_c} - C^+_{M,\sigma_c})$" with $\sigma_c = 1.96$ (the value corresponding to a 95 % confidence interval).

M	n	Backward		Forward		IPS (forward)	
2000	2000	4.83	(4.61-5.02)	5.43	(5.28-5.61)	4.82	(4.78-5.85)
2000	5000	5.34	(5.04-5.58)	5.41	(5.32-5.50)	5.19	(5.16-5.23)
2000	10000	5.45	(5.32-5.58)	5.40	(5.34-5.46)	5.40	(5.36-5.43)
2000	15000	5.42	(5.35-5.49)	5.40	(5.35-5.45)	5.45	(5.42-5.48)

Notice that the sample is very degenerate since very many paths have negligible weights, and the relevant paths are exponentially rare. Recall also that the paths all have weight 1 with the IPS algorithm.

Some free energy differences are presented in Table 4.8 for different values of n (keeping M fixed). The switchings are slow enough when the confidence intervals for free energy differences computed by constraining paths ('forward' switching) overlap with confidence intervals for free energy differences obtained by starting from a sample of constrained paths and removing progressively the constraint ('backward' switching). This is the case here for $n = 5000, 10000, 15000$ (but not when $n = 2000$). The results show that IPS agrees with the usual Jarzynski switching, the confidence interval on the results being however lower.

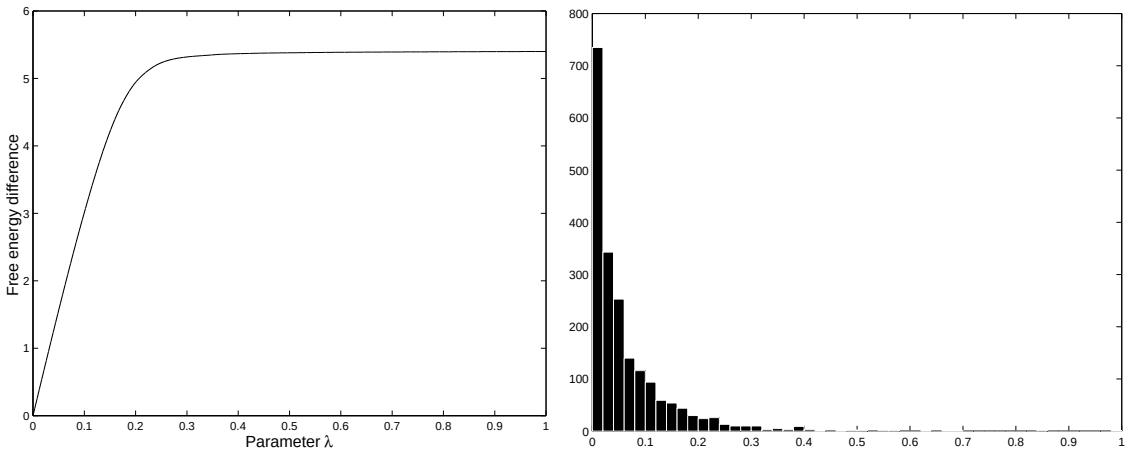


Fig. 4.13. Left: Free energy profile for a forward switching, computed for $M = 2000$ and $n = 10^4$, using a plain nonequilibrium switching. Right: Histogram of the weights w_k of the final sample as given by (4.79).

We also present in Figure 4.14 a final sample computed using a quite fast switching ($n = 1000$) with a small sample of paths ($M = 100$). Notice that all the 100 paths generated with the IPS switching are reactive, in contrast with the paths generated by a straightforward switching in the Jarzynski way. Besides, as a consequence of the degeneracy of paths, only 8 paths in 100 have a significant weight (larger than 0.05 when normalized by the total weight, see (4.79)). This simple example shows why it is difficult to compute averages over the final sample of paths when performing plain nonequilibrium switching, and why it may be interesting to resort to some selection process to prevent such a degeneracy.

In agreement with a previous study [292], the results show that the IPS algorithm allows to reduce the variance on the estimates and to end up the simulation with a well-distributed and non-degenerate sample, provided the switching is slow enough.

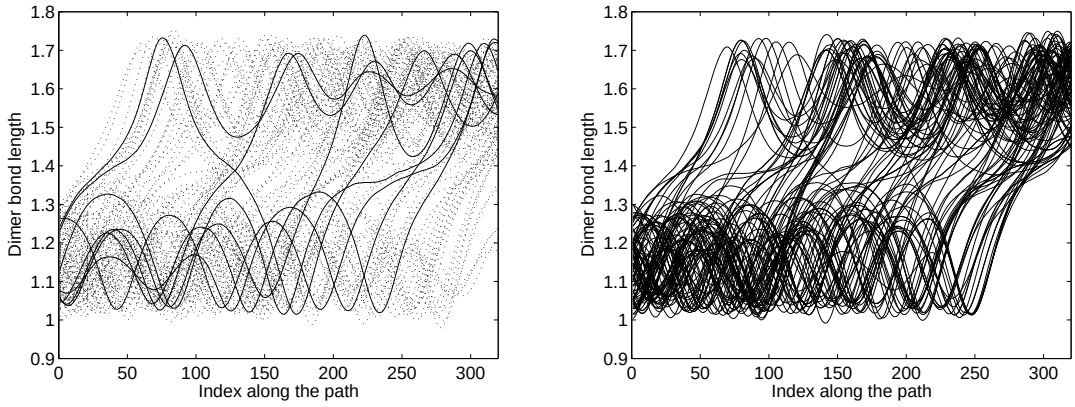


Fig. 4.14. Comparison, for a nonequilibrium switching of paths for $M = 100$ systems in $n = 1000$ steps without (Left) or with IPS (Right). Only the paths having a weight greater than 0.05 are plotted in solid lines when plain nonequilibrium switching is used (the other paths are plotted in dotted lines).

4.4 Adaptive computation of free energy differences

Methods relying on nonequilibrium dynamics follow the pioneering work of Jarzynski [187], or use some adaptive dynamics such as the Wang-Landau approach [368], the adaptive biasing force (ABF) [75, 76, 157], or the nonequilibrium metadynamics [46]. These approaches use the whole history of the exploration process to bias the current dynamics in order to force the escape from metastable sets. This is done by simultaneously estimating the free energy from an evolving ensemble of configurations of the dynamics, and using this estimate to bias the dynamics, so that the effective free energy surface explored is flattened. In the long time limit, the bias exactly gives the actual free energy profile. Adaptive methods could therefore be seen as umbrella sampling with an evolving potential. This was already noticed in a previous study presenting an adaptive dynamics as a 'self-healing umbrella sampling' [227].

To present the adaptive methods mentioned above in a general and unifying framework, it is convenient, as is done in [46], to consider ensemble of realizations (see Eq. (4.83)). The system is then described by the distribution of the configurations of this ensemble in the limit of an infinite number of replicas simulated in parallel. The key point is to reformulate the computation of the bias of adaptive dynamics, using conditional distributions (that is, distribution of the configurations for a given value of the reaction coordinate) of the latter sample. This was already proposed in [101] in the equilibrium case, and is somewhat implicit in [46]. This concept clarifies the presentation of adaptive methods, allows mathematical proofs of convergence [207] or at least, existence of a stationary state of the dynamics (still in the case of an infinite number of replicas), and suggests natural numerical strategies: the discretization may be done through a parallel implementation of several replicas of the system, which all contribute to construct the free energy profile. Such a parallel implementation was already proposed in [275] in the case of metadynamics. We show here how an additional selection process on the replicas can enhance the sampling of the reaction coordinates in comparison with a straightforward parallel implementation.

This section is organized as follows. In Section 4.4.1, we describe the general formalism for adaptive dynamics, using conditional probabilities, and show how to update the biasing potential in order to compute the free energy profile in the longtime limit, using a fixed-point strategy. Some applications of this formalism are then presented, which allow to recover the usual adaptive dynamics such as the nonequilibrium metadynamics, the Wang-Landau scheme or the ABF method. We then discuss possible parallel implementation strategies. In particular, it is shown how a selection process can enhance the straightforward parallel implementation. This is finally illustrated

by numerical results for a toy model of conformational changes. In Section 4.4.2, we then present a rigorous proof of convergence for a specific adaptive dynamics in the ABF spirit, using entropy estimates. The proof uses on a decomposition of the entropy into a macroscopic entropy (related to the distribution of the values of the reaction coordinate) and a microscopic entropy (depending on the distribution of the conditioned measures, for a fixed value of the reaction coordinate), and relies on the assumption that the conditioned measure satisfy a logarithmic Sobolev inequality, with a constant independent of the value of the reaction coordinate.

4.4.1 A general framework for adaptive methods

For a system described by a potential $V(q)$, the Boltzmann measure in the canonical ensemble is $Z^{-1}\exp\left(-\beta V(q)\right)\,dq$ (where Z is a normalization constant, the so-called partition function). We consider in this section a reaction coordinate ξ, taking values in the one dimensional torus, or in the interval $[0,1]$. In the latter case, reflecting boundary conditions for the dynamics on the two extremal values $\xi(q)=0$, $\xi(q)=1$ are used. Recall that the free energy (or potential of mean force (PMF)) to be computed is defined up to an additive constant by the normalization of a Boltzmann average of the configurations restricted to a given value of the reaction coordinate (see Section 4.1.2 for more details):

$$F(z) = -\beta^{-1}\ln\int_{\mathcal{M}}\exp(-\beta V(q))\,\delta_{\xi(q)-z}. \tag{4.80}$$

and the associated mean force is

$$F'(z) = \frac{\displaystyle\int_{\mathcal{M}} f^V(q)\,\exp(-\beta V(q))\,\delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}}\exp(-\beta V(q))\,\delta_{\xi(q)-z}}, \tag{4.81}$$

with the local force given by

$$f^V = \frac{\nabla V\cdot\nabla\xi}{|\nabla\xi|^2} - \beta^{-1}\mathrm{div}\left(\frac{\nabla\xi}{|\nabla\xi|^2}\right). \tag{4.82}$$

Here and in the sequel, we denote by F' the derivative of F with respect to z.

Adaptive dynamics are defined through the dynamics used, which dictates the distribution of the configurations at equilibrium, a biasing potential, and the way this potential is updated (see below for a heuristic derivation in the equilibrium case motivating the general setting).

Trajectories $t\mapsto Q_t$ are computed according to some dynamics which are ergodic with respect to the Boltzmann measure when the potential is time-independent. For instance, the Langevin dynamics or the overdamped Langevin dynamics may be used. We will denote by $\psi_t(q)$ the distribution (or density) of configurations at time t. This distribution will be used to update the biasing potential F_{bias}.

From a practical point of view, when M replicas $(Q_t^{i,M})_{i=1,\ldots,M}$ of the system are simulated in parallel, the density of states $\psi_t(q)$ is approximated by the instantaneous distribution of the replicas

$$\psi_t(q) = \lim_{M\to+\infty}\frac{1}{M}\sum_{i=1}^{M}\delta_{Q_t^{i,M}-q}. \tag{4.83}$$

In some cases, the density of states can also be approximated using the distribution of configurations along the trajectory, relying on some ergodic assumption.

The definition of adaptive methods requires the definition of two important quantities obtained from the distribution $\psi_t(q)$. The first one is the distribution ψ_t^ξ of the reaction coordinate values,

which is, from a mathematical perspective, the marginal law of ψ_t with respect to ξ:

$$\psi_t^\xi(z) = \int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}. \tag{4.84}$$

This quantity will be useful to propose a biasing potential (see Eqs. (4.91)-(4.93)). Another important quantity is the conditional average of some function h for some fixed value of the reaction coordinate:

$$\langle h \rangle_{t,z} = \frac{\displaystyle\int_{\mathcal{M}} h(q)\psi_t(q)\, \delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}}. \tag{4.85}$$

Such averages are used to propose biasing forces (see Eqs. (4.92)-(4.94)).

The biasing potential

In adaptive dynamics, the interaction potential is time-dependent:

$$\mathcal{V}_t(q) = V(q) - F_{\text{bias}}(t, \xi(q)). \tag{4.86}$$

The biasing potential F_{bias}, whose precise form varies according to the method under study, depends only on q through the reaction coordinate value $\xi(q)$ and is updated using the history of the configurations. It is expected that this biasing potential converges (up to an additive constant) toward the free energy F given by (4.80) in the long-time limit, so that the equilibrium distribution of the reaction coordinate is the uniform distribution.

The key idea common to all adaptive methods is to resort to a fixed point strategy, in order for the observed free energy to converge to a constant or the mean force to vanish, and the dynamics to reach equilibrium (see the updates (4.88) or (4.90) in the equilibrium case and (4.93) or (4.94) in the nonequilibrium case).

Updating the biasing potential - The equilibrium case

To derive a possible form for the biasing potential, let us first assume that the system is instantaneously at equilibrium with respect to the biased potential $\mathcal{V}_t$, *i.e.* Q_t has density $\psi_t^{\text{eq}}(q) = Z_t^{-1} \exp(-\beta \mathcal{V}_t(q))$. In this case, resorting to (4.80), the *observed free energy* (see (4.91) for a general definition) is

$$-\beta^{-1} \ln \int_{\mathcal{M}} \psi_t^{\text{eq}}(q)\, \delta_{\xi(q)-z} = F(z) - F_{\text{bias}}(t, z) + \beta^{-1} \ln Z_t. \tag{4.87}$$

Thus, for a characteristic time τ to be chosen, an update of F_{bias} of the form

$$\partial_t F_{\text{bias}}(t, z) = -\frac{\beta^{-1}}{\tau} \ln \int_{\mathcal{M}} \psi_t^{\text{eq}}(q)\, \delta_{\xi(q)-z} \tag{4.88}$$

is such that $F'_{\text{bias}}(t) \to F'$ when $t \to +\infty$ exponentially fast with rate $1/\tau$. Notice that we stated the convergence in terms of the mean force, because, in view of the constant term $\beta^{-1} \ln Z_t$ in Eq. (4.87), the potential of mean force only converges up to a constant to the true potential of mean force.

Similar considerations hold for the mean force: replacing the potential V with $\mathcal{V}_t$ given by (4.86), and resorting to (4.81)-(4.82), the observed mean force (see (4.92) for a general definition) is

$$\frac{\displaystyle\int_{\mathcal{M}} f^{\mathcal{V}_t}(q)\, \psi_t^{\text{eq}}(q)\, \delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t^{\text{eq}}(q)\, \delta_{\xi(q)-z}} = F'(z) - F'_{\text{bias}}(t, z), \tag{4.89}$$

since $f^{\mathcal{V}_t}(q) = f^V(q) - F'_{\rm bias}(t, \xi(q))$. An update of $F'_{\rm bias}(t)$ of the form

$$\partial_t F'_{\rm bias}(t,z) = \frac{1}{\tau} \frac{\int_{\mathcal{M}} f^{\mathcal{V}_t}(q)\, \psi_t^{\rm eq}(q)\, \delta_{\xi(q)-z}(dq)}{\int_{\mathcal{M}} \psi_t^{\rm eq}(q)\, \delta_{\xi(q)-z}(dq)} \tag{4.90}$$

is therefore such that $F'_{\rm bias}(t) \to F'$ when $t \to +\infty$ exponentially fast with rate $1/\tau$.

Updating the biasing potential - The nonequililibrium case

Now, in general, the system is not at equilibrium for the potential $\mathcal{V}_t$: $\psi_t \neq \psi_t^{\rm eq}$. We use the above procedure as a guideline to update the biasing potential $F_{\rm bias}(t,z)$. To derive equations for the biasing potential, let us first define two quantities. The first one is the *observed free energy* or the *observed potential of mean force*, defined as

$$F_{\rm pot,obs}(t,z) = -\beta^{-1} \ln \int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}. \tag{4.91}$$

This quantity can be interpreted as the free energy associated with the ensemble of configurations with density of states $\psi_t(q)$ (see Eq. (4.80)). The observed free energy $F_{\rm pot,obs}(t,z)$ is high when the number of visited states with reaction coordinate value z is small. In the long-time limit, the distribution of the reaction coordinate is expected to be uniform, so that the observed free energy is constant.

In the same way, the *observed mean force* is defined as the conditional average of the time-dependent biasing force for a given value of the reaction coordinate:

$$F'_{\rm force,obs}(t,z) = \frac{\displaystyle\int_{\mathcal{M}} f^{\mathcal{V}_t}(q)\, \psi_t(q)\, \delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}} = \frac{\displaystyle\int_{\mathcal{M}} f^{V}(q)\, \psi_t(q)\, \delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}} - F'_{\rm bias}(t,z). \tag{4.92}$$

This quantity can be interpreted as the mean force associated with $\psi_t(q)$ (see Eqs. (4.81)-(4.82)), minus the biasing force at time t. It is expected to vanish in the long-time limit, so that the corresponding observed free energy is also constant.

The fixed point strategy relies on two different ways of updating the bias (the *updating functions* g_t and G_t are increasing functions such that $G_t(0) = 0$):

(i) The first strategy, which may be called Adaptive Biasing Potential (ABP) method, is the generalization of (4.88) to the nonequilibrium case. The bias is updated in its potential form, preferably increased (resp. decreased) for reaction coordinate values such that the observed free energy is high (resp. low):

$$\text{(ABP)} \qquad \partial_t F_{\rm bias}(t,z) = g_t(F_{\rm pot,obs}(t,z)); \tag{4.93}$$

(ii) The second strategy, the usual ABF method, generalizes (4.90). The bias is updated through the mean force: the biasing force is increased (resp. decreased) for reaction coordinate values such that the observed mean force is positive (resp. negative):

$$\text{(ABF)} \qquad \partial_t F'_{\rm bias}(t,z) = G_t(F'_{\rm force,obs}(t,z)). \tag{4.94}$$

Let us emphasize at this point that the ABF and the ABP methods yield very different biasing dynamics, since the derivative of (4.91) with respect to z is different from (4.92) (This is not the case when the system is at equilibrium: the derivative of (4.88) with respect to z is equal to (4.90)). This difference becomes critical for multi-dimensional reaction coordinates, where the biasing force no longer derives from a potential in general.

Consistency of the method

Let us show that within this formalism, any stationary state of the ABP or ABF methods gives the true mean force F' to be computed (and therefore the true PMF up to an additive constant). For a stationary state where the biasing potential has converged to $F_{\text{bias}}(\infty)$, the ergodicity property of the dynamics ensures that samples of configurations of the system are distributed according to $\psi_\infty = Z_\infty^{-1} \exp[-\beta(V - F_{\text{bias}}(\infty, \xi))]$.

The observed free energy or mean force given by Eqs. (4.91) and (4.92) then both verify $F'_{\text{pot,obs}}(\infty, z) = F'_{\text{force,obs}}(\infty, z) = F'(z) - F'_{\text{bias}}(\infty, z)$. The updating equations Eqs. (4.93) and (4.94) yield respectively

$$g_\infty(F(z) - F_{\text{bias}}(\infty, z)) = 0, \tag{4.95}$$

$$G_\infty(F'(z) - F'_{\text{bias}}(\infty, z)) = 0, \tag{4.96}$$

so that (taking the derivative with respect to z in (4.95)) $F'_{\text{bias}}(\infty) = F'$ in both cases thanks to the strict monotonicity of the updating functions. Let us also notice that, at convergence, the values of the reaction coordinate are distributed uniformly: $\int_{\mathcal{M}} \psi_\infty(q)\, \delta_{\xi(q)-z} = 1$.

However, let us emphasize that we did not give any convergence result at this point. We merely showed that, *if the dynamics converges*, then the limiting state is the correct one. To prove convergence starting from an arbitrary initial distribution is a difficult task, and can only be done for certain dynamics (see the corresponding results in Section 4.4.2).

Application to usual adaptive dynamics and convergence results

We present in this section some applications of the above formalism, and show that the usual adaptive methods can indeed be recovered. This is summarized in Table 4.9, which gives a classification of adaptive methods.

Table 4.9. Classification of adaptive methods.

	Adaptive Biasing Force ($\partial_t F'_{\text{bias}}$)	Adaptive Biasing Potential ($\partial_t F_{\text{bias}}$)
Dimension n (V)	ABF [75, 76, 157]	ABP [368]
Dimension $n+1$ (V^μ)	m-ABF	m-ABP [46, 275]

Metadynamics

Adaptive strategies can be used with metadynamics. The configuration space is extended by considering an additional variable z representing the reaction coordinate, and the dynamics is denoted $t \mapsto (Q_t, Z_t)$. The associated extended potential incorporates a coupling between this new variable and the reaction coordinate ξ:

$$V^\mu(q, z) = V(q) + \frac{\mu}{2}(z - \xi(q))^2,$$

for some (large) $\mu > 0$. In this case, the new reaction coordinate considered is $\xi_{\text{meta}}(q, z) = z$ and the free energy is thus given by:

$$F^\mu(z) = -\beta^{-1} \ln \int_{\mathcal{M}} \exp(-\beta V^\mu(q, z))\, dq.$$

It is easy to check that, up to an additive constant, $F^\mu \to F$ as $\mu \to +\infty$, with F given by (4.80). The adaptive strategies presented above applied to this extended dynamics allow to recover the free

energy F^μ. The corresponding dynamics may be called meta-Adaptive Biasing Potential (m-ABP) and meta-Adaptive Biasing Force (m-ABF) methods.

Strategies relying on biasing potentials are reminiscent of flooding strategies [140] such as the nonequilibrium metadynamics [46]. The latter is an example of an m-ABP method, where the biasing potential is applied to the extended variable. The updating function does not depend on time and is given by $g_t(x) = -\gamma \exp(-\beta x)$ for some constant $\gamma > 0$. The ensemble of configuration used in the adaptive update is obtained from M replicas $(Q_t^{i,M}, Z_t^{i,M})$ running in parallel, so that

$$\psi_t(q,z) \simeq \frac{1}{M} \sum_{i=1}^{M} \delta_{(Q_t^{i,M}, Z_t^{i,M})-(q,z)}.$$

The resulting biasing potential at time t penalizes the values of the reaction coordinate already visited according to (see (4.93)):

$$F_{\text{bias}}(t,z) \simeq F_{\text{bias}}^M(t,z) = -\frac{\gamma}{M} \sum_{i=1}^{M} \int_0^t \delta_{Z_s^{i,M}-z}\, ds. \tag{4.97}$$

In the case of an overdamped Langevin dynamics with $M = 1$ for example, the resulting equations of motion are therefore:

$$\begin{cases} dQ_t = -\nabla V(Q_t)\, dt + \mu(Z_t - \xi(Q_t))\nabla\xi(Q_t)\, dt + \sqrt{2\beta^{-1}}\, dW_t^Q, \\ dZ_t = -\mu(Z_t - \xi(Q_t))\, dt + \sqrt{2\beta^{-1}}\, dW_t^Z - \gamma\nabla_z \left(\int_0^t \delta_{Z_s - z}\, ds \right) dt, \end{cases}$$

where the processes W_t^Q, W_t^Z are independent standard Brownian motions. When in the last equation and in (4.97) the Dirac masses $\delta_{Z_t - z}$ are discretized using Gaussian functions, the nonequilibrium metadynamics described in [46, 275] are recovered. We also refer to [46] for an error analysis.

The Wang-Landau algorithm

Another famous instance of an ABP dynamics, usually defined in discrete spaces, is the Wang-Landau algorithm [368]. The biasing potential is constructed in a similar fashion to (4.97), without extending the configuration space and with only one replica. The updating function is modified during time as $g_t(x) = -\gamma(t) \exp(-\beta x)$, so that

$$F_{\text{bias}}(t,z) = -\int_0^t \gamma(s)\, \delta_{\xi(Q_s)-z}\, ds. \tag{4.98}$$

If $\gamma(t) \to 0$ slowly enough, it is possible to prove the convergence of the dynamics, the rate of convergence of $\gamma(t)$ being controlled by the nonuniformity of the histogram of the time distribution of the reaction coordinate (see [14] for more precisions on the convergence results).

The ABF method

The usual ABF bias [157] is given by averaging the local force f^V over the configurations visited by the system. It is recovered in the formalism we propose by considering one replica of the system, and an updating function of the form $G_t(x) = \gamma x$ in the limit $\gamma \to \infty$. This gives indeed:

$$F'_{\text{bias}}(t,z) = \frac{\displaystyle\int_{\mathcal{M}} f^V(q)\, \psi_t(q)\, \delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t(q)\, \delta_{\xi(q)-z}}. \tag{4.99}$$

Since there is only one replica, the density $\psi_t(s)$ is approximated by a trajectorial distribution, for example

$$\psi_t(q) \simeq \frac{1}{T}\int_{t-T}^{t} \delta_{Q_s-q}\, ds \tag{4.100}$$

for some averaging time $T > 0$ and $t > T$.

For a rigorous convergence result of the ABF algorithm with the update (4.99) in the case of an overdamped Langevin dynamics with an infinite number of replicas, see [207] and Section 4.4.2.

Practical implementation strategies

Relying on the definition (4.83) of the distribution of configurations, adaptive dynamics can be easily parallelized by using a large number M of replicas that interact through the biasing potential or the biasing force. We first show in this section how to discretize the dynamics and the biasing potential, and then, how this implementation can be improved using some selection process.

Discretization of the biasing potential

In order to compute in practice the conditional or marginal distributions needed to update the biasing potential, there are basically two approaches, relying either on ergordic limits or on ensemble averages. Both approaches may be combined in practice in order to obtain smooth profiles. For example, when only a limited number of replicas M is used, the density $\psi_t(q)$ given by (4.83) is not regular, and some local averaging is necessary (see *e.g.* Eq. (4.101)).

We detail the implementation in the ABF case for example. The ABP case can be treated in a similar way (see also [275]). The instantaneous conditional average of some function h is typically approximated by

$$\langle h\rangle_{t,z} \simeq \langle h\rangle_{t,z}^M = \frac{\displaystyle\sum_{i=1}^M h(Q_t^{i,M})\delta_z^\epsilon(\xi(Q_t^{i,M}))}{\displaystyle\sum_{i=1}^M \delta_z^\epsilon(\xi(Q_t^{i,M}))},$$

where $Q_t^{i,M}$ is the i-th replica at time t and δ_z^ϵ is some approximation of the Dirac distribution δ_z, such as a gaussian function with standard deviation ϵ or the indicator function of an interval of size ϵ. In order to regularize these averages over the replicas, some time averagings may be used (as in (4.100)) such as

$$\langle h\rangle_{t,z} \simeq \frac{\displaystyle\int_0^t K_\tau(t-s)\left[\sum_{i=1}^M h(Q_s^{i,M})\delta_z^\epsilon(\xi(Q_s^{i,M}))\right] ds}{\displaystyle\int_0^t K_\tau(t-s)\left[\sum_{i=1}^M \delta_z^\epsilon(\xi(Q_s^{i,M}))\right] ds}, \tag{4.101}$$

or

$$\langle h\rangle_{t,z} \simeq \int_0^t K_\tau(t-s)\left[\frac{\sum_{i=1}^M h(Q_s^{i,M})\delta_z^\epsilon(\xi(Q_s^{i,M}))}{\sum_{i=1}^M \delta_z^\epsilon(\xi(Q_s^{i,M}))}\right] ds, \tag{4.102}$$

with a convolution kernel $K_\tau(t)$. For instance, $K_\tau(t) = \mathbf{1}_{t\geq 0}\tau^{-1}\mathrm{e}^{-t/\tau}$. Many other regularizations relying on a (local) ergodicity property could of course be used.

Enhancing the sampling through a selection process

A general strategy to improve the straightforward parallel implementation (4.83) is to add a selection step to duplicate "innovating" replicas (replicas located in regions where the sampling of

the reaction coordinate is not sufficient), and kill "redundant" ones. One way to perform an efficient selection is to consider an additional jump process quantified by a field $S(t,z)$ over the reaction coordinate values. Each replica trajectory $(Q_s^{i,M})$ is then weighted by $\exp(\int_0^t S(s,\xi(Q_s^{i,M}))\,ds)$, which naturally gives birth/death probabilities for the selection mechanism, in the spirit of Sequential Monte Carlo (SMC) methods [84] or Quantum Monte Carlo methods (QMC) [13] (see also Section 4.2, especially for a possible numerical implementation using birth and death times). A possible choice is

$$S = c\frac{\partial_{zz}\psi_t^\xi}{\psi_t^\xi}, \tag{4.103}$$

where c is a positive constant. This method thus enhances replicas in the convex areas of the density ψ_t^ξ, where free energy barriers still need to be overcome. When convergence has occured, ψ_t^ξ is uniform and the selection mechanism vanishes.

Consider for example the modified overdamped Langevin dynamics

$$\begin{aligned} dQ_t = &-\nabla(V + 2\beta^{-1}\ln|\nabla\xi| - F_{\text{bias}}(t,\xi))(Q_t)\,|\nabla\xi|^{-2}(Q_t)\,dt \\ &+\sqrt{2\beta^{-1}}\,|\nabla\xi|^{-1}(Q_t)\,dW_t, \end{aligned} \tag{4.104}$$

with the update (4.99): $F'_{\text{bias}}(t,z) = \langle f^V\rangle_{t,z}$. The process W_t is the standard Brownian motion. This dynamics is the usual overdamped Langevin dynamics for the potential $\mathcal{V}_t$ when $|\nabla\xi| = 1$. Notice that in the case of a metadynamics-like implementation ('m-ABF'), the modified dynamics is actually the usual overdamped Langevin dynamics since $\xi_{\text{meta}}(q,z) = z$ and thus $|\nabla\xi_{\text{meta}}| = 1$. For the dynamics (4.104), the distribution ψ_t^ξ of the reaction coordinate satisfies (see Section 4.4.2)

$$\partial_t\psi_t^\xi = \beta^{-1}\partial_{zz}\psi_t^\xi.$$

When the selection step is used with the overdamped Langevin dynamics (4.104), it can be shown that the distribution of the reaction coordinate values ψ_t^ξ still satisfies a simple diffusion equation, but with a higher diffusion constant:

$$\partial_t\psi_t^\xi = (\beta^{-1} + c)\partial_{zz}\psi_t^\xi.$$

This method thus enhances the diffusion in the reaction coordinate space, but the convergence rate is still limited by the relaxation in each submanifold $\xi(q) = z$.

Numerical results

We finally present an application of the selection strategy proposed above to the model system of conformational change in solution of Section 4.1.4. In practice, the Dirac distribution are approximated by indicator functions of intervals of size $\Delta z = 0.05$. The parameters used for these computations are $N = 16$ particles, at particle density $\rho = N/l^2 = 0.25\sigma^{-2}$, $\sigma = 1$, $w = 0.7$, $\epsilon = 1$ and $h = 20$, $\beta = 5$. We consider $M = 2000$ replicas evolving according to an overdamped Langevin dynamics, with a time step $\Delta t = 10^{-4}$. The reference computation is done with $M = 5000$ replicas and averaging the mean force profile on the time interval $[5, 10]$. The profiles are regularized in time by using (4.102) with $\tau/\Delta t = 100$. The initial conditions are such that the dimer bond lengths of all replicas are close to r_0. We consider in the sequel the interval $[z_0, z_1] = [1.1, 2.55]$ (since $r_0 \simeq 1.122$, $r_0 + 2w \simeq 2.522$ and $\Delta z = 0.05$), containing $n = 30$ bins.

We present in Figure 4.15 free energy difference profiles (averaged over $K = 100$ independent realizations) obtained with the parallel ABF dynamics (4.99), with and without the birth/death selection term (4.103) (with $c = 10$), at a fixed time $t_{\text{figure}} = 0.1$. The standard deviation of the profiles $(F'_1, \ldots, F'_K)$ for K independent realizations is

$$\sigma_{F'}(z) = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(F'_k(z) - \mathcal{F}'(z))^2},$$

where $\mathcal{F}'(z) = \frac{1}{K}\sum_{k=1}^{K} F'_k(z)$ is the mean force averaged over all the realizations. The associated 95% confidence intervals (or errors bars) are

$$[\mathcal{F}'_-(z),\ \mathcal{F}'_+(z)] = \left[\mathcal{F}'(z) - \frac{1.96}{\sqrt{K}}\sigma_{F'}(z),\ \mathcal{F}'(z) + \frac{1.96}{\sqrt{K}}\sigma_{F'}(z)\right]. \tag{4.105}$$

The curves plotted in solid lines in Figure 4.15 are the averages $\mathcal{F}'$, and the curves plotted in dashed lines are $\mathcal{F}'_-$ and $\mathcal{F}'_+$. Notice that the mean force profile obtained when the selection process is turned on is converged (since the curves $\mathcal{F}'$, $\mathcal{F}'_-$, $\mathcal{F}'_+$ and the reference curve are almost indistinguishable).

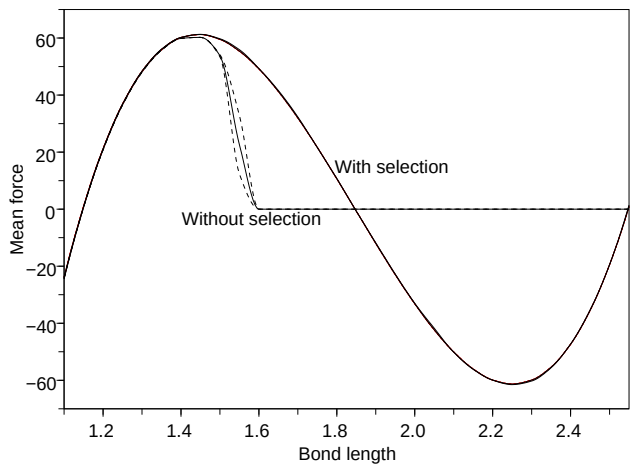


Fig. 4.15. Free energy difference profiles obtained with the parallel ABF algorithm (in reduced units), for a time $t_{\text{figure}} = 0.1$ and averaged over $K = 100$ independent realizations: with birth/death process ($c = 10$) and without birth/death process. The curve corresponding to the reference computation coincides with the curve obtained when the selection is turned on. Solid line: average mean force; dashed lines: upper and lower bounds of the 95% confidence intervals (see Eq. (4.105)).

The comparison with the reference profile shows that the selection process improves the rate of convergence of the algorithm and accelerates the exploration process on the free energy surface. Indeed, the profile obtained when the selection process is turned on is very quickly really close to the reference profile. On the other hand, with a straightforward parallelization, only a small fraction of replicas has escaped from the initial free energy metastable state at time t_{figure} to explore the free energy metastable set corresponding to bond lengths around $r_0 + 2w$.

To precise these qualitative features, we further perform two quantitative studies for several values of c:

(i) Tables 4.10 and 4.11 make precise the convergence of the profiles to the reference profile in a quantitative way. The measure of error we consider is

$$\delta F = \max_{z_0 \le z \le z_1} |\mathcal{F}(z) - F_{\text{ref}}(z)|,$$

where F_{ref} is the reference profile, and $\mathcal{F}(z) = \int_{z_1}^{z} \mathcal{F}'$ is the averaged potential of mean force, obtained as the integral of the mean force averaged over all the realizations. In practice, we consider the following approximated deviation between PMF profiles:

$$\delta F_n = \max_{0 \le i \le n} \left| \sum_{j=1}^{i} \mathcal{F}'(s_j) - F'_{\text{ref}}(s_j) \right| \Delta z. \tag{4.106}$$

A 95% confidence interval is obtained as $[\delta^- F_n, \delta^+ F_n]$, with

$$\delta^{\pm} F_n = \max_{0 \le i \le n} \left| \sum_{j=1}^{i} \mathcal{F}'(s_j) \pm \frac{1.96}{\sqrt{K}} \sigma_{F'}(s_j) + F'_{\text{ref}}(s_j) \right| \Delta z.$$

(ii) Figure 4.16 presents the fraction of replicas which have crossed the free-energy barrier (averaged aver the $K = 100$ realizations), *i.e.* the instantaneous fraction of particles such that $r \ge r_0 + w$. Notice that we expect this fraction to converge to 0.5 (up to some errors due to statistical fluctuations and to the binning of $[z_0, z_1]$).

Table 4.10. Deviation δF_n from the reference PMF profile (given by Eq. (4.106)) as a function of the selection parameter c ($c = 0$ when the selection is turned off) and the simulation time t_{simu}. The 95% confidence interval $[\delta^- F_n,\ \delta^+ F_n]$ is given in brackets.

c	$t_{\text{simu}} = 0.05$	0.1	0.2	0.4
0	9.51 (7.73-11.3)	18.0 (14.8-21.2)	19.5 (18.3-20.7)	0.066 (0.056-0.075)
2	20.4 (17.0-23.8)	5.69 (5.55-5.82)	0.020 (0.016-0.023)	0.034 (0.029-0.038)
5	22.9 (20.9-24.9)	0.22 (0.19-0.25)	0.027 (0.022(0.032)	0.026 (0.022-0.031)
10	10.4 (10.4-10.4)	0.035 (0.029-0.041)	0.028 (0.023-0.032)	0.032 (0.027-0.037)

Table 4.11. Deviation δF_n from the reference PMF profile (and associated error bars) when $c = 10$ for different number of replicas ($K = 50$ realizations).

number of replicas	$t_{\text{simu}} = 0.05$	0.1	0.4
1000	23.3 (20.4-26.3)	0.45 (0.39-0.50)	0.064 (0.054-0.074)
2000	11.2 (11.2-11.2)	0.034 (0.025-0.042)	0.032 (0.024-0.039)
10,000	2.05 (1.54-2.56)	0.026 (0.019-0.033)	0.022 (0.016-0.028)

As can be seen from the different escaping profiles of Figure 4.16, the selection process really accelerates the transition from one free energy metastable state to the other. This is due to the fact that the birth and death jump process triggers non local moves, as opposed to the traditional diffusive exploration of adaptive dynamics. The numerical results of Table 4.10 show that it is very interesting to consider a selection process, especially at the early stages of the simulation. This selection is even more efficient when the number of replicas increases (see Table 4.11). In conclusion, the selection process seems to be an efficient tool to improve the exploration power of the adaptive dynamics.

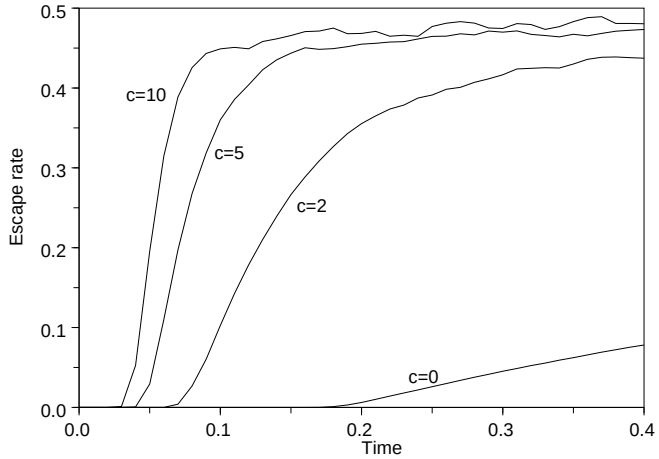


Fig. 4.16. Average fraction of the replicas in the region $r \geq r_0 + w$ as a function of time, for $c = 0$ (no selection), $c = 2$, $c = 5$, $c = 10$.

4.4.2 Rigorous convergence results for the Adaptive Biasing Force method

We present in this section a proof of convergence for the following dynamics, which is of ABF type:

$$dQ_t = -\nabla(V - F_{\text{bias}}(t,\xi) + 2\beta^{-1}\ln(|\nabla\xi|))(Q_t)\,|\nabla\xi|^{-1}(Q_t)\,dt + \sqrt{2\beta^{-1}}|\nabla\xi|^{-1}(Q_t)dW_t, \quad (4.107)$$

using the update (4.99) for the biasing force, that is

$$F'_{\text{bias}}(t,z) = \frac{\displaystyle\int_{\mathcal{M}} f^V(q)\,\psi_t(q)\,\delta_{\xi(q)-z}}{\displaystyle\int_{\mathcal{M}} \psi_t(q)\,\delta_{\xi(q)-z}}. \quad (4.108)$$

We assume in this section that the density ψ_t of the distribution of X_t is well-defined at all times. The proof presented here is actually restricted to the case

$$q = (z,\tilde{q}) \in \mathcal{M} = \mathbb{T} \times \mathbb{R}^{n-1}, \qquad \xi(q) = z,$$

$\mathbb{T}$ denoting the one-dimensional torus $\mathbb{R}/\mathbb{Z}$. In this case, $\Sigma_z = \{(z,\tilde{q}),\ \tilde{q} \in \mathbb{R}^{n-1}\}$, and $|\nabla\xi(q)| = 1$ so that the dynamics considered coincides with the usual overdamped dynamics when the biasing term is added. The case of a general one-dimensional reaction coordinate $\xi : \mathbb{R}^n \to \mathbb{R}$ is treated in [A1], where a convergence result for higher dimensional reaction coordinates is also stated, provided the temperature is large enough.

After a brief review on the most important results for convergence results relying on entropy estimates, we present a mathematical convergence result in the simplified setting considered in this section, and finally give the corresponding proof.

Some background on logarithmic Sobolev inequalities and their applications in statistical physics

The aim of this preliminary section is to give some background on entropy techniques with a focus on logarithmic Sobolev inequalities, which can be used to show the convergence to the equi-

librium state. More material can be read in the review papers by Guionnet and Zegarlinski [143], Ledoux [202] and Arnold, Markowich, Toscani and Unterreiter [10] (this last paper having rather a PDE approach).

For simplicity, we will consider an invariant measure of Boltzmann-Gibbs type, having a density with respect to the Lebesgue measure:

$$\psi_\infty(q)\,dq = Z^{-1}\,\mathrm{e}^{-\beta V(q)}\,dq, \qquad Z = \int_{\mathcal{M}} \mathrm{e}^{-\beta V(q)}\,dq,$$

and the overdamped Langevin dynamics on the configuration space $\mathcal{M}$:

$$dQ_t = -\nabla V(Q_t)\,dt + \sqrt{\frac{2}{\beta}}\,dW_t. \tag{4.109}$$

It can be assumed without loss of generality that $\beta = 1$ (replacing the potential V by βV). The density $\psi(t,\cdot) \equiv \psi_t(\cdot)$ of the law of Q_t evolves according to the Fokker-Planck equation

$$\partial_t \psi_t = \nabla \cdot \left(\psi_\infty \nabla \left(\frac{\psi_t}{\psi_\infty} \right) \right).$$

Notice that ψ_t is the density of a probability measure, so that $\int_{\mathcal{M}} \psi_t = 1$. Since ψ_∞ is a stationary solution of the above equation, it is expected that $\psi_t(q) \to \psi_\infty(q)$ as $t \to +\infty$. This is indeed the case when the dynamics is ergodic and an exponential rate of convergence can even be obtained when a convenient Lyapounov function can be found (see Section 3.2.3). However, the Lyapounov condition (3.45) may be difficult to check.

An alternative way to obtain exponential convergence of the density ψ_t to the target density is to resort to entropy estimates. Consider the convex function

$$\Phi(x) = x \ln x - x + 1,$$

and define the relative entropy of ψ_t with respect to ψ_∞ as

$$H(\psi_t \,|\, \psi_\infty) = \int_{\mathcal{M}} \Phi\left(\frac{\psi_t}{\psi_\infty}\right) \psi_\infty = \int_{\mathcal{M}} \ln\left(\frac{\psi_t}{\psi_\infty}\right) \psi_t \tag{4.110}$$

since $\int_{\mathcal{M}} \psi_t = 1$. Jensen's inequality shows that

$$H(\psi_t \,|\, \psi_\infty) = \int_{\mathcal{M}} \Phi\left(\frac{\psi_t}{\psi_\infty}\right) \psi_\infty \geq \Phi\left(\int_{\mathcal{M}} \frac{\psi_t}{\psi_\infty}\, \psi_\infty \right) = \Phi(1) = 0.$$

An alternative proof of the non-negativity of the entropy can be done by remarking that $\Phi \geq 0$. Actually, $\Phi(x) > 0$ if and only of $x \neq 1$, so that $H = 0$ if and only if $\psi_t = \psi_\infty$ almost everywhere.

Straightforward computations also show that

$$\frac{d}{dt} H(\psi_t \,|\, \psi_\infty) = -I(\psi_t \,|\, \psi_\infty), \tag{4.111}$$

where I is the Fisher information of ψ_t with respect to ψ_∞: Denoting $f_t = \psi_t/\psi_\infty$,

$$I(\psi_t \,|\, \psi_\infty) = \int_{\mathcal{M}} \frac{|\nabla f_t|^2}{f_t}\, \psi_\infty \geq 0.$$

Equality (4.111) therefore implies the decay of the relative entropy. An exponential decay rate can be obtained when ψ_∞ satisfies a logarithmic Sobolev inequality (LSI) with constant ρ.

Definition 4.1. *The probability measure $\psi_\infty(q)\,dq$ satisfies a logarithmic Sobolev inequality with constant $\rho > 0$ (in short: LSI(ρ)) if*

$$\forall f \in \mathrm{L}^1(\psi_\infty),\ f \geq 0,\ \int_{\mathcal{M}} f\psi_\infty = 1, \qquad \int_{\mathcal{M}} \Phi(f)\,\psi_\infty \leq \frac{1}{2\rho}\int_{\mathcal{M}} \frac{|\nabla f|^2}{f}\,\psi_\infty. \tag{4.112}$$

In other words, for all probability measures absolutely continuous with respect to the Lebesgue measure, with density $\phi(q)\,dq$,

$$H(\phi\,|\,\psi_\infty) \leq \frac{1}{2\rho} I(\phi\,|\,\psi_\infty).$$

Then, combining (4.111) and (4.112), it follows, using a Gronwall inequality:

$$0 \leq H(\psi_t\,|\,\psi_\infty) \leq H(\psi_0\,|\,\psi_\infty)\,\mathrm{e}^{-2\rho t}.$$

The convergence $\psi_t \to \psi_\infty$ can be precised using the Csizár-Kullback inequality:

$$\int_{\mathcal{M}} |\psi_t - \psi_\infty| \leq 2\sqrt{H(\psi_t\,|\,\psi_\infty)},$$

which implies an exponentially fast convergence of ψ_t to ψ_∞ in $\mathrm{L}^1(\mathcal{M})$.

Obtaining logarithmic Sobolev inequalities

To prove convergence results for the density of the process such as (4.109), it therefore suffices to show that a LSI of the form (4.112) holds for the target measure $\psi_\infty(q)\,dq = Z^{-1}\,\exp(-V(q))\,dq$ (recall that we assumed $\beta = 1$ thoughout this section). A LSI can for instance be obtained in the following cases:

(i) when the potential V satisfies a strict convexity condition of the form $\mathrm{Hess}(V) \geq \rho\,\mathrm{Id}$ with $\rho > 0$, then a LSI with constant ρ holds, as first shown by Bakry and Emery [19];

(ii) when $\psi_\infty = \prod_{i=1}^M \psi_\infty^i$ and each measure $\psi_\infty^i(q)\,dq$ satisifies a LSI with constant ρ_i, then ψ_∞ satisfies a LSI with constant $\rho = \min\{\rho_1, \ldots, \rho_M\}$ (see Gross [139]);

(iii) when a LSI with constant ρ is satisfied by $Z_V^{-1}\,\mathrm{e}^{-V(q)}\,dq$, then $Z_{V+W}^{-1}\,\mathrm{e}^{-(V(q)+W(q))}\,dq$ (with W bounded) satisfies a LSI with constant $\tilde{\rho} = \rho\,\mathrm{e}^{\inf W - \sup W}$. This property expresses some stability with respect to bounded pertubations (see Holley and Stroock [169]);

(iv) there are also results on a global LSI for the measure when a marginal and the corresponding conditional law satisfy a LSI (see Blower and Bolley [33]), or when all the marginals satisfy a LSI under some weak coupling assumption (see Otto and Reznikoff [263]).

A PDE formulation and a precise statement of the result

Since only the law of the process Q_t at a fixed time t is used in (4.107)-(4.108), it is possible to recast the dynamics in terms of a nonlinear partial differential equation (PDE) on the density $\psi(t,\cdot)$ of Q_t (recall that $\xi(q) = \xi(z,\tilde{q}) = z$):

$$\boxed{\begin{cases} \partial_t \psi = \mathrm{div}\left(\nabla(V - F_{\mathrm{bias}}(t,z))\psi + \beta^{-1}\nabla\psi\right), \\[2ex] F'_{\mathrm{bias}}(t,z) = \dfrac{\displaystyle\int_{\mathbb{R}^{n-1}} \partial_z V(z,\tilde{q})\,\psi(t,z,\tilde{q})\,d\tilde{q}}{\displaystyle\int_{\mathbb{R}^{n-1}} \psi(t,z,\tilde{q})\,d\tilde{q}}. \end{cases}} \tag{4.113}$$

Measure of the convergence

Let us introduce the longtime limit of the distribution of X_t:

$$\psi_\infty = \exp(-\beta(V - F \circ \xi)),$$

and the longtime limit of the marginal and conditional laws:

$$\psi_\infty^\xi(z) = \int_{\mathbb{R}^{n-1}} \psi_\infty(z, \tilde{q})\, d\tilde{q} \equiv 1, \qquad d\mu_{\infty,z}(\tilde{q}) = \frac{\psi_\infty(z, \tilde{q})\, d\tilde{q}}{\psi_\infty^\xi(z)}.$$

The "distance" between ψ (respectively ψ^ξ) and ψ_∞ (respectively ψ_∞^ξ) is measured using the relative entropy $H(\psi|\psi_\infty)$ defined in (4.110) (respectively $H(\psi^\xi|\psi_\infty^\xi)$). In the following, the "total" entropy is denoted by

$$E(t) = H(\psi(t, \cdot)|\psi_\infty),$$

the "macroscopic entropy" by

$$E_M(t) = H(\psi^\xi(t, \cdot)|\psi_\infty^\xi),$$

the "local entropy" at a fixed value z of the reaction coordinate by

$$e_m(t, z) = H(\mu_{t,z}|\mu_{\infty,z}) = \int_{\mathbb{R}^{n-1}} \ln\left(\frac{\psi(t, z, \tilde{q})}{\psi^\xi(t, z)} \Big/ \frac{\psi_\infty(z, \tilde{q})}{\psi_\infty^\xi(z)}\right) \frac{\psi(t, z, \tilde{q})\, d\tilde{q}}{\psi^\xi(t, z)},$$

and finally the "microscopic entropy" by

$$E_m(t) = \int_{\mathbb{T}} e_m(t, z)\psi^\xi(t, z)\, dz.$$

It is straightforward to obtain the following result which can be seen as a property of extensivity of the entropy:

Lemma 4.2 (Extensivity of the entropy). *The total entropy can be decomposed as the sum of the macroscopic and the microscopic entropies:*

$$E(t) = E_M(t) + E_m(t).$$

Remark 4.5 (On the choice of the entropy). *In the case of linear Fokker Planck equations, it is well known that one can obtain exponential decay to equilibrium by considering various entropies of the form* $\int h\left(\frac{d\mu}{d\nu}\right) d\mu$, *where* h *is typically a strictly convex function such that* $h(1) = 0$ *(see [10] for more assumptions required on* h*). For example, the classical choice* $h(x) = \frac{1}{2}(x-1)^2$ *is linked to Poincaré type inequalities and leads to* L^2*-convergence, while the function* $h(x) = x \ln x - x + 1$ *used here to build the entropy is linked to logarithmic Sobolev inequalities and leads to* $L^1 \ln L^1$*-convergence. However, for the study of the non-linear Fokker Planck equation (4.113), it seems that the choice* $h(x) = x \ln x - x + 1$ *is important to derive the estimates, since the extensivity property of Lemma 4.2 is fundamental for the proof presented here.*

Let us also introduce another way to compare two probability measures, namely the Wasserstein distance with quadratic cost:

$$W(\mu, \nu) = \sqrt{\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^{n-1} \times \mathbb{R}^{n-1}} |\tilde{q} - \tilde{q}'|^2\, d\pi(\tilde{q}, \tilde{q}')}$$

where $\Pi(\mu, \nu)$ denotes the set of coupling probability measures, namely probability measures on $\mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ such that their marginals are μ and ν. We need the following definition:

Definition 4.2. *The probability measure ν satisfies a Talagrand inequality with constant $\rho > 0$ (in short: T(ρ)) if for all probability measures μ such that $\mu \prec \nu$ (i.e. μ is absolutely continuous with respect to ν),*

$$W(\mu, \nu) \leq \sqrt{\frac{2}{\rho} H(\mu|\nu)}.$$

In the last definition, we implicitly assume that the probability measures have finite moments of order 2, which will be always the case for all the probability measures we consider. We will need the following important result (see [264, Theorem 1]).

Lemma 4.3. *If ν satisfies LSI(ρ), then ν satisfies T(ρ).*

Convergence results

Proposition 4.6. *The marginal ψ^ξ satisfies the following diffusion equation on $\mathbb{T}$:*

$$\partial_t \psi^\xi = \frac{1}{\beta} \partial_{z,z} \psi^\xi \tag{4.114}$$

and

$$\forall t \geq 0, \qquad I(\psi(t, \cdot) \,|\, \psi_\infty) \leq I(\psi(0, \cdot) \,|\, \psi_\infty) \exp(-8\pi^2 \beta^{-1} t). \tag{4.115}$$

The proof of (4.114) is straightforward (by inegrating (4.113) with respect to $\tilde{q} \in \mathbb{R}^{n-1}$), and implies the convergence of the marginals (see Lemma 4.4 for the complete proof of this proposition). To prove the global convergence, we need some additional assumptions (on the potential V):

Theorem 4.3. *Let $(\psi, F'_{\text{bias}}(t))$ be a smooth solution to (4.113), and assume*

(H1) The function V is such that $\|\partial_{z,\tilde{q}} V\|_{L^\infty} \leq M < \infty$;

(H2) There exists $\rho > 0$ such that for all $z \in \mathcal{M}$, the conditional measure $\mu_{\infty,z}$ satisfies LSI(ρ).

Then,

(i) the "microscopic entropy" E_m satisfies

$$E_m(t) \leq C^2 \exp(-2\lambda t) \tag{4.116}$$

where $C = 2 \max\left(\sqrt{E_m(0)}, M\beta|\rho - 4\pi^2|^{-1} \sqrt{\frac{I_0}{2\rho}}\right)$ with $I_0 = I(\psi(0, \cdot) \,|\, \psi_\infty)$, and

$$\lambda = \beta^{-1} \min(\rho, 4\pi^2).$$

In the special case $\rho = 4\pi^2$, it holds $\sqrt{E_m(t)} \leq \left(\sqrt{E_m(0)} + M\sqrt{\frac{I_0}{2\rho}}\, t\right) \exp(-4\pi^2 \beta^{-1}\, t)$.

(ii) The mean force observed at time t $F'_{\text{bias}}(t)$ converges to the mean force F' in the following sense:

$$\forall t \geq 0, \qquad \int_{\mathbb{T}} |F'_{\text{bias}}(t) - F'|^2(z) \psi^\xi(t, z)\, dz \leq \frac{2M^2}{\rho} E_m(t). \tag{4.117}$$

Therefore, there exist $\overline{C}, \overline{t} > 0$ such that

$$\forall t \geq \overline{t}, \qquad \int_{\mathbb{T}} |F'_{\text{bias}}(t) - F'|(z)\, dz \leq \overline{C} \exp(-\lambda t). \tag{4.118}$$

This theorem therefore shows that $F'_{\text{bias}}(t)$ converges exponentially fast to F' at a rate $\lambda = \beta^{-1} \min(\rho, 4\pi^2)$. The limitations on the rate λ are linked to the rate of convergence at the macroscopic level, on the equation (4.114) satisfied by ψ^ξ, and the rate of convergence at the microscopic level, which depends on the constant ρ of the logarithmic Sobolev inequalities satisfied by the conditional measures $\mu_{\infty,z}$. This constant depends of course on the choice of the reaction

coordinate. In our framework, we could state that a "good reaction coordinate" is such that ρ is as large as possible.

Notice also that a consequence of (4.116), (4.115) and Lemma 4.2 is that the "total entropy" E also decays exponentially fast to zero, with the same rate λ. Therefore, by the Csiszár-Kullback inequality, $\psi(t,\cdot)$ converges exponentially fast to ψ_∞ in $L^1(\mathbb{R}^n)$ norm.

Remark 4.6 (On the initial condition). *If $\psi^\xi(0,\cdot)$ is zero at some points or is not sufficiently smooth, then $F'_{\text{bias}}(0)$ may be not well defined or $I(\psi^\xi(0,\cdot)|\psi^\xi_\infty)$ may be infinite. But since we show that ψ^ξ satisfies a simple diffusion equation (see item 1 in Theorem 4.3), these difficulties disappear as soon as $t>0$. Therefore, up to considering the problem for $t \geq t_* > 0$, we can suppose that $\psi^\xi(0,\cdot) > 0$.*

It can be checked that the assumptions (H1) and (H2) are satisfied in this context for a potential V of the following form:

$$V(z,\tilde{q}) = V_0(z,\tilde{q}) + V_1(z,\tilde{q})$$

where $\alpha = \inf_{\mathbb{T}\times\mathbb{R}^{n-1}} \partial_{\tilde{q},\tilde{q}} V_0 > 0$, $\|V_1\|_{L^\infty} < \infty$, $\|\partial_{z,\tilde{q}}(V_0+V_1)\|_{L^\infty} < \infty$, with the choice $M = \|\partial_{z,\tilde{q}} V\|_{L^\infty}$, $\rho = (\inf_{\mathbb{T}\times\mathbb{R}^{n-1}} \partial_{\tilde{q},\tilde{q}} V_0)\exp(-\text{osc } V_1)$, where osc $V_1 = \sup_{\mathbb{T}\times\mathbb{R}^{n-1}} V_1 - \inf_{\mathbb{T}\times\mathbb{R}^{n-1}} V_1$. In words, the potential V is a uniformly α-convex potential in the $\tilde{q}$ variable (therefore satisfying a LSI thanks to the Bakry-Emery criterion), perturbed by some bounded potential. The (almost) α-convexity in the variables orthogonal to the reaction coordinate is indeed natural enough since it is expected that the metastable features of the potential are in the reaction coordinate variable.

Proofs of Proposition 4.6 and Theorem 4.3

To simplify the presentation of the proof, we assume $\beta = 1$, up to the following change of variable: $\tilde{t} = \beta^{-1}t$, $\tilde{\psi}(\tilde{t},q) = \psi(t,q)$, $\tilde{V}(q) = \beta V(q)$.

Lemma 4.4 (Convergence of the Fisher information). *Let ϕ be a positive function defined for $t \geq 0$ and $z \in \mathbb{T}$, satisfying*

$$\partial_t \phi = \partial_{z,z}\phi \quad \text{on } \mathbb{T}, \qquad \int_{\mathbb{T}} \phi = 1. \tag{4.119}$$

Denoting by $\phi_\infty \equiv 1$ the longtime limit of ϕ, it holds

$$\forall t \geq 0, \qquad I(\phi(t,\cdot)\,|\,\phi_\infty) \leq I(\phi(0,\cdot)\,|\,\phi_\infty)\exp(-8\pi^2 t).$$

Proof. Denoting by $u = \sqrt{\phi}$, it follows

$$I(\phi\,|\,\phi_\infty) = \int_{\mathbb{T}} |\partial_z \ln\phi|^2 \phi = 4\int |\partial_z u|^2.$$

Moreover, from the diffusion equation (4.119),

$$\partial_t u = \partial_{z,z} u + \frac{(\partial_z u)^2}{u}.$$

Therefore,

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{T}}(\partial_z u)^2 &= 2\int_{\mathbb{T}}\partial_{z,z,z}u\,\partial_z u + 2\int_{\mathbb{T}}\partial_z\left(\frac{(\partial_z u)^2}{u}\right)\partial_z u,\\
&= -2\int_{\mathbb{T}}(\partial_{z,z}u)^2 - 2\int_{\mathbb{T}}\frac{(\partial_z u)^2}{u}\partial_{z,z}u,\\
&= -2\int_{\mathbb{T}}(\partial_{z,z}u)^2 - 2\int_{\mathbb{T}}\frac{\partial_z((\partial_z u)^3)}{3u},\\
&= -2\int_{\mathbb{T}}(\partial_{z,z}u)^2 - \frac{2}{3}\int_{\mathbb{T}}\frac{(\partial_z u)^4}{u^2},
\end{aligned}$$

so that finally

$$\frac{d}{dt}\int_{\mathbb{T}}(\partial_z u)^2 \le -8\pi^2\int_{\mathbb{T}}(\partial_z u)^2,$$

where we have used the Poincaré-Wirtinger inequality on $\mathbb{T}$, applied to $\partial_z u$: For any function $f \in H^1(\mathbb{T})$,

$$\int_{\mathbb{T}}\left(f - \int_{\mathbb{T}} f\right)^2 \le \frac{1}{4\pi^2}\int_{\mathbb{T}}(\partial_z f)^2.$$

This Poincaré inequality is obtained by studying the spectral gap of the operator $\partial_{z,z}$ on $[0,1]$. □

We now turn to the proof of Theorem 4.3. One fundamental lemma for the following is

Lemma 4.5. *The difference between the "current mean force" $F'_{\text{bias}}(t)$ and the mean force F' can be expressed in term of the densities as*

$$F'_{\text{bias}}(t) - F' = \int_{\mathbb{R}^{n-1}}\partial_z \ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,d\tilde{q} - \partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right).$$

Proof. This is a simple computation:

$$\begin{aligned}
\int_{\mathbb{R}^{n-1}}\partial_z \ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,d\tilde{q} - \partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right) &= \int_{\mathbb{R}^{n-1}}\partial_z \ln\psi\frac{\psi}{\psi^\xi}\,d\tilde{q} - \int_{\mathbb{R}^{n-1}}\partial_z \ln\psi_\infty\frac{\psi}{\psi^\xi}\,d\tilde{q} - \partial_z \ln\psi^\xi,\\
&= \int_{\mathbb{R}^{n-1}}\frac{\partial_z\psi}{\psi^\xi}\,d\tilde{q} + \int_{\mathbb{R}^{n-1}}\partial_z(V-F)\frac{\psi}{\psi^\xi}\,d\tilde{q} - \partial_z \ln\psi^\xi,\\
&= F'_{\text{bias}}(t) - F',
\end{aligned}$$

which concludes the proof. □

We will also use the following estimates:

Lemma 4.6. *Under the assumptions (H1)–(H2), it holds, for all $t \ge 0$ and for all $z \in \mathbb{T}$,*

$$|F'_{\text{bias}}(t,z) - F'(z)| \le \|\partial_{z,\tilde{q}}V\|_{L^\infty}\sqrt{\frac{2}{\rho}e_m(t,z)}.$$

Proof. For any coupling measure $\pi \in \Pi(\mu_{t,z}, \mu_{\infty,z})$,

$$\begin{aligned}
|F'_{\text{bias}}(t,z) - F'(z)| &= \left|\int_{\mathbb{R}^{n-1}\times\mathbb{R}^{n-1}}\partial_z V(z,\tilde{q}) - \partial_z V(z,\tilde{q}')\,\pi(d\tilde{q},d\tilde{q}')\right|,\\
&\le \|\partial_{z,\tilde{q}}V\|_{L^\infty}\int|\tilde{q}-\tilde{q}'|\,\pi(d\tilde{q},d\tilde{q}')\\
&\le \|\partial_{z,\tilde{q}}V\|_{L^\infty}\sqrt{\int|\tilde{q}-\tilde{q}'|^2\,\pi(d\tilde{q},d\tilde{q}')}.
\end{aligned}$$

Taking now the infimum over all $\pi \in \Pi(\mu_{t,z}, \mu_{\infty,z})$ and using (2) together with Lemma 4.3, it follows

$$|F'_{\text{bias}}(t,z) - F'(z)| \leq \|\partial_{z,\tilde{q}} V\|_{L^\infty} W(\mu_{t,z}, \mu_{\infty,z}) \leq \|\partial_{z,\tilde{q}} V\|_{L^\infty} \sqrt{\frac{2}{\rho} H(\mu_{t,z} \mid \mu_{\infty,z})},$$

which concludes the proof. □

Lemma 4.7. *When (H2) is satisfied,*

$$\forall t \geq 0, \qquad E_m(t) \leq \frac{1}{2\rho} \int_{\mathbb{T}\times\mathbb{R}^{n-1}} \left| \partial_z \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi.$$

Proof. Using (H2), it follows

$$E_m = \int_{\mathbb{T}} e_m \psi^\xi \, dz \leq \int_{\mathbb{T}} \frac{1}{2\rho} \int_{\mathbb{R}^{n-1}} \left| \partial_z \ln\left(\frac{\psi}{\psi^\xi} \Big/ \frac{\psi_\infty}{\psi^\xi_\infty}\right)\right|^2 \frac{\psi}{\psi^\xi} \, d\tilde{q} \; \psi^\xi \, dz,$$

which yields the result since ψ^ξ/ψ^ξ_∞ does not depend on $\tilde{q}$. □

We are now in position to prove the first assertion (4.116) of Theorem 4.3. The equation on ψ can be rewritten as:

$$\partial_t \psi = \operatorname{div}\left(\psi_\infty \nabla \left(\frac{\psi}{\psi_\infty}\right)\right) + \partial_x((F' - F'_{\text{bias}}(t))\psi).$$

Therefore, after integration by parts, using a Cauchy-Schwarz inequality and Lemma 4.5,

$$\begin{aligned}
\frac{d}{dt} E_m &= \frac{d}{dt} E - \frac{d}{dt} E_M, \\
&= -\int_{\mathcal{M}} \left|\nabla \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi + \int_{\mathcal{M}} (F'_{\text{bias}}(t) - F') \partial_z \ln\left(\frac{\psi}{\psi_\infty}\right) \psi + \int_{\mathbb{T}} \left|\partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi, \\
&= -\int_{\mathcal{M}} \left|\partial_{\tilde{q}} \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi - \int_{\mathcal{M}} \left|\partial_z \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi \\
&\quad + \int_{\mathbb{T}} \left(\int_{\mathbb{R}^{n-1}} \partial_z \ln\left(\frac{\psi}{\psi_\infty}\right) \psi \, d\tilde{q}\right)^2 \frac{1}{\psi^\xi} \, dz - \int_{\mathcal{M}} \partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right) \partial_z \ln\left(\frac{\psi}{\psi_\infty}\right) \psi \\
&\quad + \int_{\mathbb{T}} \left|\partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi, \\
&\leq -\int_{\mathcal{M}} \left|\partial_{\tilde{q}} \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi - \int_{\mathbb{T}} \partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right) \psi^\xi (F'_{\text{bias}}(t) - F').
\end{aligned}$$

Using now Lemmata 4.6 and 4.7,

$$\begin{aligned}
\frac{d}{dt} E_m &\leq -2\rho E_m + \sqrt{\int_{\mathbb{T}} |F'_{\text{bias}}(t) - F'|^2 \psi^\xi} \sqrt{\int_{\mathbb{T}} \left|\partial_z \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi}, \\
&\leq -2\rho E_m + \|\partial_{z,\tilde{q}} V\|_{L^\infty} \sqrt{\frac{2}{\rho} E_m} \sqrt{I(\psi^\xi \mid \psi^\xi_\infty)}.
\end{aligned}$$

With Lemma 4.4, it then follows

$$\frac{d}{dt} \sqrt{E_m} \leq -\rho \sqrt{E_m} + \|\partial_{z,\tilde{q}} V\|_{L^\infty} \sqrt{\frac{I(\psi^\xi(0,\cdot) \mid \psi^\xi_\infty)}{2\rho}} \exp(-4\pi^2 t),$$

from which (4.116) is deduced.

Let us now turn to the proof of the second item of Theorem 4.3. Notice first that $\|\psi(t,\cdot) - \psi_\infty\|_{\mathrm{L}^\infty} \to 0$ when $t \to +\infty$. This results from the exponentially fast $\mathrm{H}^1(\mathbb{R}^3)$ convergence of $\psi_t^\xi \to \psi_\infty^\xi$ (which can be proved using Lemma 4.4) and the inequality

$$\left\| f - \int_{\mathbb{T}} f \right\|^2_{\mathrm{L}^\infty(\mathbb{T})} \leq \int_{\mathbb{T}} (\partial_z f)^2$$

applied to $f = \psi^\xi$. Since $\psi_\infty^\xi \equiv 1$, it holds

$$\begin{aligned}\int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'| = \textstyle\int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|\psi_\infty^\xi &= \int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|\psi^\xi - \int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|(\psi^\xi - \psi_\infty^\xi) \\ &\leq \int |F'_{\mathrm{bias}}(t) - F'|^2 \psi^\xi + \|\psi(t,\cdot) - \psi_\infty\|_{\mathrm{L}^\infty} \int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|.\end{aligned}$$

Thus, for t sufficiently large, $\int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|$ is bounded from above by $c \int_{\mathbb{T}} |F'_{\mathrm{bias}}(t) - F'|^2 \psi^\xi$ (for some $c > 0$), which yields (4.118) (using (4.117) and (4.116)).

Part III

Shock Waves: a Multiscale Approach

5

A reduced model for shock waves

5.1 A simplified one-dimensional model **192**
5.1.1 Shock waves in one-dimensional lattices 192
5.1.2 An augmented one-dimensional model 197
5.1.3 The stochastic limit 205
5.1.4 Extension to the reactive case 209
5.2 A reduced model based on Dissipative Particle Dynamics **212**
5.2.1 Previous mesoscopic models 212
5.2.2 A reduced model in the inert case 213
5.2.3 The reactive case 218

Multimillion atom simulations are nowadays common in molecular dynamics (MD) studies. However, the time and space scales numerically tractable are still far from being macroscopic, so that reduced models are of primary interest when multiscale phenomena are considered. In particular, the simulation of shock waves is a challenging task, involving very small time and space scales and large energies near the shock front, and much larger time and space scales and lower energies for the relaxation of the shocked materials, including the evolution of dislocations loops for example.

The situation is even worse for detonation waves (Roughly speaking, a detonation wave is a shock wave combined with very exothermic chemical reactions, see [103] for a fundamental reference). The simulation of detonation requires the description of a thin shock front, moving at a high velocity, usually using a complicated empirical potential able to treat the chemical events happening (dissociation, recombination). To this end, toy molecular models were proposed at the early stages of the molecular simulation of detonation (see *e.g.* [269]), until the first all-atom studies in the 90's [38, 39]. Such computations are nowadays common (see for example [327] for a state of the art study), but are still limited in spatial and temporal sizes, so that a reduced model for detonations is of interest.

Some reduced models for shock waves were proposed, for polycrystalline materials [163] or resorting to mesoparticles with internal degrees of freedom [326] (see a brief overview of all those methods in Section 5.2.1). The latter approach seems to be the most promising and the most general one, and consists in replacing a complex molecule by a single particle. The introduction of an internal degree of freedom describing in a mean way the behavior of several degrees of freedom is reminiscent from Dissipative Particle Dynamics (DPD) models [98,170], which aim at describing complex fluids through some mesodynamics with some additional variables.

We present in this chapter reduced model for shock and detonation waves described at the microscopic level. Starting in Section 5.1 from a very simple one-dimensional (1D) model where the main features of shock waves are already present, we show how a model reduction of dimensionality

can be performed under some decoupling or low-coupling assumptions. Though the initial model is deterministic, the obtained model is stochastic: more precisely, the many-body interactions are replaced by some generalized friction (with memory) depending on the relative velocities of neighboring particles (which is reminiscent of DPD models), and the system is governed by a generalized Langevin equation instead of the usual Hamiltonian dynamics. However, the temperature jumps across the shock front are not reproduced correctly.

Building on this one-dimensional model, a simplified DPD dynamics preserving the total energy of the system is proposed in Section 5.2. Within such a model, temperature jumps across the shock front can be treated. It is also a convenient framework for an extension to chemically reactive shock waves (detonations).

5.1 A simplified one-dimensional model

We begin in Section 5.1.1 with some introduction to 1D lattice motion, and briefly report on some theoretical results and numerical experiments on piston-impacted shocks. It is shown that, in the absence of a specific treatment, the shock profiles generated significantly differ from shock waves. Especially, their thicknesses grow linearly with time [166,359], there is no usual equilibration downstream the shock front [87,168,359], and relaxation waves do not behave as expected. Indeed, one would expect the shock wave to be a self-similar jump separating two domains at local thermal equilibrium at different temperatures. The relaxation waves should then catch up the shock front and weaken the shock wave until it disappears. So, we have to introduce higher-dimensional effects, at least in an averaged way. This is performed in Section 5.1.2. The connection of the chain with a heat bath consisting of a large number of harmonic oscillators, seems to be a good remedy for spurious 1D effects. The shocks generated have constant thicknesses and relaxation waves appear to be properly modelled. We also present the stochastic limit of this model in Section 5.1.3, and an extension to the reactive case in Section 5.1.4.

5.1.1 Shock waves in one-dimensional lattices

The aim of this section is to derive and assess the validity of a simplified microscopic model of shock waves which can useful for a more general derivation. Shock waves are intrinsically propagative phenomena. It is thus reasonable to describe them within a 1D macroscopic theory. In some cases depending on the geometry, this approximation has proven to be correct [73].

A 1D lattice seems an appropriate model that could, in addition, allow for some mathematical treatment and thus a better theoretical understanding of the phenomena and mechanisms at play. Indeed, many mathematical results are known about the behavior of waves in 1D lattices, concerning the existence of localized waves [117,315], the form of those waves in the high-energy limit [115] or in the low-energy limit [116], or the behavior under shock [104]. There also exist extended results for a particular interaction between sites, the Toda potential [344] : the structure of a 1D shock is then precisely known, at least in some regime [359].

Description of the lattice model

Consider a one-dimensional chain of particles with nonlinear nearest-neighbor interactions, described by a potential V. Initially, the particles are at rest at positions $X_n(0) = nd$, which is an equilibrium state for the system. All the masses are set to 1. The normalized displacement of the n-th particle from its equilibrium position is $x_n(t) = \frac{1}{d}(X_n(t) - X_n(0))$. The following normalization conditions [166] for the interaction potential V can be used:

$$V(0) = 0, \quad V'(0) = 0, \quad V''(0) = 1. \tag{5.1}$$

The first condition is more a shift on the energy reference, the second one expresses the fact that $x = 0$ is the equilibrium position, and the last one amounts to a rescaling of time. The so-called "reduced relative displacement" is defined as $\delta x_n(t) = x_{n+1}(t) - x_n(t)$.

The Hamiltonian of the system is:

$$H_{\mathrm{S}}(\{q_n, p_n\}) = \sum_{n=-\infty}^{\infty} V(q_{n+1} - q_n) + \frac{1}{2}\dot{p}_n^2, \tag{5.2}$$

where $(q_n, p_n) = (x_n, \dot{x}_n)$. The Newton equations of motion read:

$$\ddot{x}_n = V'(x_{n+1} - x_n) - V'(x_n - x_{n-1}). \tag{5.3}$$

The potential taken here can either have a physical origin, like the 1D Lennard-Jones potential:

$$V_{\mathrm{LJ}}(x) = \frac{1}{8}\left(\frac{1}{(1+x)^4} - \frac{2}{(1+x)^2}\right), \tag{5.4}$$

or more mathematical motivations, like the one-parameter Toda potential [344]:

$$V^{\mathrm{b}}_{\mathrm{Toda}}(x) = \frac{1}{b^2}\left(\mathrm{e}^{-bx} - 1 + bx\right). \tag{5.5}$$

Define $b = -V'''(0)$. The parameter b measures at the first order the anharmonicity of the system. For the Lennard-Jones potential $b = 9$, and for the Toda potential, the parameter b introduced in the definition (5.5) is indeed equal to $-\dfrac{d^3V^{\mathrm{b}}}{dx^3}(0)$.

Shock waves in the 1D lattice

A brief review of the existing mathematical and numerical results

A shock can be generated using a "piston" : the first particle is considered as being of infinite mass and constantly moving at velocity u_p. We refer to [90] for a pioneering study of those shocks in 1D lattices, to [164,166,168] for careful numerical experiments and formal analysis, and to [359] for a rigorous mathematical study in the Toda case. All of these studies identify the parameter $a = bu_p$ as critical. When $a < 2$, the velocity of the downstream particles converge to the piston velocity, in analogy with the behavior of a harmonic lattice[1] (see Figure 5.1). When $a > 2$, the particles behind the shock experience an oscillatory motion (see Figure 5.2). This behavior is quite similar to what is happening in hard-rod fluids (see [168] for a more precise description of that phenomenon), and has to be linked to the exchange of momenta happening when two particles collide in a 1D setting. This was also noticed for other potentials such as the Lennard-Jones potential, and can be used to define specific 1D thermodynamical averages [87].

In the case of a strong shock ($a > 2$) and in the Toda case, the displacement pattern is particularly well understood from a mathematical point of view [359]: the lattice can be decomposed in three regions. In the first one, for $n > c_{\max}t$, the particles have "almost" not felt the shock yet, and their displacements are exponentially small. The second region, whose thickness grows linearly in time ($c_{\min}t < n < c_{\max}t$), is composed of a train of solitons. Recall that solitons are particular solutions of the Toda lattice model, and correspond to localized waves [344]. In the third region ($n < c_{\min}t$), the lattice motion converges to an oscillatory pattern of period 2 (binary wave). The motion behind the shock is asymptotically described by the evolution of a single oscillator (see [87] for a precise description of this behavior). There is no local thermal equilibrium in the usual sense (*i.e.* the distribution of the velocities is not of Boltzmann form). This was already mentioned in [168].

[1] Note that we use $b = 2\alpha$ with the notation of [166].

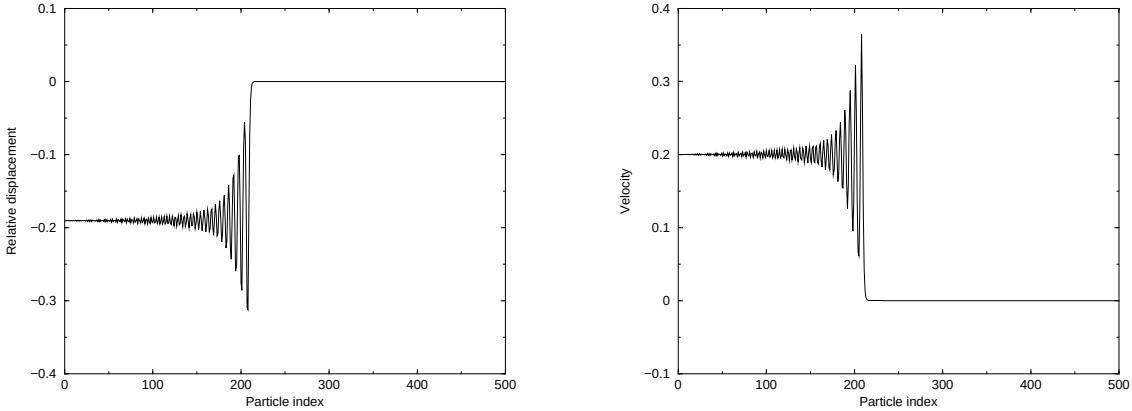


Fig. 5.1. Relative displacement (left) and velocity profiles (right) versus particle index for a weak shock at a representative time: number of particles $N_{\rm part} = 500$, Toda parameter $b = 1$, piston velocity $u_p = 0.2$, so that $a = 0.2$. The particle are taken initially at rest at their equilibrium positions.

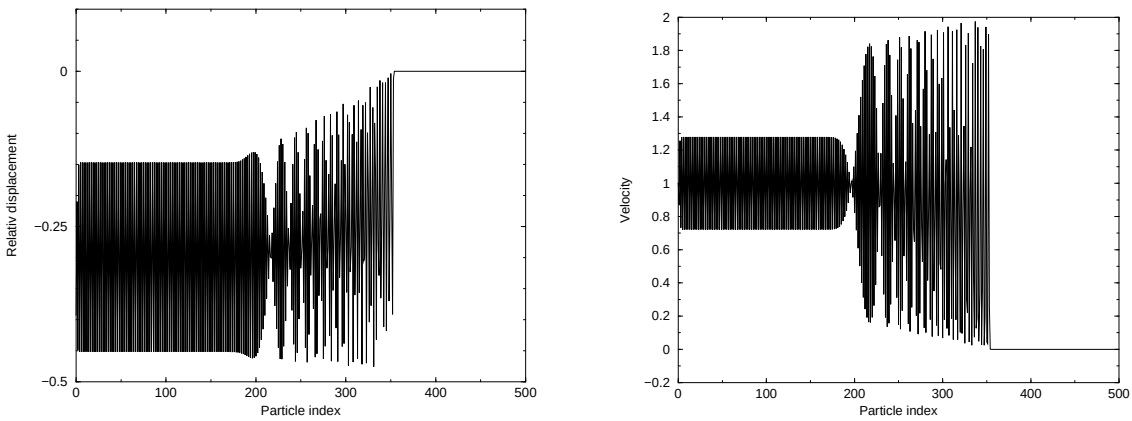


Fig. 5.2. Relative displacement (left) and velocity profiles (right) versus particle index for a strong shock at time $T = 100$: $b = 10$, $u_p = 1$, so that $a = 10$. The particles are initially at rest.

Density plots.

To get a better understanding of the shock patterns, it is convenient to represent the system in terms of local density. This local density can be obtained as a function of the local average of the interatomic distances, both in space and time. We restrict ourselves to a local average in space. More precisely, the local averaged interatomic distance of the n-th length is denoted by $\overline{\delta x_n}$, and given by $\overline{\delta x_n} = \sum_{i=-\infty}^{+\infty} \alpha_j \, \delta x_{n+j}$. The local density ρ_n is then defined as $\rho_n = \left(1 + \overline{\delta x_n}\right)^{-1}$. The weights $\{\alpha_j\}$ are chosen in practice to be non negative and of sum equal to one. For example: $\alpha_j = C^{-1} \cos\left(\frac{j}{2M+1}\pi\right)$ for $-M \le j \le M$, $\alpha_j = 0$ otherwise, and with $C = \sum_{j=-M}^{M} \cos\left(\frac{j}{2M+1}\pi\right)$. The integer M is the local range of averaging. Figure 5.3 presents the densities corresponding to the relative displacement patterns of Figures 5.1 and 5.2.

Simulation of piston compression

We first implement a preliminary thermalization. The particles are taken initially at rest at their equilibrium positions. We then generate displacements x_n and velocities $\dot{x}_n$ from the probability density

$$d\nu = \bigotimes_{n=-\infty}^{\infty} Z^{-1} \mathrm{e}^{-\frac{1}{2}\beta_x (x_n^2 + \dot{x}_n^2)} \, dx_n \, d\dot{x}_n, \tag{5.6}$$

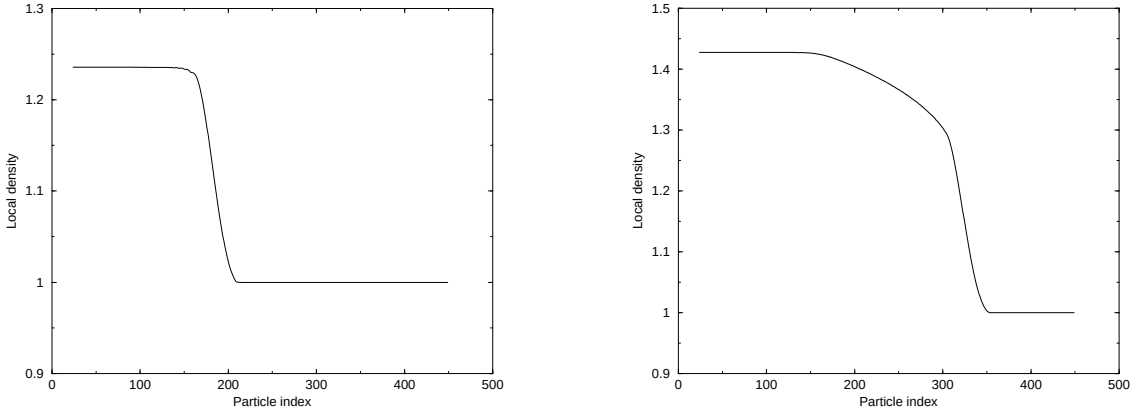


Fig. 5.3. Density patterns for the relative displacement pattern of the weak shock of Figure 5.1 (left) and the strong shock of Figure 5.2 (right). The local averaging range is $M = 50$.

with $Z = 2\pi/\beta_x$. The initial displacements and velocities are then of order $\dfrac{1}{\sqrt{\beta_x}}$. Notice that we take small initial displacements, so we approximate the full potential $V(x)$ by its harmonic part $\frac{1}{2}x^2$. This approximation is of course justified only at the beginning of the simulation, when displacements are small enough. After this initial perturbation, we let the system free to evolve during a typical time $T_{\text{init}} = 10$. The simulations were performed using a Velocity Verlet scheme, the time step being chosen to have a relative energy conservation $\dfrac{\Delta E}{E}$ of about 10^{-3}. At time T_{init} the piston impact begins: the first particle is kept moving toward the right at constant velocity u_p.

Let us emphasize that the shock patterns are robust, in the sense that they remain essentially unchanged when initial thermal pertubations are supplied. This point was already noted in [168] where the authors gave numerical evidence of that fact. While rigorously proven only in the Toda lattice case for a lattice initially at rest at equilibrium, the above shock description seems then to remain qualitatively valid for a quite general class of potentials and with random initial conditions. A comparison of the different profiles is made in Figures 5.4 and 5.5. The profiles are indeed quite conserved, especially the density profiles.

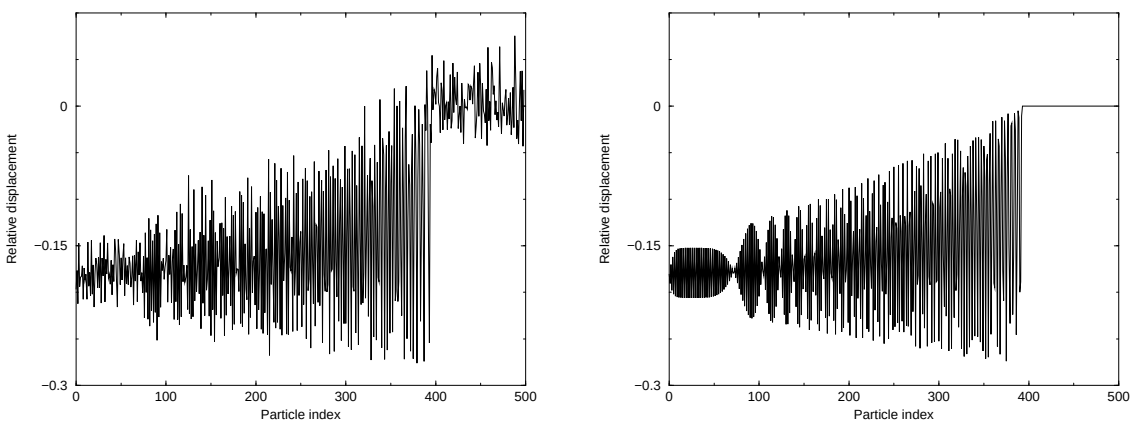


Fig. 5.4. Relative displacement profiles for a thermalized strong shock using a Toda potential with $b = 10$, and comparison with the reference profile corresponding to a lattice initially at rest. The piston speed is $u_p = 0.3$ (so that $a = 3$), $\frac{1}{\sqrt{\beta_x}} = 0.02$.

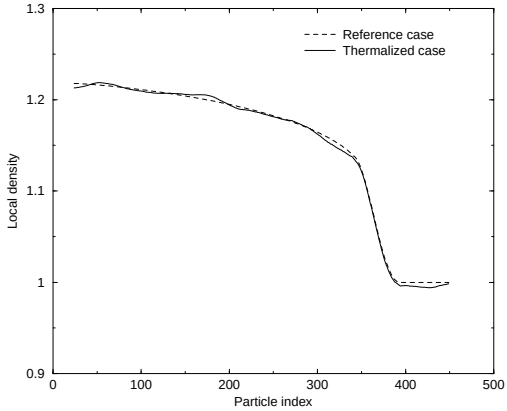


Fig. 5.5. Local density profiles corresponding to Figure 5.4 with $M = 50$. Dashed line: reference profile. Solid line: Thermalized profile. Notice that both patterns almost coincide.

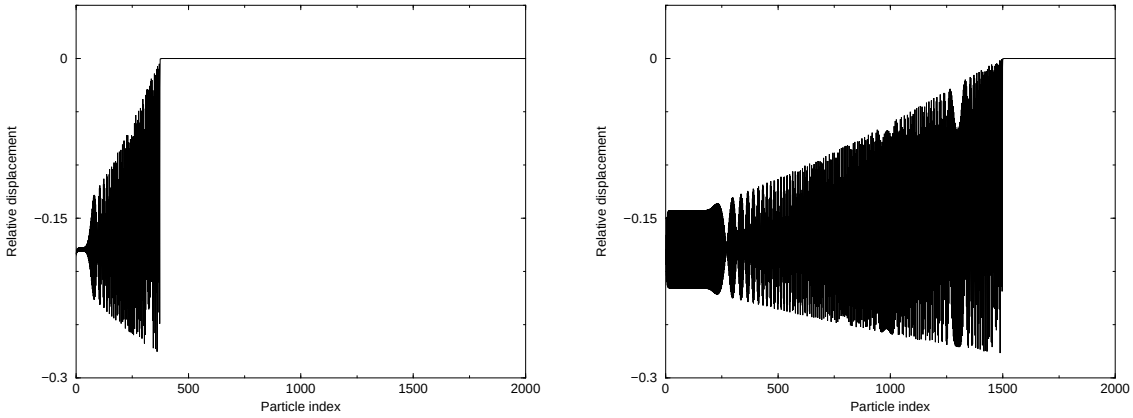


Fig. 5.6. Relative displacement patterns for the same conditions as in Figure 5.4 (reference case). Left: Snapshot at time $T_1 = 200$. The shock front corresponds (roughly) to the zone between particle $N_{\min} = c_{\min}T_1 = 60$ and particle $N_{\max} = c_{\max}T_1 = 350$. Right: Snapshot at time $T_2 = 800$. The shock front corresponds to the zone between particle number $N_{\min} = 250$ and particle number $N_{\max} = 1500$. Thus the shock front is indeed growing linearly in time.

For strong shocks ($a > 2$), the shock front thickens linearly with time as can be seen in Figure 5.6. This is in contradiction with what is observed in shock propagation experiments as well as in 3D numerical simulations. Moreover the velocity distribution behind the shock front shows that the downstream particles experience a (quasi-)oscillatory motion in the range $[0, 2u_p]$. This is of course not the case for 3D simulations, where the particle velocities are much less correlated, and appears to be a pure 1D effect.

We emphasize once again that initial thermal perturbations are not sufficient to remedy these spurious 1D effects since the patterns obtained in Figures 5.4 and 5.5 are very similar. In the sequel we are going to build a 1D model that enables us to get rid of these undesired effects.

Simulation of relaxation waves

In order to study the relaxation waves, the piston is removed after a compression time t_0, and the systems evolves freely during time $t_1 - t_0$.

The results are once again not physically satisfactory. The soliton train of Figure 5.7, which was less visible in Figure 5.4, is not destroyed by the relaxation waves. It travels on and widens since the solitons move away from each others (the distance between the fastest ones, that is,

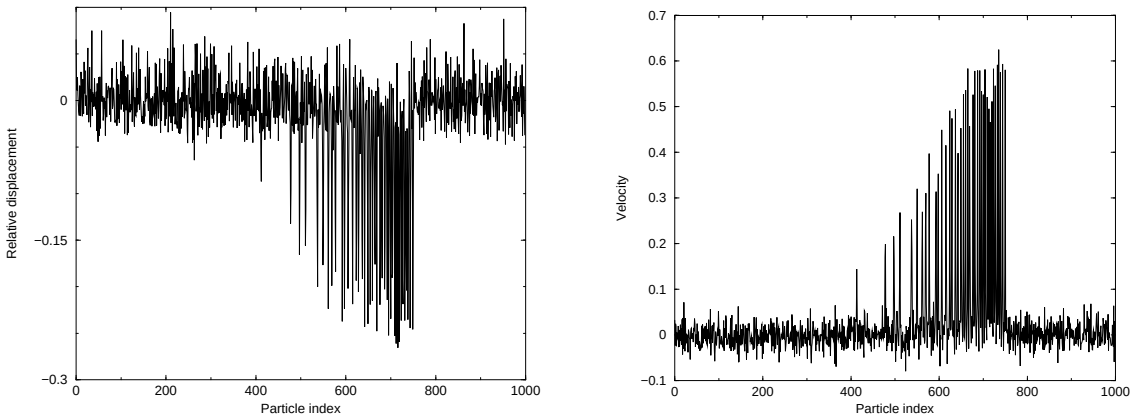


Fig. 5.7. Relative displacement and speed profiles for the same parameters as for Figure 5.4. The compression time is now $t_0 = 50$, and the relaxation time is $t_1 - t_0 = 350$.

the more energetic ones, and the slowest ones, increases). We emphasize that the energy remains localized in those waves, so there is no damping of these solitons. Rarefaction is only observed in the region behind the soliton train.

On the other hand, in 3D simulations or in experiments, one observes a progressive damping of the whole compressive wave. This is a second spurious effect of the 1D model we would like to get rid of and that the model of Sections 5.1.2 and 5.1.3 will be able to deal with.

5.1.2 An augmented one-dimensional model

The results of the previous Section indicate the need for a modeling of perturbations arising from the transverse degrees of freedom existing in higher dimensional simulations. Such perturbations will interfere with the shock front composed of a soliton train, and possibly damp this soliton train. Perturbations in the longitudinal direction, such as thermal initialization for the x_n, cannot do this, as shown by Figures 5.4 and 5.5.

Actually, some facts are already known about the influence of 3D effects on shock waves. In [162,167] Holian *et. al* pointed out the fact that even a 1D shock considered in a 3D system (a piston compression along a principal direction of a crystal for example) may not look like the typical 1D pattern of Figures 5.1 or 5.2. If the crystal is at zero temperature, then the compression pattern in 3D is the same as the 1D one, with a soliton train at the front. But if positive temperature effects are considered, the interactions of the particles with their neighbors - especially in the transverse directions - lead to the destruction of the coherent soliton train at the front, and a steady-regime can be reached (shock with constant thickness).

Therefore, 1D models are often supplemented with a *postulated* dissipation. The corresponding damping term in the equations of motion usually accounts for radiative damping [160,313,314], or may compensate thermal fluctuations [9] from an external heat bath for a system at equilibrium. Let us point out that purely dissipative models may stabilize shock fronts. However, temperature effects then completely disappear. In particular, no jump in kinetic temperature can be observed in purely dissipative 1D simulations. Besides, we also aim here at motivating the usually postulated dissipation and memory terms, and show that they arise naturally as effects of (conveniently chosen) higher dimensional degrees of freedom.

There is no existing model (to our knowledge) that could both account for higher dimensional effects in non equilibrium dynamics and be mathematically tractable. We introduce a classical *deterministic* heat bath model, as an idealized way to couple the longitudinal modes of the atom chain to other modes. This model is justified to some extent by heuristic considerations in Sec-

tion 5.1.2. We are then able to derive a generalized Langevin equation describing the evolution of the system, and recover a stochastic model in some limiting regime.

Form of the perturbations arising from higher dimensional degrees of freedom

Consider the system described in Figure 5.8, which is still a 1D atom chain, but where each particle in the 1D chain also interacts with two particles outside the horizontal line. These particles aim at mimicking some effects of transverse degrees of freedom. The transverse particles are placed in the middle of the springs and have only one degree of freedom, namely their ordinates y_n. The particles in the 1D chain are still assumed to have only one degree of freedom as well. This means that we constrain them to remain on the horizontal line. The interactions between the particles in the chain and the particles outside the chain are ruled by a pairwise interaction potential, for example the same potential as for interactions in the 1D chain.

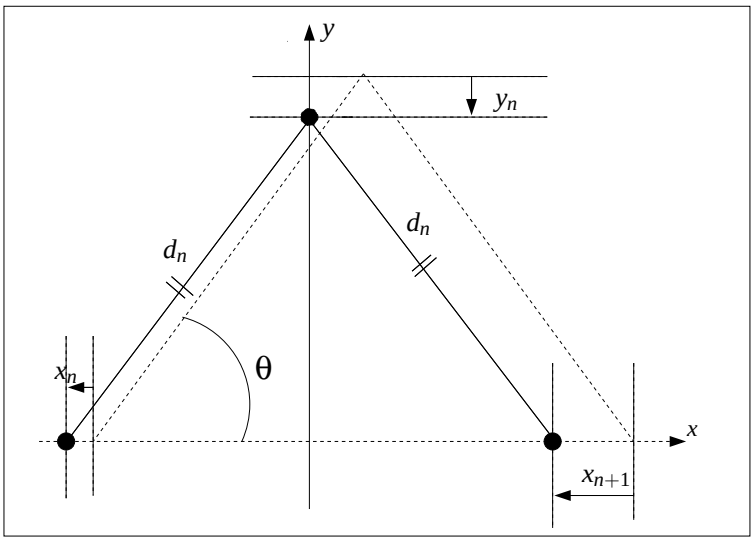


Fig. 5.8. Notations for the interaction of a transverse particle with particles on the 1D atom chain.

Consider small displacements around equilibrium positions. The pairwise interaction potentials can therefore be taken harmonic. Up to a normalization, and for a displacement x from equilibrium position, $V(x) = \frac{1}{2}x^2$.

We first turn to the case $\theta = \frac{\pi}{3}$ corresponding to a 2D regular lattice. At first order,

$$d_n = \left[\left(\frac{1}{2}(1 + x_{n+1} - x_n)\right)^2 + \left(\frac{\sqrt{3}}{2} + y_n\right)^2\right]^{1/2} \simeq 1 + \frac{1}{4}(x_{n+1} - x_n) + \frac{\sqrt{3}}{2}y_n.$$

We now focus on the evolution of x_n. All the equalities written below have to be understood as equalities holding at first order in $\mathrm{O}(|x_n|), \mathrm{O}(|y_n^j|)$. Considering only interactions with the neighboring particles on the horizontal line, and the additional interaction with the particle y_n,

$$\ddot{x}_n = \frac{9}{8}(x_{n+1} - 2x_n + x_{n-1}) + \frac{\sqrt{3}}{4}(y_n - y_{n-1}).$$

The equation governing the evolution of y_n is:

$$\ddot{y}_n = -\frac{3}{2}y_n - \frac{\sqrt{3}}{2}(x_{n+1} - x_n).$$

More generally, consider the system of Figure 5.8 with an arbitrary angle θ. The equilibrium distance is now $d^0 = \frac{1}{2\cos\theta}$, and the corresponding normalized harmonic potential is $V(d) = \frac{1}{2}(\frac{d}{d^0} - 1)^2$. The normalized distance $\bar{d}_n = \dfrac{d_n}{d^0}$ is

$$\bar{d}_n = 1 + \cos^2\theta(x_{n+1} - x_n) + 2\sin\theta\cos\theta \cdot y_n.$$

The additional longitudinal force exerted on x_n by y_n is then

$$f_n = \cos^2\theta\left[\cos\theta(x_{n+1} - x_n) + 2\sin\theta \cdot y_n\right].$$

Summing over N particles that do not interact with each other, each one being characterized by an angle θ_i, the additional force on x_n is seen to be of the form

$$F_n = A_N(x_{n+1} - 2x_n + x_{n-1}) + \sum_{i=1}^{N} K_i(y_n^i - y_{n-1}^i),$$

with $K_i = 2\cos^2\theta_i \sin\theta_i$ and $A_N = \sum_{i=1}^{N}\cos^3\theta_i$. So, the equation of motion for x_n is

$$\ddot{x}_n = (1 + A_N)(x_{n+1} - 2x_n + x_{n-1}) + \sum_{i=1}^{N} K_i(y_n^i - y_{n-1}^i). \tag{5.7}$$

The equations for the y_n^i can be obtained in the same way as before:

$$\ddot{y}_n^i = -a_i y_n^i - 2K_i(x_{n+1} - x_n). \tag{5.8}$$

These linear perturbations are only valid for small displacements, *i.e.* when the approximation of the full potential by its harmonic part is justified. Notice moreover that we discard any type of interaction of the y particles with each others. However, this motivates an attempt to take into account missing degrees of freedom by introducing a heat bath whose form will lead to equation of motion similar to (5.7) - (5.8). We now turn to this task.

Description of the heat bath model

We consider the following Hamiltonian for a coupled system consisting of the system under study (S) and a heat bath (B) described by bath variables $\{y_n^j\}$ ($n \in \mathbb{Z},\ j = 1, \ldots, N$). To use a heat bath is classical but was never done in the context of 1D chains. The full Hamiltonian reads:

$$H(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}) = H_{\mathrm{S}}(\{q_n,\, p_n\}) + H_{\mathrm{SB}}(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}), \tag{5.9}$$

where $(q_n, p_n,\ \tilde{q}_n^j, \tilde{p}_n^j) = (x_n, \dot{x}_n, y_n^j, m_j\dot{y}_n^j)$, H_{S} is given by (5.2), and

$$H_{\mathrm{SB}}(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}) = \sum_{n=-\infty}^{\infty}\sum_{j=1}^{N}\frac{1}{2m_j}(\tilde{p}_n^j)^2 + \frac{1}{2}k_j\left[\gamma_j(q_{n+1} - q_n) + \tilde{q}_n^j\right]^2. \tag{5.10}$$

The interpretation is as follows. Each spring length $\delta x_n = x_{n+1} - x_n$ is thermostated by a heat bath $\{y_n^j\}$, in the spirit of [108,379]. The parameter k_j is the spring constant of the j-th oscillator, m_j its mass, γ_j weights the coupling between Δx_n and y_n^j. Note that although more general cases can be considered [198,212], the coupling is taken bilinear in the variables, for it allows for an exact mathematical treatment. Indeed, a generalized Langevin equation (GLE) can be easily recovered

(see [108, 379] for seminal examples). It is also the only case where the limit $N \to \infty$ can be rigorously justified. Other physical motivations may be presented, such as the representation of extra variables in Fourier modes leading to a Hamiltonian similar to (5.9), see [44]. These extra degrees of freedom allow for some "transverse" radiation of the energy.

Derivation of the generalized Langevin equation

General procedure

Up to a rescaling of y_n^j, we may assume that all masses m_j are 1. The only parameters left for the coupling are the coupling factors γ_j. Introducing the pulsations ω_j given by $\omega_j = k_j^{1/2}$, the equations of motion read:

$$\ddot{x}_n = g_N(x_{n+1} - x_n) - g_N(x_n - x_{n-1}) + \sum_{j=1}^{N} \gamma_j \omega_j^2 (y_n^j - y_{n-1}^j), \tag{5.11}$$

$$\ddot{y}_n^j = -\omega_j^2 \left[y_n^j + \gamma_j (x_{n+1} - x_n) \right], \tag{5.12}$$

where

$$g_N(x) = V'(x) + \left(\sum_{j=1}^{N} \gamma_j^2 \omega_j^2 \right) x. \tag{5.13}$$

Notice the structural similarities of (5.11) with (5.7) and of (5.12) with (5.8).

The solutions $\{y_n^j\}$ of (5.12) are then integrated and inserted in (5.11) for $\{x_n\}$. This procedure is the classical Mori-Zwanzig projection [250, 379]. The integrability of the system is clear (once initial conditions in velocities and displacements are set) when the force g_N is globally Lipschitz. This is for example the case when the sum $\sum_{j=1}^{N} \gamma_j^2 \omega_j^2$ is finite, and when V' is globally Lipschitz, which is indeed true for the Toda potential (5.5). For the Lennard-Jones potential (5.4) it remains true as long as the energy of the system is finite (since the potential diverges when $x \to -1$, the bound on the total energy implies $x > x_0 > -1$, and a bound on the Lipschitz constant can be given by $V'(x_0)$). The computation gives:

$$y_n^j(t) = y_n^j(0) \cos(\omega_j t) + \frac{\dot{y}_n^j(0)}{\omega_j} \sin(\omega_j t) + \int_0^t \gamma_j \omega_j \sin(\omega_j s)(x_{n+1} - x_n)(t - s)\, ds.$$

Integrating by parts and inserting in (5.11):

$$\boxed{\begin{aligned} \ddot{x}_n(t) = {} & V'(x_{n+1} - x_n) - V'(x_n - x_{n-1}) \\ & + \int_0^t K_N(s)(\dot{x}_{n+1} - 2\dot{x}_n + \dot{x}_{n-1})(t - s)\, ds + r_n^N(t), \end{aligned}} \tag{5.14}$$

where

$$K_N(t) = \sum_{j=1}^{N} \gamma_j^2 \omega_j^2 \cos(\omega_j t),$$

and

$$\begin{aligned} r_n^N(t) = {} & \sum_{j=1}^{N} (y_n^j(0) - y_{n-1}^j(0)) \gamma_j \omega_j^2 \cos(\omega_j t) + (\dot{y}_n^j(0) - \dot{y}_{n-1}^j(0)) \gamma_j \omega_j^2 \frac{\sin(\omega_j t)}{\omega_j} \\ & + \gamma_j^2 k_j \cos(\omega_j t)(x_{n+1} - 2x_n + x_{n-1})(0). \end{aligned}$$

Formally, (5.14) looks like a generalized Langevin equation (GLE), provided r_n^N is a random forcing term. The dissipation term involves a memory kernel K_N and an "inner" friction $\dot{x}_{n+1} - 2\dot{x}_n + \dot{x}_{n-1}$.

The derivation made here shows that the usually postulated dissipation and memory arise naturally as effects of higher dimensional degrees of freedom. The dissipation term, classical in elasticity theory and postulated by some studies [160,314], is derived here, as memory effects, that were also considered in [314], since the corresponding model was that of a viscoelastic material. So, we are left with a description of the system only in terms of $\{x_n\}$. To further specify the terms, we have to describe the choice of the heat bath spectrum $\{\omega_j\}$, the coupling constant γ_j and the initial conditions for the bath variables.

Choice of the constants

We choose the values [199]:

$$\omega_j = \Omega\left(\frac{j}{N}\right)^k, \quad \gamma_j^2\omega_j^2 = \lambda^2 f^2(\omega_j)\ (\Delta\omega)_j, \quad f^2(\omega) = \frac{2\alpha}{\pi}\frac{1}{\alpha^2+\omega^2}, \tag{5.15}$$

where $(\Delta\omega)_j = \omega_{j+1} - \omega_j$, $\alpha, \lambda > 0$ and $k > 0$.

The function f^2 is defined this way for reasons that will be made clear in Section 5.1.3. The heat bath spectrum $\{\omega_j\}$ is more dense as N increases. The exponent k accounts for the repartition of the pulsations. More general choices could be made, involving randomly chosen pulsations [199]. However, we restrict ourselves to the case of deterministic pulsations. We emphasize here once again that the constants chosen and the form of the coupling are not new. A similar choice is made in [199]. The novelty is in the application to a 1D chain, where independent heat baths are considered, each heat bath corresponding to a spring length.

We now motivate (5.15). Notice that an upper bound to the heat bath spectrum is imposed. This is related to the discreteness of the medium. Indeed, for a system at rest with particles distant from 1, the higher pulsation allowed is π, corresponding to an oscillatory motion of spatial period 2. When particles come closer (for example if the mean distance between particles is $a < 1$), the higher pulsation increases to the value $\frac{\pi}{a}$ since the lowest spatial period is now $2a$. Taking then lower bound d_m for the minimal distance between neighboring particles, we get an upper bound for the spectrum, namely $\Omega = \frac{\pi}{d_m}$.

The choice of the coupling constants between the system and the bath is an important issue. The only purpose of the heat bath in a 1D shock simulation is to mimic some effects of dimensionality, such as energy transfer to the tranverse modes. This energy transfer can be quantified using (5.12). Indeed, the total energy transfer for a harmonic oscillator of pulsation ω subjected to an external forcing σ is known [44]. More precisely, consider the following harmonic oscillator:

$$\ddot{z} + \omega^2 z = h(t), \tag{5.16}$$

where h is an external time-dependent forcing term. Then the total energy transfered by the external forcing to the system (from $t = -\infty$ to $t = +\infty$ for a system at rest at $t = -\infty$) is $\Delta E = \frac{1}{2}|\hat{h}(\omega)|^2$. The energy transfer to the heat bath occurs as described by (5.12). This gives a total energy transfer for a spring $x_{n+1} - x_n$ considered initially at rest:

$$\Delta E_n = \frac{1}{2}\sum_{j=1}^{N}\gamma_j^2\omega_j^4|\widehat{\Delta x_n}(\omega_j)|^2. \tag{5.17}$$

As a first approximation, a shock profile can be described as a self-similar jump: $\Delta x_n(t) = \delta H(n - ctn)$, where $\delta < 0$ is the jump amplitude, c the shock speed, and H is the Heaviside function. Then, $|\widehat{\Delta x_n}(\omega)| = \omega^{-1}$. The energy transfer (5.17) is therefore

$$\Delta E_n = \frac{\delta^2}{2}\sum_{j=1}^{N}\gamma_j^2\omega_j^2.$$

With the spectrum (5.15), the condition $\Delta E_n \to C$ with $0 < C < \infty$ is satisfied:

$$\Delta E_n = \frac{\delta^2\lambda^2}{2}\sum_{j=1}^{N} f^2(\omega_j)(\Delta\omega)_j \to \frac{\delta^2\lambda^2}{2}\int_0^{\Omega} f^2 = \lambda^2\delta^2\sigma(\Omega).$$

The last expression is bounded since f^2 is integrable (recall $\int_0^{\infty} f^2 = 1$). The function σ is a C^∞ function. Notice that the above convergence results from the convergence of the Riemann sum appearing on the left.

Choice of the initial conditions.

We consider initial conditions $\{y_n^j(0),\ \dot{y}_n^j(0)\}$ randomly drawn from a Gibbs distribution with inverse temperature β_y. This distribution is conditioned by the initial data $\{x_n,\ \dot{x}_n\}$. More precisely, set

$$y_n^j(0) = -\gamma_j(x_{n+1} - x_n)(0) + (\beta_y k_j)^{-1/2}\xi_j^n, \tag{5.18}$$

$$\dot{y}_n^j(0) = (\beta_y)^{-1/2}\eta_j^n, \tag{5.19}$$

where $\xi_j^n, \eta_j^n \sim \mathcal{N}(0,1)$ are independently and identically distributed (i.i.d.) random Gaussian variables. With these choices,

$$r_n^N(t) = \frac{1}{\sqrt{\beta_y}}\sum_{j=1}^{N} \omega_j\gamma_j\cos(\omega_j t)(\xi_n^j - \xi_{n-1}^j) + \omega_j\gamma_j\sin(\omega_j t)(\eta_n^j - \eta_{n-1}^j). \tag{5.20}$$

The probability space is induced by the mutually independent sequences of i.i.d. random variables $\xi_n^j,\ \eta_n^j$. Denote D the linear operator acting on sequences $Z = \{z_n\}$ through $DZ = \{z_n - z_{n-1}\}$. So,

$$r_n^N(t) = \frac{\lambda}{\sqrt{\beta_y}}\sum_{j=1}^{N} f(\omega_j)\cos(\omega_j t)D\xi_n^j + f(\omega_j)\sin(\omega_j t)D\eta_n^j\ (\Delta\omega)_j^{1/2}.$$

For fixed N, the above expressions give

$$\mathbb{E}(r^N(t)(r^N(s))^T) = \frac{1}{\beta_y}K_N(t-s)DD^T \tag{5.21}$$

where $r^N = (\dots, r_n^N, \dots)$ and the linear operator DD^T acts on sequences Z as $DD^T z = \{z_{n+1} - 2z_n + z_{n-1}\}$. This relation is known as the fluctuation-dissipation relation, linking the random forcing term and the memory kernel. Notice that the noise term is correlated both in time and in space. The behavior of the system when $N \to \infty$ is then an interesting issue, that can help us to get a better understanding of the phenomenas at play (see Section 5.1.3).

Numerical results

The equations of motion (5.11), (5.12) are integrated numerically for a given N, using a classical velocity-Verlet scheme. The system is initialized with velocities and displacements generated from (5.18) and (5.19) in the y-coordinates, and from (5.6) in the x coordinates. Note that the quantities $\frac{1}{\beta_x}$ and $\frac{1}{\beta_y}$ may differ. The system is then first let to evolve freely, so that the coupling between transverse and longitudinal directions starts.

Shock waves are generated using a piston in the same fashion as in Section 5.1.1, giving Figures 5.9 and 5.10. We then study relaxation waves (Figure 5.11). The time-step Δt is chosen to ensure a relative energy conservation of 10^{-3} in the absence of external forcing. Typically,

$\Delta t = 0.01$. The spectrum density parameter k in (5.15) is taken to be $k = 1$. Other choices lead to the same kind of simulation results. Notice that, if L represents the size of the 1D chain, the algorithmic complexity scales as $\mathrm{O}(LN)$.

Sustained shock waves

Figures 5.9 and 5.10 show the different patterns obtained in the case of a system coupled to a heat bath. Notice that the upper bound to the spectrum, Ω, is of order π since the shock is not too strong, and hence the medium is not too compressed. The parameter α is taken less or equal to Ω so that K_Ω and $\sigma(\Omega)$ are sufficiently close from their limiting values.

The parameter λ was varied in the range $[0, 5]$. If λ is too small, the coupling is too weak and the profiles look like the pure 1D ones (Note that we recover the purely 1D model with Hamiltonian (5.2) when $\lambda = 0$). If λ is too high, the forcing may be too strong, leading to the collapse of two neighboring particles if the time step is not small enough. A good choice of λ involves a good rate of energy transfer to the transverse modes. The choice of λ is completely empirical, but it would be desirable to estimate it from full 3D simulations.

The results show that the introduction of transverse degrees of freedom has important consequences on the pure 1D pattern. The soliton train at the front is destroyed, and the shock thickness is constant along time, instead of growing in time as in the pure 1D case. Thus a steady regime can now be reached, and these simulations really seem to deserve the name "shock waves". In contrast to the pure 1D model results, these simulations have now the same qualitative behavior as 3D simulations or experiments.

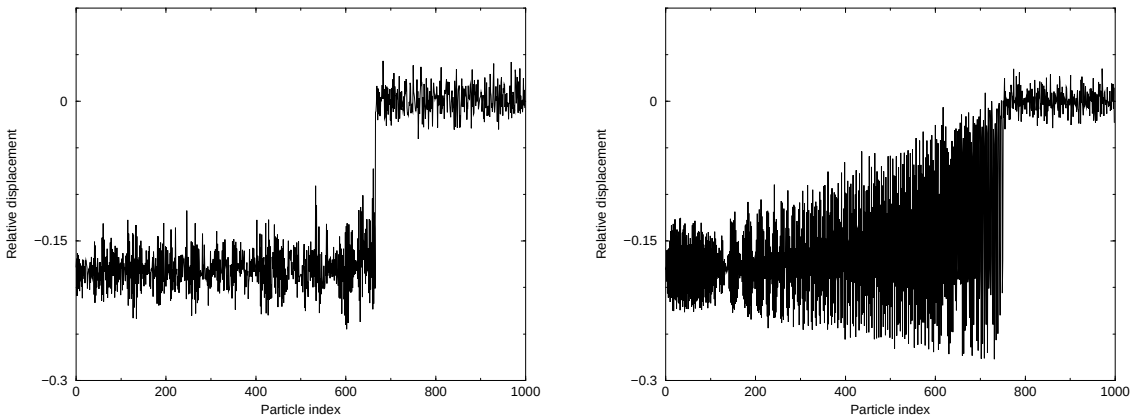


Fig. 5.9. Relative displacement profiles for the system coupled to a heat bath (left), and comparison with a thermalized shock (right). For the thermalized shock, the parameters are $u_p = 0.3$, $b = 10$ and $\frac{1}{\sqrt{\beta_x}} = 0.01$. For the system coupled to a heat bath, the additional parameters are $\frac{1}{\sqrt{\beta_y}} = 0.02$, $\alpha = 5$, $\Omega = 10$, $\lambda = 0.5$. The number of transverse oscillators is $N = 25$.

Rarefaction waves

As can be seen in Figure 5.11, a rarefaction wave develops and progressively weakens the shock (notice that the velocities decrease and that the relative displacement increase compared to Figures 5.9 and 5.10). This is indeed the expected physical behavior for a viscous fluid. This dissipation can be interpreted as energy transfer to the transverse modes.

Besides, no soliton train survives, contrarily to the pure 1D case, where the solitons are not destroyed and move on unperturbed. In the pure 1D case, there is no weakening of the initial wave, only dispersion. Once again, to our knowledge, this is the first time a 1D discrete model behaves as expected.

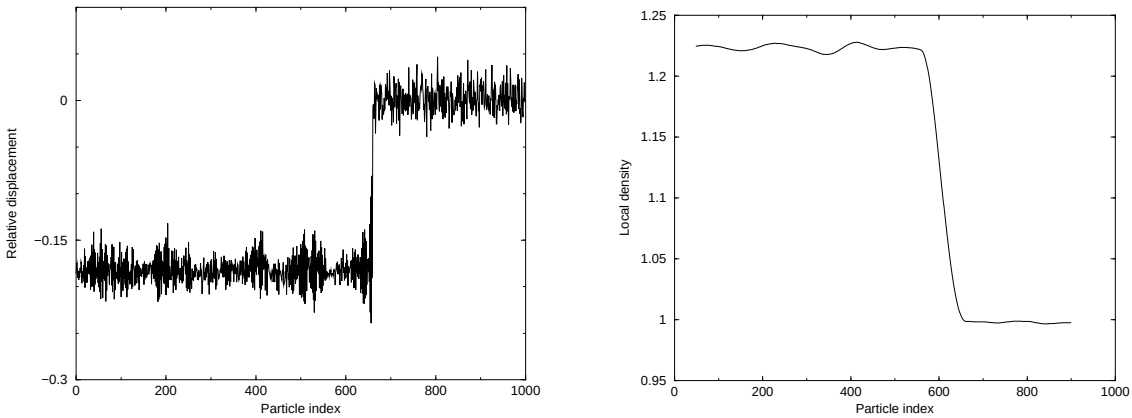


Fig. 5.10. Same parameters as for Figure 5.9, except for the system coupled to a heat bath, $N = 100$. Left: Relative displacement profile. Right: Local density as a function of the particle index.

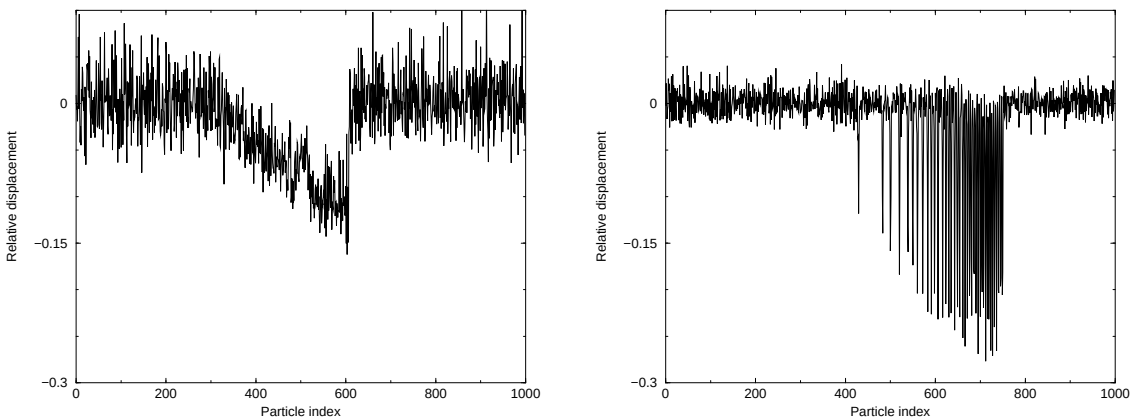


Fig. 5.11. Relative displacement profiles for the system coupled to a heat bath (left) and the thermalized 1D system (right). The parameters for the system coupled to a heat bath are $\frac{1}{\sqrt{\beta_y}} = 0.04$, $\alpha = 2$, $\Omega = 5$, $\lambda = 0.5$. The system is compressed during $t_0 = 50$. The relaxation time is $t_1 - t_0 = 350$.

Generalizations of the system-bath interaction

Beyond nearest-neighbor interactions

The Hamiltonian of the system can be written in an abstract form as

$$H(x,\ y_N) = \frac{1}{2}|\dot{x}|^2 + F(x) + \frac{1}{2}\dot{y_N}^T M \dot{y_N} + \frac{1}{2}|\mathcal{A}x - \mathcal{B}y_N|^2 \tag{5.22}$$

where $x = (\ldots, x_{n-1}, x_n, x_{n+1}, \ldots)$ and $y_N = (\ldots, y^1_{n-1}, \ldots, y^N_{n-1}, y^1_n, \ldots, y^N_n, \ldots)$. The matrix M is a mass matrix (operator), $\mathcal{A}$ and $\mathcal{B}$ are general operators, $F(x) = \sum_{n=-\infty}^{\infty} V(x_{n+1} - x_n)$. We chose previously $\mathcal{B}$ diagonal. But more generally, $\mathcal{B}$ could be considered as tridiagonal: this could model the interaction of two neighboring heat baths linked to neighboring spring lengths.

Nonlinear coupling with the heat-bath

When the shock strength increases, the heuristic derivation performed in this section (relying on small displacements) is no longer valid. The approach can however be generalized by considering a nonlinear coupling between the transverse particles and the particles in the chain. It is hoped that the thermalization will be more efficient this way, in particular, stronger shocks could be

sustained with less transverse oscillatory degrees of freedom. We therefore consider the following Hamiltonian:

$$H(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}) = H_{\mathrm{S}}(\{q_n,\, p_n\}) + H_{\mathrm{NLB}}(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}), \tag{5.23}$$

with $(q_n, p_n, \tilde{q}_n^j, \tilde{p}_n^j) = (x_n, \dot{x}_n, y_n^j, \dot{y}_n^j)$, H_{S} still given by (5.2), and

$$H_{\mathrm{NLB}}(\{q_n,\, p_n,\, \tilde{q}_n^j,\, \tilde{p}_n^j\}) = \sum_{n=-\infty}^{\infty} \sum_{j=1}^{N} \frac{1}{2}(\tilde{p}_n^j)^2 + k_j U[\gamma_j(q_{n+1} - q_n) + \tilde{q}_n^j], \tag{5.24}$$

where U is a nonlinear function to be specified. The Hamiltonian (5.9) is recovered when $U(x) = \frac{1}{2}x^2$. Typically,

$$U(x) = V_{\mathrm{LJ}}(1 + x),$$

so that the interactions with the transverse oscillators are similar than the interactions in the chain. We still consider the distribution of stiffnesses k_j and coupling constants γ_j given by (5.15). Figure 5.12 presents numerical results obtained for a strong shock ($u_p = 1$). Satisfactory shock profiles are obtained with $N = 8$ additional degrees of freedom only.

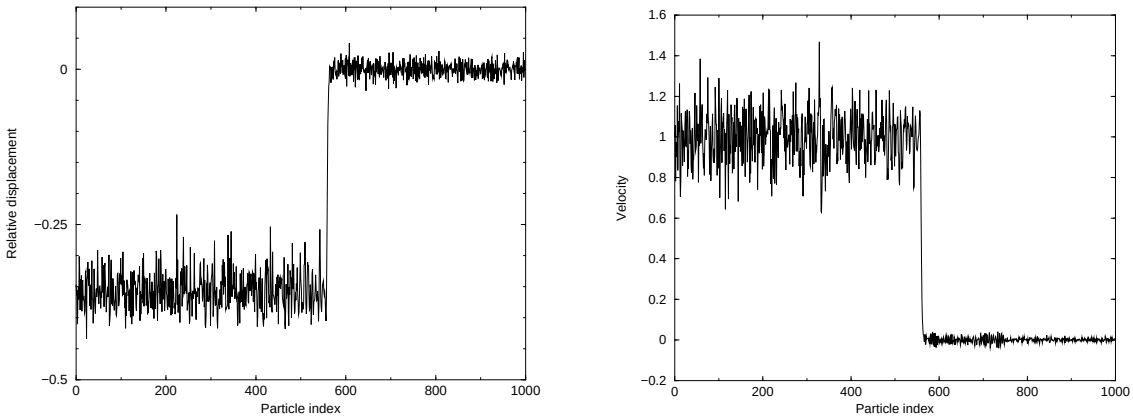


Fig. 5.12. Displacement profiles (Left) and velocity profiles (Right) for a strong shock ($u_p = 1$) for the deterministic model (5.9) using a nonlinear coupling, with $N = 8$, the parameters of the spectrum (5.15) being $k = 1$, $\Omega = 10$, $\alpha = 5$ and $\lambda = 0.2$.

5.1.3 The stochastic limit

The model developed in the previous section shows how the introduction of a certain number of transverse degrees of freedom leads to compression profiles very different from the purely one-dimensional results. In particular, some energy relaxation is possible due to the heat bath formed by the transverse oscillators. However, even when the heat bath is nonlinearly coupled, several degrees of freedom have to be introduced and numerically resolved for each longitudinal degree of freedom. Therefore, it is interesting to replace the deterministic heat bath with many oscillators by its average action. Mathematically, this amounts to replacing the deterministic system (5.14) by a stochastic differential equation (SDE) of lower dimension. The only remaining unknowns are the positions of the particles $(\ldots, x_n(t), \ldots)$.

Limit of the dynamics (5.14) when $N \to \infty$

Limit of the dissipation term

The memory kernel can be seen as a Riemann sum. The limit is then:

$$K_N(t) = \lambda^2 \sum_{j=1}^{N} f^2(\omega_j)\cos(\omega_j t)(\Delta\omega)_j \to \lambda^2 \int_0^{\Omega} f^2(\omega)\cos(\omega t)\ dt = \lambda^2 K_\Omega(t) \tag{5.25}$$

when $N \to \infty$, the convergence holding in $L^1[0,T]$, $T > 0$.

The special choice (5.15) implies $K_\Omega(t) \to \mathrm{e}^{-\alpha t}$ when $\Omega \to \infty$ in $L^\infty(\mathbb{R}_+)$. The memory kernel is then exponentially decreasing.

Limit of the fluctuation term

The limit $N \to \infty$ gives the convergence of the noise term in a weak sense in $C[0,T]$ toward a stochastic integral:

$$r_n^N(t) \to \lambda r_n^\Omega(t) = \frac{\lambda}{\sqrt{\beta_y}} \int_0^{\Omega} f(\omega)\cos(\omega t) D\, dW_\omega^{n,1} + f(\omega)\sin(\omega t) D\, dW_\omega^{n,2} \tag{5.26}$$

where $W_\omega^{n,1}, W_\omega^{n,2}$ $(n \in \mathbb{Z})$ are independent standard Brownian motions.

Limit of the equation

Formally, a stochastic integro-differential equation (SIDE) is obtained in the limit $N \to \infty$:

$$\boxed{\begin{aligned} \ddot{x}_n(t) &= V'(x_{n+1} - x_n) - V'(x_n - x_{n-1}) \\ &\quad + \lambda^2 \int_0^t K_\Omega(s)(\dot{x}_{n+1} - 2\dot{x}_n + \dot{x}_{n-1})(t-s)\ ds + \lambda r_n^\Omega(t), \end{aligned}} \tag{5.27}$$

with

$$K_\Omega(t) = \int_0^{\Omega} f^2(\omega)\cos(\omega t)\, d\omega,$$

$$r_n^\Omega(t) = \frac{1}{\sqrt{\beta_y}} \int_0^{\Omega} f(\omega)\cos(\omega t)\, D\, dW_\omega^{n,1} + f(\omega)\sin(\omega t)\, D\, dW_\omega^{n,2},$$

and the fluctuation-dissipation relation

$$\mathbb{E}(r^\Omega(t)(r^\Omega(s))^T) = \frac{1}{\beta_y} K_\Omega(t-s) D D^T, \tag{5.28}$$

where $r^\Omega = (\dots, r_n^\Omega, \dots)$. The way the solutions of (5.14) converge to the solutions of (5.27) can be made rigorous by a direct adaptation of the results of [199]: the convergence of x_n^N solution of (5.14) to x_n solution of (5.27) is weak in $C^2[0,T]$ (in the sense of continuous random processes, see below).

The SIDE (5.27) can be rewritten as a stochastic differential equation (SDE). In the limiting case $\Omega \to \infty$, a Markovian limit can indeed be recovered when considering an additional variable [199]. Notice that when $\Omega \to \infty$, $K_\Omega(t) \to K(t) = \mathrm{e}^{-\alpha t}$. Denoting $Q = (\dots, x_{n-1}, x_n, x_{n+1}, \dots)$, $P = (\dots, \dot{x}_{n-1}, \dot{x}_n, \dot{x}_{n+1}, \dots)$, $V(Q) = \sum_{n=-\infty}^{\infty} V(x_{n+1} - x_n)$ and $R = (\dots, R_{n-1}, R_n, R_{n+1}, \dots)$, $\lambda = \sqrt{\alpha\xi}$, the previous SIDE (5.27) is equivalent to the following SDE:

$$\begin{aligned} dQ_t &= P_t\,dt, \\ dP_t &= (R_t - \nabla V(Q_t))\,dt, \\ dR_t &= -\alpha(R_t + \xi DD^T P_t)\;dt + \alpha\sqrt{2\beta^{-1}\xi}\,DdW_t, \end{aligned} \tag{5.29}$$

where W is a standard Brownian motion, and with initial conditions $r_n(0) \sim \lambda\beta^{-1/2}\mathcal{N}(0,1)$.

The limiting equation (5.26) shows the main effects of the heat-bath interaction: The pure 1D equation (5.3) is supplemented by two terms, one dissipation term with an exponentially decreasing memory, and a random forcing. Therefore the heat bath acts first as an energy trap, absorbing some of the energy of the shock when it passes. This energy is then given back to the system through the random forcing term to an amount precised by (5.28). This allows the equilibration of the downstream domain.

Proof of convergence

The proof of the convergence of the solutions of (5.14) to the solutions of (5.27) can be done as in [199], by a straightforward extension to the multi-dimensional case (in order to deal with convergence of sequences). Denote by x_n^N the solution of (5.14) for a given number N of transverse variables. We set $\delta x_n^N = x_{n+1}^N - x_n^N$. The solution of (5.27) is noted x_n. We set $\lambda = 1$ to simplify notations. The extension to more general values of λ is straightforward. The space of real sequences in noted $\mathcal{H} = \mathbb{R}^{\mathbb{N}}$, and is equiped with the usual l^∞-norm. For a sequence $z = \{z_n\} \in \mathcal{H}$:

$$|z|_{l^\infty} = \sup_{n\in\mathbb{Z}} |z_n|.$$

The space $\mathcal{H}$ endowed with this norm is then a separable complete metric space.

Consider the array of spring lenghts

$$Q_N = \begin{pmatrix} \vdots \\ \delta x_n^N \\ \vdots \end{pmatrix},$$

and the array of random forcing terms

$$G_N = \frac{1}{\beta_y}\begin{pmatrix} \vdots \\ r_n^N \\ \vdots \end{pmatrix}.$$

We similarly define Q and G for the sequence $\{x_n\}$.

Recall that the linear operator D, acting on sequences $z = \{z_n\} \in \mathcal{H}$, is defined by $Dz = \{Dz_n\} = \{z_n - z_{n-1}\}$. It follows $|DD^T z|_{l^\infty} \le 4|z|_{l^\infty}$. Equation (5.14) can be rewritten as (recall $\lambda = 1$)

$$\ddot{Q}_N = DD^T F(Q_N) + \int_0^t K_N(s) DD^T \dot{Q}_N(t-s)\;ds + DG_N(t).$$

Introducing $\mathcal{K}_N(t) = \int_0^t K_N(s)\;ds$ and integrating the convolution term by parts, (5.14) becomes

$$\ddot{Q}_N - \left(DD^T F(Q_N) + \int_0^t \mathcal{K}_N(s) DD^T \ddot{Q}_N(t-s)\;ds\right) = DG_N(t) - DD^T \dot{Q}_N(0)\mathcal{K}_N(t). \tag{5.30}$$

This equation can be rewritten under a fixed point form as

$$(\mathrm{Id} + R_N)\ddot{Q}_N(t) = h_N(t). \tag{5.31}$$

As F is Lipschitz, $||R_N||$ is small for small T. An usual Picard argument gives the existence and uniqueness of $\ddot{Q}_N \in C([0,T],\mathcal{H})$ solving (5.31) for T small enough (see [148], Section 12, for an analogous proof). Standard results also give the continuity of $\ddot{Q}_N$ on $\mathcal{K}_N \in L^1[0,T]$ and $U_N = DG_N - DD^T Q_N(0)\mathcal{K}_N \in C([0,T],\mathcal{H})$. The mapping $(K_N, U_N) \mapsto Q_N$ is then continuous from $L^1[0,T] \times C([0,T],\mathcal{H})$ to $C([0,T],\mathcal{H})$ with the corresponding norms.

The convergence of K_N in $L^1[0,T]$ is straightforward, and implies the convergence of $\mathcal{K}_N$ in $L^1[0,T]$. The convergence of U_N results from the convergence of $\mathcal{K}_N \in L^1[0,T]$ and from the convergence of G_N to G (in a way to precise). We refer to [125], Section VI.4., Theorem 2. Considering the collection of continuous real-valued stochastic processes G_N with values in $\mathcal{H}$ (which is a separable complete metric space), we have to show:

(i) The finite-dimensional distributions of G_N weakly converge to those of G, which is a continuous process.

(ii) A tightness inequality of the form

$$\forall t, t+u \in [0,T], \quad \mathbb{E}\left[|G_N(t+u) - G_N(t)|^2_{l^\infty}\right] \leq C|u|.$$

Then it follows $G_N \Rightarrow G$ in $C([0,T],\mathcal{H})$-weak.

These two points are straightforward generalizations of the proof in [199] (in the case of non-random pulsations ω_j) when extended to sequences with values in $\mathcal{H}$, giving the convergence $U_N \Rightarrow U$ in $C([0,T],\mathcal{H})$-weak. The convergences of K_N to K in $L^1[0,T]$ and U_N to U in $C([0,T],\mathcal{H})$ in a weak sense then give the convergence of $\ddot{Q}_N$ in $C([0,T],\mathcal{H})$ in a weak sense. Therefore, $Q_N \Rightarrow Q$ in $C^2([0,T],\mathcal{H})$-weak. This implies the convergence in a weak sense for all the components of Q_N for T small enough.

For general t, consider $\mathrm{e}^{-\gamma t} Q_N$ for γ large enough, and rescale appopriately the operators appearing in (5.31). The proof then follows the same lines.

Numerical implementation

The SDE (5.29) is of the form

$$dX_t = Y(X_t)\,dt + \Sigma dW_t, \tag{5.32}$$

where W_t is a standard Wiener process, with the notations

$$X_t = (Q_t, P_t, R_t), \quad Y(X_t) = (P_t, R_t - \nabla V(Q_t), -\alpha R_t + \alpha\xi DD^T P_t), \quad \Sigma = \alpha\sqrt{\frac{2\xi}{\beta}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix}.$$

The integration is done using the following splitting of the vector field Y:

$$Y(X) = Y_{\text{Newton}}(X) + Y_{PR}(X) + Y_{RR}(X) + Y_{RP}(X),$$

with $Y_P(X) = (0, R, 0)$, $Y_R(X) = (0, 0, -\alpha R + \alpha\xi DD^T P)$ and $Y_{\text{Newton}}(X) = (P, -\nabla V(Q), 0)$. Denote also by $\phi^{\Delta t}_{\text{Newton}}$, $\phi^{\Delta t}_P$ and $\phi^{\Delta t}_R$ the associated numerical flows. When $\Sigma = 0$, a consistant numerical scheme is

$$\Psi^{\Delta t} = \Phi^{\Delta t/2}_R \circ \Phi^{\Delta t/2}_P \circ \Phi^{\Delta t}_{\text{Newton}} \circ \Phi^{\Delta t/2}_P \circ \Phi^{\Delta t/2}_R.$$

The flow $\phi^{\Delta t}_{\text{Newton}}$ is approximated by the Velocity-Verlet scheme $\Phi^{\Delta t}_{\text{Newton}}$. The flows $\phi^{\Delta t}_P$ and $\phi^{\Delta t}_R$ can be analytically integrated, so that:

$$\Phi^{\Delta t}_P(Q_0, P_0, R_0) = (Q_0, P_0 + R_0\Delta t, R_0).$$

$$\Phi^{\Delta t}_R(Q_0, P_0, R_0) = \left(Q_0, P_0, \mathrm{e}^{-\alpha\Delta t} R_0 - \xi(1 - \mathrm{e}^{-\alpha\Delta t}) DD^T P_0\right).$$

The random noise is added at the beginning and at the end of the time step. Denoting by i the index of the particlesn and by n the integration index, the following scheme can be proposed:

$$\begin{cases} r_i^{n+1/2} &= \mathrm{e}^{-\alpha\Delta t/2} r_i^n - \xi(1-\mathrm{e}^{-\alpha\Delta t/2})(DD^T p^n)_i + \sqrt{\dfrac{\alpha\xi(1-\mathrm{e}^{-\alpha\Delta t})}{\beta}}(DZ^n)_i, \\ p_i^{n+1/2} &= p_i^n - \dfrac{\Delta t}{2}\nabla V(Q^n) + \dfrac{\Delta t}{2} r_i^{n+1/2}, \\ q_i^{n+1} &= q_i^n + \Delta t p_i^{n+1/2}, \\ p_i^{n+1} &= p_i^{n+1/2} - \dfrac{\Delta t}{2}\nabla V(Q^{n+1}) + \dfrac{\Delta t}{2} r_i^{n+1/2}, \\ r_i^{n+1} &= \mathrm{e}^{-\alpha\Delta t/2} r_i^{n+1/2} - \xi(1-\mathrm{e}^{-\alpha\Delta t/2})(DD^T p^{n+1})_i + \sqrt{\dfrac{\alpha\xi(1-\mathrm{e}^{-\alpha\Delta t})}{\beta}}(DZ^{n+1})_i, \end{cases} \tag{5.33}$$

where $\{Z^n\}_{n\in\mathbb{N}} = \{(\dots, z_i^n, \dots)\}_{n\in\mathbb{N}}$ and $(z_i^n)_{n\in\mathbb{N}, i\in\mathbb{Z}}$ are i.i.d. standard random gaussian variables.

Numerical results

Profiles obtained with a compression at fixed piston velocity u_p for one realization of (5.29) are presented in Figure 5.13, as well as averages obtained over 100 realizations (see Figure 5.14). Although the profiles show sharp transitions, the temperature (given by fluctuations in velocities or positions downtream the shock front) is not correct since it is the same as before the shock. This is contrast with simulation results obtained with a few transverse oscillatory degrees of freedom. We will see in Section 5.2 how to maintain changes in the temperature across the shock interface, as observed in all-atom simulations.

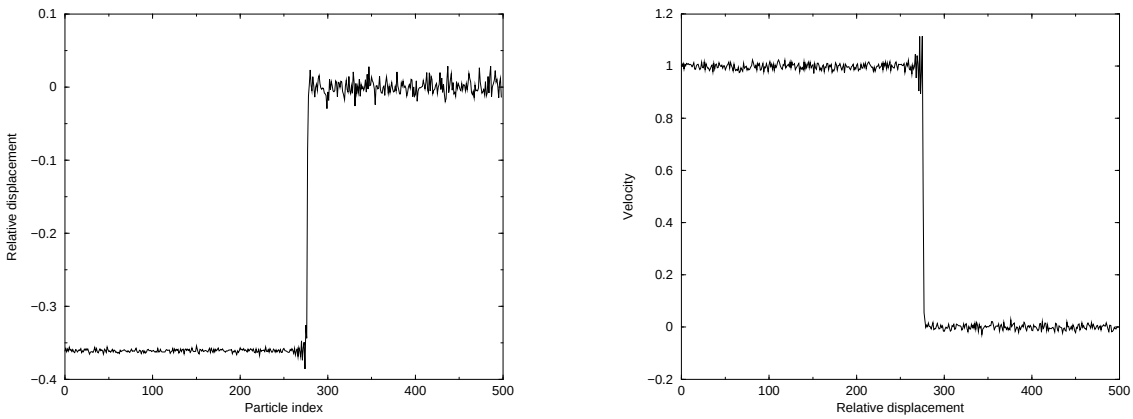


Fig. 5.13. Displacement profiles (Left) and velocity profiles (Right) for a single realization of a sustained shock compression at $u_p = 1$ for (5.29), the parameters being $\alpha = 10$, $\beta^{-1/2} = 0.01$ and $\xi = 1$.

5.1.4 Extension to the reactive case

We extend here the one-dimensional stochastic model for shock waves to the reactive shock waves, where chemical reactions are triggered when the shock passes. The exothermicity of these reactions first enhances, then sustains the propagation of the shock. The physical theorey behind these reactive waves is the ZND theory [103, 343] of detonation waves, which decomposes the wave into three regions: an upstream unperturbed region, a shock front (or reaction zone) of constant

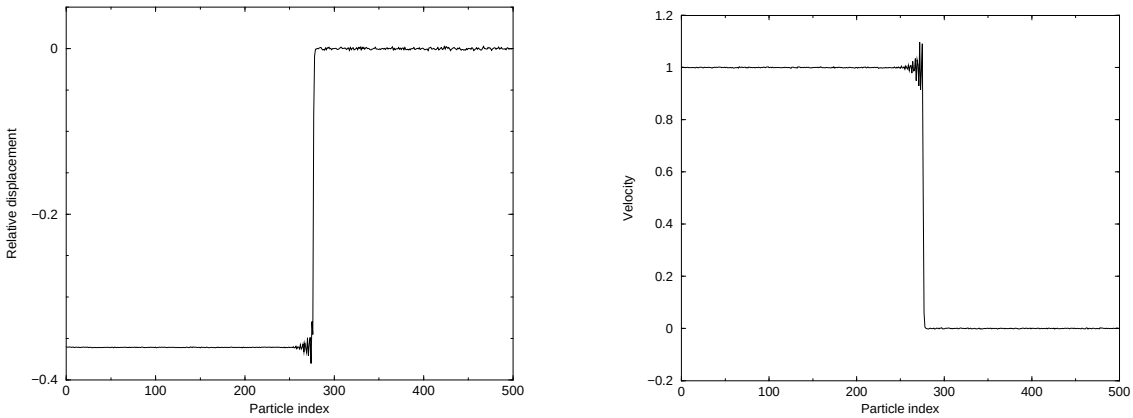


Fig. 5.14. Average over 100 realizations with the same conditions as for Figure 5.13.

width where chemical reactions happen, followed by an autosimilar rarefaction wave. To give some orders of magintude for real materials, the width of the reaction zone ranges between several micrometers to several millimeters, and the speed of the shock front may reach several km/s.

Modelling of reaction waves

We consider a reactive potential in the vein of [361]. To this end, an additional parameter r_n is introduced for each interatomic bond $\Delta x_n = x_{n+1} - x_n$, and models the reaction rate of the zone between x_{n+1} and x_n. The interaction potential is also a function of this additional variable, and since the reaction is exothermic, the ground state of the reaction products is lower than the ground state of the reactants. We therefore consider the following interaction potential:

$$V_r(x) = (1 + Kr)V_{\mathrm{LJ}}(x) - V_{\mathrm{LJ}}(d_c) = \frac{1 + Kr}{8}\left(\frac{1}{(1+x)^4} - \frac{2}{(1+x)^2}\right) - V_{\mathrm{LJ}}(d_c). \tag{5.34}$$

The potential stiffens as the reaction goes on. The reaction starts when enough energy has been stored in the media, for example when the media is compressed enough (a less naive ignition of the reaction is proposed in Section 5.2.3). For the bond Δx_n, this corresponds to the first time t^* such that $\Delta x_n < d_c$, where $d_c < 0$ is a parameter (critical distance). By construction, the potential is continous at $x = d_c$. For $t \geq t^*$, the kinetics of the reaction is assumed to be

$$\frac{dr_n}{dt}(t) = D \quad \text{if} \quad 0 \leq r_n(t) \leq 1, \qquad \frac{dr_n}{dt}(t) = 0 \quad \text{otherwise} ,$$

or possibly

$$\frac{dr_n}{dt}(t) = D(1 - r_n(t))$$

for a first-order kinetics. The bond $\Delta x_n(t)$ is then described by the potential $V_{r_n(t)}$, using (5.34). The exothermicity of the reaction is ensured provided $d_c < 0$, and is parametrized by K and d_c. Figure 5.15 presents an example of modification of the potential when a reaction occurs.

Modification of the parameters in the generalized Langevin equation

The derivation of (5.29) uses parameters describing some absorption spectrum. However, as the chemical reaction goes on, the mechanical properties of the media evolve, and so, the parameters of the absorption spectrum should evolve as well. Since the interaction potentials get stiffer by a factor $1 + Kr_n$, we arbitrarily modify the distribution of the pulsations $\{\omega\}$, and replace ω^2

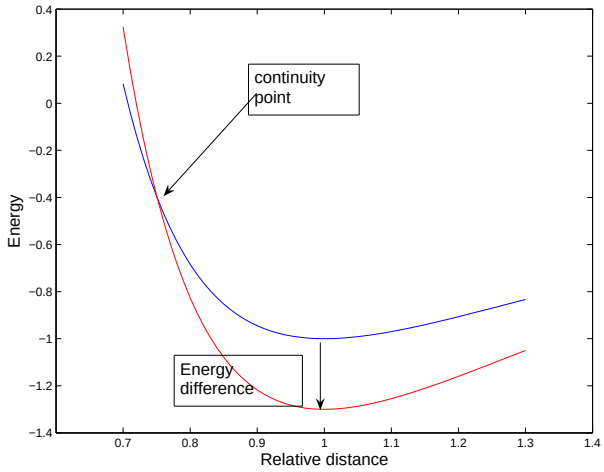


Fig. 5.15. Modification of the potential during the reaction (initial potential: upper curve, final potential: lower curve). Note that the equilibrium position is preserved, but the ground state is lower.

par $(1 + Kr_n)\omega^2$. analogously, α is replaced by $\alpha\sqrt{1 + Kr_n}$ and λ by $\lambda\sqrt{1 + Kr_n}$, while keeping the $\{\gamma_j\}$ unchanged.

Numerical results

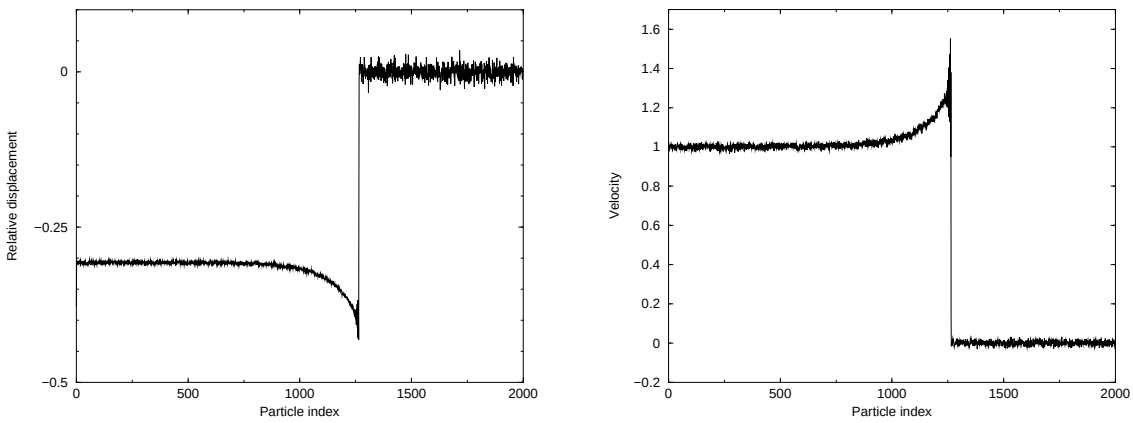


Fig. 5.16. Sustained compression of reactive shock waves. Displacement profiles (Left) and velocity profiles (Right) for a single realization of a sustained shock compression. The parameters are the same as for Figure 5.13, with $K = 1$, $d_c = -0.3$, $D = 0.025$ and a first-order reaction kinetics.

Profiles reminiscent of classical ZND profiles are recovered, with shocks stronger than in the non-reactive case and propagating faster (see Figure 5.16). The shock is also followed by a relaxation wave. When the piston is removed, a steady-state shock front is finally obtained, which is not weakened by the downstream rarefactions (see Figure 5.17). However, the material returns to equilibrium after some relaxation period, whereas a fluid behavior is expected when detonation takes place (the order in the material being completely lost because of the large energy release). Therefore, the 1D model, even augmented, is not convenient to model detonation of real materials.

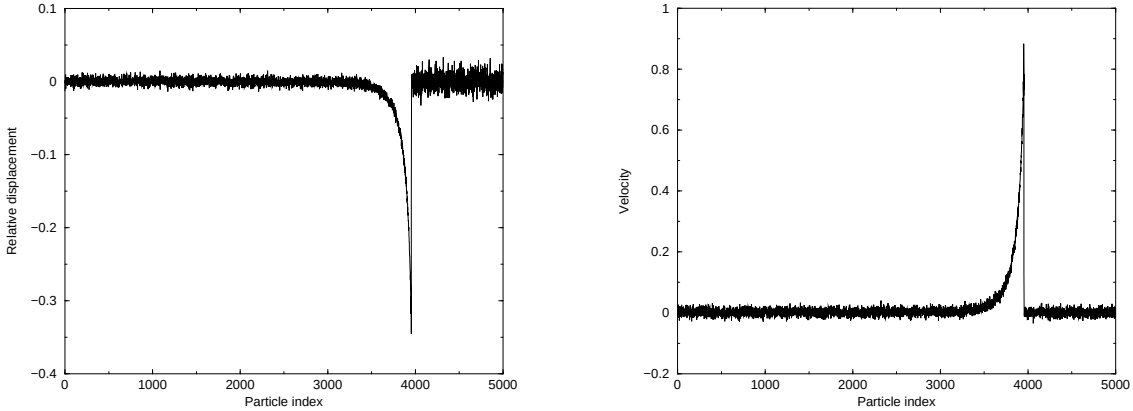


Fig. 5.17. Same parameters as for Figure 5.16, a compression time $T_{\text{comp}} = 20$ and a relaxation time $T_{\text{relax}} = 1500$.

5.2 A reduced model based on Dissipative Particle Dynamics

The reduced model (5.29) obtained in Section 5.1 is reminiscent of DPD models since the friction depends on the relative velocities of the particles. However, the temperature effects are not correctly taken into account. Let us emphasize at this point that keeping thermal fluctuations in the microscopic models is of paramount importance to obtain the right relaxation profiles behind the shock front [162, 323].

It is not possible to resort directly to the classical DPD models to simulate shock waves. Indeed, the dissipative and random forces arising in DPD are linked through some fluctuation-dissipation relation, using a local temperature. But when a shock wave passes, energy is transfered to the material, and the local temperature changes. Therefore, it is necessary to consider DPD models where the fluctuation-dissipation relation is not fixed *a priori*, but evolves depending on the physical events that have happened. DPD with conserved energy [15, 95] are such models.

DPD models, introduced in [170], have been put on firm thermodynamics ground in [98]. Some derivations from molecular dynamics where proposed in a simplified case in [94], the more convincing general derivation being at the moment [106]. These studies motivate the modelling of the mean action of the projected degrees of freedom through some dissipative forces (depending on the relative velocities of the particles, so that the global momentum is conserved), balanced by some random forces. Ergodicity of the dynamics can be shown in some simplified cases [307]. Therefore, DPD dynamics are well established and motivated reduced models.

Coarser models such as SPH (Smoothed particle hydrodynamics) [217,246] are routinely used to simulate shock waves at the hydrodynamic level, and can also be formulated in a DPD framework (the so-called Smoothed dissipative particle dynamics [96]). However, these models require the knowledge of an equation of state $E_{\text{int}} = E_{\text{int}}(S, P)$ giving the internal energy as a function of entropy and pressure, for instance. Therefore, SPH-like models cannot be considered when the coarse-grained model is still at the microscopic level.

We present in this section a dynamics strongly inspired by those models, and show that it provides an interesting mesoscopic model for the simulation of shock waves (see Section 5.2.2 and [324]). It also opens the way for an extension to detonation waves, where exothermic chemical reactions are triggered as the shock passes, with the shock sustained and enhanced through the energy released (see Section 5.2.3 and [222]).

5.2.1 Previous mesoscopic models

We review here some mesoscopic models [163, 326] for shock waves, obtained through a coarse-graining from microscopic (all-atom) models. The model from [163] is more empirical and has been

derived to recover certain properties of polycrystalline materials. One particle stand for a grain in this case, and some assumptions are made on the mechanical behavior at grain boundaries. The model from [326] considers the elementary coarse-graining, in which a complex molecule is replaced by a single fictitious particle with internal degrees of freedom (internal energy).

In both [163,326], the dissipation forces acting on the i-th particle are of the form $-\gamma(v_i - \bar{v}_i)$, where $\bar{v}_i$ is a local average of the velocities around the particle. We will focus in the sequel on the model [326], in which the Hamiltonian equations of motions are then perturbed by additional terms:

$$\begin{cases} \dfrac{dq_i}{dt} = \dfrac{p_i}{m_i} - \chi_i \nabla V_{q_i}(q), \\ \dfrac{dp_i}{dt} = -\nabla V_{q_i}(q) - \dfrac{\eta_i}{m_i}(v_i - \bar{v}_i). \end{cases}$$

It is assumed that the variations of mechanical energy are exactly compensated by the variations of internal energy. Associating an internal energy ϵ_i to each particle (see Section 5.2.2), it follows

$$dE_{\text{tot}} = dE_{\text{mec}} + \sum_{i=1}^{N} d\epsilon_i = d\left[\frac{1}{2}p^T M p + V(q)\right] + \sum_{i=1}^{N} d\epsilon_i = 0.$$

Therefore,

$$\frac{d\epsilon_i}{dt} = \eta_i(v_i - \bar{v}_i) \cdot v_i + \chi_i |\nabla V_{q_i}(q)|^2.$$

The authors of [326] then argue that this energy transfer is not Galilean invariant (in view of the first term on the right hand side in the above equation: $v_i - \bar{v}_i$ is Galilean invariant, but v_i is not), even if the dynamics is. To remedy this problem, they restrain themselves to dissipation on the position variable q only, and do not consider dissipation in the momenta ($\eta_i = 0$). A stable dynamics is obtained by considering a coefficient χ_i depending on the difference between the internal and the external (translational or mechanical temperature), and a Berendsen-like feedback. The resulting dynamics is not completely satisfactory from a physical viewpoint since it has a structure very different of Newton's equation. It is also not clear whether an invariant measure exists.

It is however possible to preserve the Galilean invariance by considering pair friction forces, depending on the relative velocities of the particles as is done in DPD models. In this case, the energy exchanges can indeed be symmetrized, and the resulting process is totally Galilean invariant. The resulting dynamics, of DPD form, are physically more natural then the damped dynamics of [326].

5.2.2 A reduced model in the inert case

Description of the model

All atom simulations are performed resorting to Newton's equations of motion. The corresponding microscopic systems are deterministic, Galilean invariant, and have some invariants, such as the total energy. While stochastic models are natural models to describe systems with reduced dynamics (since the information lost by the averaging process is modelled by some random process), it is however not clear that such a stochastic model can reproduce, even in a mean way, a deterministic dynamics with invariants.

It turns out however that DPD models are stochastic dynamics which are Galilean invariant and preserve total momentum. Some refinements were also proposed in order to conserve the total energy of the system, a model called 'DPD with conserved energy' (DPDE [15,95]).

We consider a system of N particles in a space of dimension d, described by their positions $(q_1, \ldots, q_N)$ and momenta $(p_1, \ldots, p_N)$, with associated mass matrix $M = \text{Diag}(m_1, \ldots, m_N)$,

interacting through a potential $\mathcal{V}$. We assume for simplicity that the interactions between the particles are pairwise and depend only on the relative distances, so that $\mathcal{V}(q) = \sum_{i<j} V(|q_i - q_j|)$. Denoting by $\bar{T}$ the reference temperature and $\beta = 1/(k_{\mathrm{B}}\bar{T})$, the DPD equations read [98, 170]

$$\begin{cases} dq_i = \dfrac{p_i}{m_i}\, dt, \\ dp_i = \displaystyle\sum_{j\neq i} -\nabla V(r_{ij})\, dt - \gamma\chi^2(r_{ij})(v_{ij}\cdot e_{ij})e_{ij} + \sqrt{\frac{2\gamma}{\beta}}\chi(r_{ij})dW_{ij}\, e_{ij}, \end{cases} \tag{5.35}$$

with $\gamma > 0$, $r_{ij} = |q_i - q_j|$, $e_{ij} = (q_i - q_j)/r_{ij}$, $v_{ij} = \frac{p_i}{m_i} - \frac{p_j}{m_j}$, χ a weight function (with support in $[0, r_c]$ where r_c is a cut-off radius), and where W_{ij} are 1-dimensional independent Wiener processes such that $W_{ij} = W_{ji}$.

Notice that, since the dissipation term depends only on the relative velocities, the dynamics are globally Galilean invariant. Besides, the total momentum is preserved. However, the total energy fluctuates, so that some refinements in the model are required. Relying on the general DPD picture, DPD with conserved energy were introduced in [15, 95]. The idea is that the variations of the total mechanical energy $H(q,p) = \frac{1}{2}p^T Mp + \mathcal{V}(q)$ through the dissipative forces are compensated by some reservoir energy variable attached to each particle. Introducing an internal energy ϵ_i for each particle, the evolution of the internal energies are constructed such that

$$dH(q,p) + \sum_i d\epsilon_i = 0.$$

An associated entropy $s_i = s(\epsilon_i)$ and an internal temperature can be also defined for each particle as

$$T_i = \left(\frac{\partial s_i}{\partial \epsilon_i}\right)^{-1}.$$

For example, when the internal degrees of freedom are purely harmonic, $T(\epsilon) = \epsilon/C_v$, where C_v is the specific heat at constant volume. More generally, this microscopic state law should be computed using all-atom MD or *ab initio* simulations.

The model we consider is strongly inspired from DPD models with conserved energy [15, 95], so that all the properties of the usual DPD models with conserved energy can be straightforwardly transposed to this case. The derivation of the model is done as in [15, 95]. The main differences here is that (i) we present the dynamics for particles of unequal masses, and (ii) do not project the dissipatives and random forces along the lines of center of the particles. The generalization to particles of unequal masses is done by considering dissipation forces depending on the relative velocities, and not on the relative momenta. This is important if mixtures composed of (say) two molecules are simulated, and each molecule is replaced by a single particle, whose mass is the total mass of the molecule. The dissipative and random forces could be projected as well to conserve angular momentum, but we restrict ourselves to the simpler and more general case when these forces are not projected, since we are only interested in Galilean invariance, and have in mind an extension to reduced models for reactive shock waves, which do not necessarily preserve angular momentum, even if the dissipative and random forces are projected. Such a model is also closer to the Langevin picture of the previous reduced models for shock waves [163, 326].

We finally neglect the thermal conduction here, since the contribution to the evolution of the internal energy arising from the dissipation forces is expected to be dominant in the nonequilibrium zone near the shock front. Heat diffusion plays a role only after the relaxation towards equilibrium in the shocked zone is achieved.

The equations of motion for the system read:

$$\begin{cases} dq_i = \dfrac{p_i}{m_i}\,dt, \\ dp_i = \displaystyle\sum_{j,\,j\neq i} -\nabla V(r_{ij})\,dt - \gamma_{ij}\chi^2(r_{ij})v_{ij}\,dt + \sigma_{ij}\chi(r_{ij})dW_{ij}, \end{cases} \tag{5.36}$$

where χ is still a weight function (with support in $[0, r_c]$ where r_c is a cut-off radius), and W_{ij} are now d-dimensional independents Wiener processes such that $W_{ij} = -W_{ji}$. The friction γ_{ij} and the fluctuation magnitude σ_{ij} will be precised below. As for DPD models with conserved energy, the dynamics is postulated in a manner such that the total energy $E(q,p,\epsilon) = H(q,p) + \sum_i \epsilon_i$ is preserved. The evolution of $dH = -\sum_i d\epsilon_i$ is inferred from (5.36) using Itô rule (see [95] for more details). Therefore, we consider the following dynamics:

$$\boxed{\begin{aligned} dq_i &= \frac{p_i}{m_i}\,dt, \\ dp_i &= \sum_{j,\,j\neq i} -\nabla V(r_{ij})\,dt - \gamma_{ij}\chi^2(r_{ij})v_{ij}\,dt + \sigma_{ij}\chi(r_{ij})dW_{ij}, \\ d\epsilon_i &= \frac{1}{2}\sum_{j,\,j\neq i}\left(\chi^2(r_{ij})\gamma_{ij}v_{ij}^2 - \frac{d\sigma_{ij}^2}{2}\left(\frac{1}{m_i}+\frac{1}{m_j}\right)\chi^2(r_{ij})\right)dt - \sigma_{ij}\,\chi(r_{ij})v_{ij}\cdot dW_{ij}, \end{aligned}} \tag{5.37}$$

with the fluctuation-dissipation relation [15, 95] :

$$\sigma_{ij} = \sigma, \quad \gamma_{ij} = \sigma^2\beta_{ij}/2, \quad \beta_{ij}^{-1} = 2k_{\mathrm{B}}(T_i^{-1} + T_j^{-1})^{-1}.$$

It is then easily checked that measures of the form

$$d\rho(q,p,\epsilon) = \frac{1}{Z_{P,E}}\mathrm{e}^{-\beta H(q,p)}\exp\left(\sum_i \frac{s(\epsilon_i)}{k_{\mathrm{B}}} - \beta\epsilon_i\right)\delta_{E=E_0}\,\delta_{P=P_0}\;dq\,dp\,d\epsilon \tag{5.38}$$

are invariant [15]. This measure expresses the fact that the translational degrees of freedom are distributed according to a classical Boltzmann statistics, whereas the internal energies are distributed according to some free energy statistics. The total momentum $P_0 = \sum_i p_i$ and the total energy $E_0 = E(q,p,\epsilon)$ are also preserved by construction.

If the dynamics is ergodic for the measure (5.38) and in the limit $N \to +\infty$, it holds

$$k_{\mathrm{B}}\langle T_{\mathrm{kin}}\rangle = \beta^{-1}, \qquad k_{\mathrm{B}}(\langle T_{\mathrm{int}}^{-1}\rangle)^{-1} = \beta^{-1},$$

with

$$T_{\mathrm{kin}} = \frac{1}{k_{\mathrm{B}}dN}\sum_{i=1}^{N}\frac{p_i^2}{m_i}, \qquad \frac{1}{T_{\mathrm{int}}} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i},$$

and $\langle A\rangle = \int A(q,p)\,\rho(q,p,\epsilon)\,dq\,dp\,d\epsilon$. Indeed, as $T_i^{-1} = s'(\epsilon_i)$, and assuming $s(\epsilon) \to -\infty$ when $\epsilon \to 0$, $s(\epsilon)/\epsilon \to 0$ when $\epsilon \to +\infty$ (which is the case when $s(\epsilon) = C_v \ln\epsilon$),

$$\left\langle \frac{1}{k_{\mathrm{B}}T_i}\right\rangle = \frac{\displaystyle\int_0^{+\infty}\frac{s'(\epsilon_i)}{k_{\mathrm{B}}}\exp\left(\frac{s(\epsilon_i)}{k_{\mathrm{B}}} - \beta\epsilon_i\right)d\epsilon_i}{\displaystyle\int_0^{+\infty}\exp\left(\frac{s(\epsilon_i)}{k_{\mathrm{B}}} - \beta\epsilon_i\right)d\epsilon_i} = \beta.$$

Notice that these relationships provide estimators for the local thermodynamic temperature $\beta^{-1}/k_{\mathrm{B}}$ through the arithmetic average kinetic temperatures, and the *harmonic* average inter-

nal temperatures. Let us emphasize that a straightforward arithmetic average over the internal temperatures would give wrong results (the corresponding estimator being biased).

A deterministic version of the model

We intend here to introduce a deterministic version of our model, which allows to bridge the gap between a previous mesoscopic deterministic model [326] (see also Section 5.2.1) and the DPD framework for shock waves. The model proposed in [326] introduces damping forces on the position variables directly (and not on the momentum variables as would be expected) in order to preserve the Galilean invariance. Indeed, the damping terms in the momentum variable are considered to be of the form $-\gamma(v_i - \bar{v}_i)$, where $\bar{v}_i$ is a local average of the velocities around the particle, which makes the Galilean invariance of the dissipated energy difficult to preserve. If on the other hand the dissipation term in the momentum variable implies only pairwise velocity differences as for DPD models, the Galilean invariance follows immediately. The following equations of motion then mix the deterministic equations of motion of [326] and the DPD philosophy:

$$\begin{cases} dq_i = \dfrac{p_i}{m_i}\,dt, \\ dp_i = \displaystyle\sum_{j,\,j\neq i} -\nabla V(r_{ij})\,dt - \gamma\frac{T_{ij}^{\text{ext}} - T_{ij}^{\text{int}}}{\bar{T}}\omega(r_{ij})v_{ij}\,dt, \\ d\epsilon_i = \dfrac{1}{2}\displaystyle\sum_{j,\,j\neq i} \gamma\frac{T_{ij}^{\text{ext}} - T_{ij}^{\text{int}}}{\bar{T}}\omega(r_{ij})v_{ij}^2\,dt, \end{cases}$$

where T_{ij}^{ext} is the average temperature in the kinetic degrees of freedom of particles i and j (for example, $T_{ij}^{\text{ext}} = (T_i^{\text{ext}} + T_j^{\text{ext}})/2$ with $T_i^{\text{ext}} = 2p_i^2/k_{\text{B}}dm_i$ the kinetic temperature associated with particle i) and T_{ij}^{int} is the average internal temperatures of particles i and j (for example, $T_{ij}^{\text{int}} = (T_i^{\text{int}} + T_j^{\text{int}})/2$). The function ω is still a weighting function, and γ determines the strength of the coupling.

Notice that the dissipation term is in fact a dissipation term only when $T_{ij}^{\text{ext}} > T_{ij}^{\text{int}}$, and an anti-dissipation term otherwise (and so, is a Nosé-like feedback). This ensures that the internal and external (kinetic thus potential terms) energies equilibriate in all cases. However, the thermodynamic properties of such a model are less clear to state than for the previous stochastic model, and so, we stick to the model (5.37).

Numerical discretization

We use splitting formulas inspired from [305,306]. Recall that the integration of the equation of motion (5.37) is not straightforward since the dissipation terms depend on the relative velocities. We decompose (5.37) into elementary SDEs, and denote by $\phi_{\Delta t}$ the (stochastic) flow map for a time Δt. The elementary SDEs are the usual deterministic Newton part and the dissipation part, which read respectively

$$\begin{cases} dq = M^{-1}p\,dt, \\ dp = -\nabla V(q)\,dt \end{cases} \quad \text{and} \quad \forall i < j, \quad \begin{cases} dp_i = -\gamma_{ij}\chi^2(r_{ij})v_{ij}\,dt + \sigma\chi(r_{ij})\,dW_{ij}, \\ dp_j = -dp_i, \\ d\epsilon_i = -\frac{1}{2}d\left(\frac{p_i^2}{2m_i} + \frac{p_j^2}{2m_j}\right), \\ d\epsilon_j = d\epsilon_i. \end{cases}$$

Denoting by $\phi_{\text{Newton},\Delta t}$ and $\phi_{\text{diss},\Delta t}^{i,j}$ $(1 \leq i < j \leq N)$ the associated stochastic flow maps, an approximation of $\phi_{\Delta t}$ is

$$\phi_{\Delta t} \simeq \phi_{\text{diss},\Delta t}^{1,2} \circ \cdots \circ \phi_{\text{diss},\Delta t}^{N-1,N} \circ \phi_{\text{Newton},\Delta t}.$$

The Newton flow $\phi_{\mathrm{Newton},\Delta t}$ is approximated using a Velocity-Verlet scheme. For an approximation $\Phi^{i,j}_{\mathrm{diss},\Delta t}$ $(i < j)$ of the dissipation part, we first update the velocities at fixed internal temperatures using a Verlet-like algorithm as proposed in [306]. The energy is then updated as

$$\epsilon_i^{n+1} - \epsilon_i^n = \epsilon_j^{n+1} - \epsilon_j^n = \frac{1}{2}\left(\frac{(p_i^{n+1})^2}{2m_i} + \frac{(p_j^{n+1})^2}{2m_j} - \frac{(p_i^n)^2}{2m_i} - \frac{(p_j^n)^2}{2m_j}\right),$$

so that the total energy is indeed conserved by this step. Of course, this integration scheme could be refined, especially the dissipation part.

Application to shock waves

Some numerical simulations of DPD models with conserved energy where proposed in [16, 282], but were concerned only with the computation of thermal conductivities. The corresponding nonequilibrium states were stabilized using steady temperature gradients. The dissipation terms in the DPDE equations of motions were discarded, and only the diffusive part was retained. We present in this section profiles obtained from simulations of shock waves, for which the diffusive part of the dynamics can be discarded, but the dissipative part is of paramount importance to reproduce qualitative and quantitative features of all-atom shock waves. This situation is somehow complementary to the cases studied in [16,282], and, to our knowledge, was never considered before for some physical application.

We consider the crystalline polymer (PVDF) system of [326], the corresponding reduced system being modeled by a two-dimensional (2D) triangular lattice of mesoparticles. Results for the all-atom model can also be found in [326].

The effective interaction potential between mesoparticles is a pairwise Rydberg potential of the form [326] $V(r) = V_{\mathrm{R}}(\lambda(r/r_0) - 1))$ with $V_{\mathrm{R}}(d) = -\epsilon\,(1 + d + \alpha d^3)\,\mathrm{e}^{-d}$. The parameters given by [326] were fitted to reproduce the stress in an uniaxial compression: $\lambda = 7.90$, $\alpha = 0.185$, $r_0 = 5.07$ Å, $\epsilon = 1.612 \times 10^{-20}$ J, $m = 64.03 \times 10^{-3}$ kg/mol. We also choose a cut-off radius $R_{\mathrm{cut}} = 15$ Å for the pairwise interactions. The microscopic state law is obtained by assuming that C_v is independent of the temperature: $\epsilon = C_v T$, with here $C_v = 16\ k_{\mathrm{B}}$ since we represent a three-dimensional molecule formed of 6 atoms by a 2D mesoparticle. In general, the heat capacity is a function of the temperature $C_v = C_v(T)$, and should be parametrized by equilibrium simulations.

We use the simple weight function $\chi(r) = (1 - r/R_{\mathrm{cut}})^2$ if $r \geq R_{\mathrm{cut}}$, $\chi(r) = 0$ otherwise, the cut-off radius R_{cut} being the same as the one used for the potential. Of course, many other weight functions could be used. We also set $\gamma = 1.5 \times 10^{-14}$ kg/s and $\Delta t = 10^{-14}$ s. In these preliminary tests of the model, the parameter γ was varied to obtain a good agreement with the all-atom results. However, it is expected that γ is linked to some physical quantity, such as the decay rate of the relative velocities autocorrelation in an all-atom simulation, and could therefore be estimated using some preliminary small equilibrium simulations.

We first prepare an initial state according to the invariant measure (5.38). To this end, we sample independently the internal energies according to the measure $Z_\epsilon^{-1}\exp(-\beta\epsilon + s(\epsilon)/k_{\mathrm{B}}) = Z_\epsilon^{-1}\epsilon^{C_v/k_{\mathrm{B}}}\exp(-\beta\epsilon)$, and the initial configuration in phase-space by thermalizing a lattice initially at rest, using a Langevin dynamics. In this study, the initial temperature is $T_0 = 300$ K, and the edge of the triangles in the triangular lattice is $a = 5.13$ Å.

We then produce a shock using a piston at velocity $u_{\mathrm{p}} = 3000$ m/s. Figure 5.18 presents the relaxation behind the shock front for the 2D triangular lattice of mesoparticles subjected to the dynamics (5.37). The results are in good agreement with the all-atom results of [326]. In particular, the final temperature is very close to the all-atom value (whereas it is of course greatly overestimated by the mesoscopic dynamics without coupling), and the time required for the internal temperatures and kinetic temperatures to equilibriate is almost the time needed in all-atom studies.

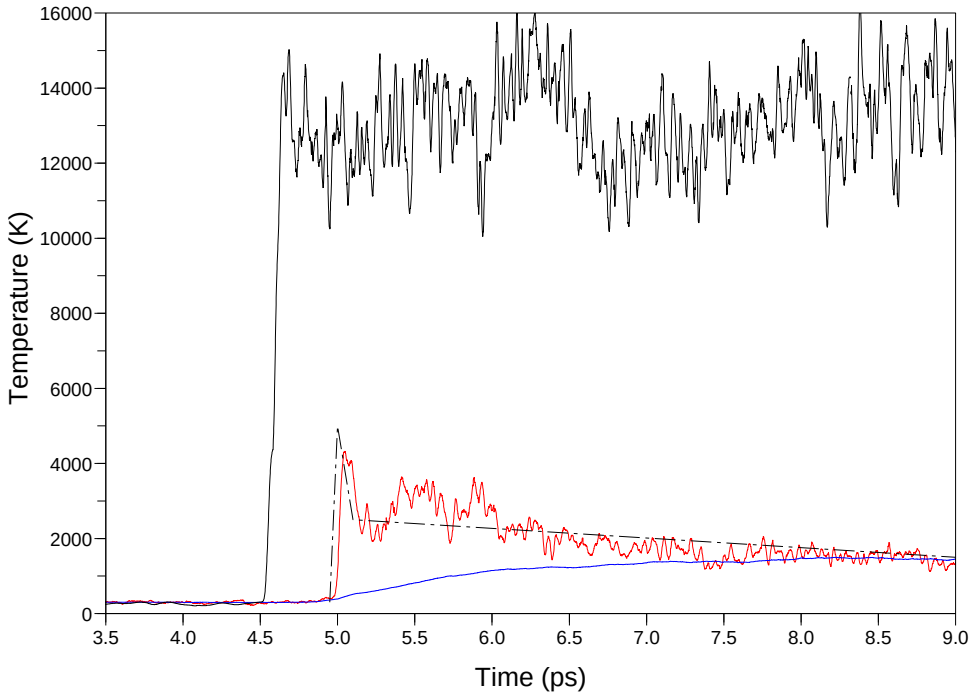


Fig. 5.18. Temporal evolution of the temperature of a thin slab of material as the shock runs trough it: mean kinetic temperature $\hat{T}_{\text{kin}}$ in the direction of the shock (intermediate curve, red), mean internal temperature $\hat{T}_{\text{int}}$ (lower curve, blue). The corresponding results when the coupling with the internal degress of freedom is turned off are also shown (upper curve, black), and a cartoon representation of the all-atom result from [326] for the kinetic temperature $\hat{T}_{\text{kin}}$ is also plotted (dark dash dotted line).

5.2.3 The reactive case

In the reactive case, exothermic chemical reactions are triggered when the shock passes, and the energy liberated sustains the shock. To model detonation at the mesoscopic level, we introduce an additional variable per mesoparticle, namely a progress variable. The dynamics can then be split into three elementary physical processes:

(i) the translational dynamics of the particles, given by the dynamics of inert materials (see Eq. (5.37));

(ii) the evolution of the chemical reaction through some kinetics;

(iii) the exothermicity of the reaction: energy transfers between chemical energy and mechanical and internal energies have to be precised.

Treating the exothermicity

In the reactive case, chemical reactions are triggered when the shock passes. To model the progress of the reaction, an additional degree of freedom, a progress variable λ_i, is attached to each particle. For the model reaction

$$2\text{AB} \rightleftarrows \text{A}_2 + \text{B}_2, \tag{5.39}$$

the state $\lambda = 0$ corresponds to a molecule AB, whereas the state $\lambda = 1$ corresponds to $A_2 + B_2$. Representing the progress of the chemical reaction by a real-value parameter makes sense when the mesoparticle represent a blob of material, but seems questionable when a mesoparticle stands for a single molecule. Therefore, the progress variable should be seen as some dissociation probability, or progress along some free energy profile.

Since the model reaction (5.39) involves two species on each side, we postulate for example a reversible evolution of order 2:

$$\frac{d\lambda_i}{dt} = \sum_{j \neq i} \omega_r(r_{ij}) \left[K_1(T_{ij})(1-\lambda_i)(1-\lambda_j) + K_2(T_{ij})\lambda_i\lambda_j \right], \tag{5.40}$$

the function ω_r being a weight function (with support in $[0, r_{\text{reac}}]$), and the mean temperature $T_{ij} = (T_i + T_j)/2$. The reaction constants K_1, K_2 are assumed to depend only on internal temperatures of the particles. For example, a possible form in the Arrhénius spirit is:

$$K_1(T) = A_1 \, \mathrm{e}^{-E_1/k_\mathrm{B}T}, \qquad K_2(T) = A_2 \, \mathrm{e}^{-E_2/k_\mathrm{B}T}. \tag{5.41}$$

The total increment of the progress variable is therefore the sum of all elementary pair increments, which is very much in the DPD spirit. Other kinetics (for example, using some local averaged internal temperatures $\langle T \rangle_i$ and local averaged progress variables $\langle \lambda \rangle_i$) are of course possible.

For very exothermic reactions, $E_2 \gg E_1$, and both energies are large since the activation energy is usually large for energetic materials. The increment of a given progress variable is non-negligible only if the material is locally heated enough. In practice, this can be achieved when a strong shock is initiated in the system. If this shock is not strong enough, chemical reactions do not occur fast enough, and since the energy released is not sufficient, the shock wave is weakened until it disappears. On the contrary, if the shock wave is strong enough, the chemical reactions happen close enough from the detonation front, and the energy released sustains the shock wave.

The progress of the reaction also modifies the mechanical properties of the material. In particular, reaction products usually have a larger specific volume than reactants (at fixed thermodynamic conditions). Therefore, some expansion is expected. The changes in the nature of the molecules are taken into account by introducing two additional parameters k_a, k_E and using some mixing rule such as Berthelot's rule. When the interaction potential is of Lennard-Jones form, the interaction between the mesoparticles i and j separated by a distance r_{ij} is then given by

$$V(r_{ij}, \lambda_i, \lambda_j) = 4E_{ij} \left(\left(\frac{a_{ij}}{r_{ij}} \right)^{12} - \left(\frac{a_{ij}}{r_{ij}} \right)^{6} \right), \tag{5.42}$$

with $E_{ij} = E\sqrt{(1+k_E\lambda_i)(1+k_E\lambda_j)}$, $a_{ij} = a\left(1 + k_a \frac{\lambda_i + \lambda_j}{2}\right)$. When the reaction is complete, the material initially described by a Lennard-Jones potential of parameters a, E is then described by a Lennard-Jones of parameters $a' = a(1+k_a)$ and $E' = E(1+k_E)$.

We denote by ΔE_{exthm} the exothermicity of the reaction (5.39). It is expected that $\Delta E_{\text{exthm}} = E_2 - E_1$. We assume that the energy is liberated progressively during the reaction, in a manner that the total energy of the system (chemical, mechanical, internal) is preserved:

$$dH_{\text{tot}}(q, p, \epsilon, \lambda) = d\left[\sum_{1 \le i < j \le N} V(r_{ij}, \lambda_i, \lambda_j) + \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \epsilon_i + (1-\lambda_i)\Delta E_{\text{exthm}} \right] = 0.$$

In order to propose a dynamics satisfying this condition, we have to make an additional assumption about the evolution of the system. Negelecting diffusive processes, we require that, during the elementary step corresponding to exothermicity, the total energy of a given mesoparticle does not

change[2]:

$$d\left[\frac{1}{2}\sum_{i\neq j} V(r_{ij},\lambda_i,\lambda_j)\right] + d\left(\frac{p_i^2}{2m_i}\right) + d\epsilon_i - \Delta E_{\text{exthm}} d\lambda_i = 0. \tag{5.43}$$

We then consider evolutions of momenta and internal energies balancing the variations in the total energy due to the variations of λ (exothermicity, changes in the potential energies). This is analogous to the fact that the variations of kinetic energy in (5.37) are compensated by variations of internal energies. The variations in total energy are distributed between internal energies and kinetic energies following some predetermined ratio $0 < c < 1$. For the internal energies,

$$d\epsilon_i = -c\left(d\left[\frac{1}{2}\sum_{i\neq j} V(r_{ij},\lambda_i,\lambda_j)\right] - \Delta E_{\text{exthm}} d\lambda_i\right).$$

For the momenta, we consider a process Z_i^p such that $dp_i = dZ_i^p$ with

$$d\left(\frac{p_i^2}{2m}\right) = -(1-c)\left(d\left[\frac{1}{2}\sum_{i\neq j} V(r_{ij},\lambda_i,\lambda_j)\right] - \Delta E_{\text{exthm}} d\lambda_i\right).$$

We explain in the next section how this is done in practice (see Eq. (5.46)).

Let us emphasize at this point that there are many other possible ways to treat the exothermicity. For instance, it would be possible to consider instantaneous reactions (jump processes for which λ changes from 0 to 1) occuring at random times, the probability of reaction depending on the progress variable. However, it is not clear whether such a dynamics is reversible, since the reverse reaction requires particles to have large kinetic and internal energies. In comparison, the process described here is progressive and therefore, much more reversible.

Finally, we propose the following dynamics to describe reactive shock waves:

$$\boxed{\begin{aligned}
dq_i &= \frac{p_i}{m_i}\,dt,\\
dp_i &= \sum_{j,\,j\neq i} -\nabla_{q_i} V(r_{ij},\lambda_i,\lambda_j)\,dt - \gamma_{ij}\chi^2(r_{ij})v_{ij}\,dt + \sigma\chi(r_{ij})dW_{ij} + dZ_i^p,\\
d\epsilon_i &= \frac{1}{2}\sum_{j,\,j\neq i}\left(\chi^2(r_{ij})\gamma_{ij}v_{ij}^2 - \frac{d\sigma^2}{2}\left(\frac{1}{m_i}+\frac{1}{m_j}\right)\chi^2(r_{ij})\right)dt\\
&\qquad\qquad -\sigma\,\chi(r_{ij})v_{ij}\cdot dW_{ij} + dZ_i^\epsilon,\\
d\lambda_i &= \sum_{j\neq i}\omega_r(r_{ij})\left[K_1(T_{ij})(1-\lambda_i)(1-\lambda_j) + K_2(T_{ij})\lambda_i\lambda_j\right]dt,
\end{aligned}} \tag{5.44}$$

where dZ_i^p, dZ_i^ϵ are such that (5.43) holds, *i.e.* the total energy is conserved. The fluctuation-dissipation relation relating γ_{ij} and σ is the same as for (5.37). Notice also that the inert dynamics (5.37) is recovered when $A_1 = A_2 = 0$, starting from $\lambda_i = 0$ for all i.

Numerical implementation

The numerical integration of (5.44) is done using a decomposition of the dynamics into elementary stochastic differential equations. We denote by ϕ^t_{inert} the flow associated with the dynamics (5.37), and by ϕ^t_{reac} the flow associated with the remaining part of the dynamics (5.44):

[2] Of course, during the elementary step corresponding to the dynamics (5.37), the total energy changes.

$$\forall 1 \leq i \leq N, \quad \begin{cases} d\lambda_i = \sum_{j \neq i} \omega_r(r_{ij}) \left[K_1(T_{ij})(1-\lambda_i)(1-\lambda_j) + K_2(T_{ij})\lambda_i\lambda_j \right] dt, \\ dp_i = dZ_i^p, \\ d\epsilon_i = dZ_i^\epsilon. \end{cases} \tag{5.45}$$

A one-step integrator for a time-step Δt is constructed as $(q^{n+1}, p^{n+1}, \epsilon^{n+1}, \lambda^{n+1}) = \Phi_{\text{reac}}^{\Delta t} \circ \Phi_{\text{inert}}^{\Delta t}(q^n, p^n, \epsilon^n, \lambda^n)$. A possible numerical flow $\Phi_{\text{inert}}^{\Delta t}$ is given in Section 5.2.3.

Let us now construct a numerical flow $\Phi_{\text{reac}}^{\Delta t}$ approximating the flow $\phi_{\text{reac}}^{\Delta t}$. Denoting $(q^{n+1}, \tilde{p}^n, \tilde{\epsilon}^n, \lambda^n) = \Phi_{\text{inert}}^{\Delta t}(q^n, p^n, \epsilon^n, \lambda^n)$, we first integrate the evolution equation on the progress variables λ_i using a first-order explicit integration:

$$\tilde{\lambda}_i^{n+1} = \lambda_i^n + \left[\sum_{j \neq i} \omega_r(r_{ij}^{n+1}) K_1(\tilde{T}_{ij}^n)(1-\tilde{\lambda}_i^n)(1-\tilde{\lambda}_j^n) + K_2(\tilde{T}_{ij}^n)\tilde{\lambda}_i^n\tilde{\lambda}_j^n \right] \Delta t.$$

We then set $\lambda_i^{n+1} = \min(\max(0, \tilde{\lambda}_i^{n+1}), 1)$ in order to ensure that the progress variable remains between 0 and 1. Once all progress variables are updated, the variation δE_i^n in the total energy of particle i due to the variations of $\{\lambda_j\}$ is computed as

$$\delta E_i^n = (\lambda_i^{n+1} - \lambda_i^n)\Delta E_{\text{exthm}} + \frac{1}{2} \sum_{j \neq i} \left(V(r_{ij}^{n+1}, \lambda_i^{n+1}, \lambda_j^{n+1}) - V(r_{ij}^{n+1}, \lambda_i^n, \lambda_j^n) \right).$$

The conservation of total energy is then ensured through variations of internal and kinetic energies. The internal energies are updated as $\epsilon_i^{n+1} = \tilde{\epsilon}_i^n + c\,\delta E_i^n$. The update of p_i^{n+1} is done by adding to p_i^n a vector with random direction, so that the final momentum is such that the kinetic energy is correct. More precisely, when the dimension of the physical space is $d = 2$ for example, an angle θ_i^n is chosen at random in the interval $[0, 2\pi]$, the angles $(\theta_i^n)_{i,n}$ being idependent and identically distributed (i.i.d.) random variables. The new momentum p_i^{n+1} is then constructed such that

$$p_i^{n+1} = p_i^n + \alpha^n(\cos\theta^n, \sin\theta^n), \qquad \frac{(p_i^{n+1})^2}{2m_i} = \frac{(\tilde{p}_i^n)^2}{2m_i} + (1-c)\,\delta E_i^n. \tag{5.46}$$

Solving this equation in α^n gives the desired result.

Numerical results

We present in this section numerical results obtained for the dynamics (5.44) for a two-dimensional fluid. A shock is initiated using a piston of velocity u_{p} during a time t_{p}. The initial conditions for the positions q_i, momenta p_i and internal energies ϵ_i are sampled as proposed in Section 5.2.2.

We consider the following parameters, inspired by the nitromethane example, where the molecule CH_3NO_2 is replaced by a mesoparticle in a space of 2 dimensions. The parameters can be classified in four main categories, the ones describing the mechanical properties of the material, the parameters used to characterize the inert dynamics and the chemical kinetics, and the parameters related to the exothermicity. We consider here a system with

(i) (Material parameters) a molar mass $m = 80$ g/mol, described by a Lennard-Jones potential of parameter $E_{\text{LJ}} = 3 \times 10^{-21}$ J (melting temperature around 220 K) and $a = 5$ Å, with a cut-off radius $r_{\text{cut}} = 15$ Å for the computation of forces. The changes in the parameters of the Lennard-Jones material during the reaction follow (5.42), using $k_E = 0$ and $k_a = 0.2$ (pure expansion).

(ii) (Parameters of the inert dynamics) The microscopic state law is $\epsilon = C_v T$ with $C_v = 10\ k_{\mathrm{B}}$ (*i.e.*, 20 degrees of freedom are not represented). The friction is $\gamma = 10^{-15}$ kg/s, and the dissipation weighting function $\chi(r) = (1 - r/r_{\mathrm{c}})$, with $r_{\mathrm{c}} = r_{\mathrm{cut}}$.

(iii) (Chemical kinetics) For the chemical reaction (5.40), reaction constants are computed using (5.41) with $Z_1 = Z_2 = 10^{17}$ s^{-1}, $E_1/k_{\mathrm{B}} = 15000$ K, the exothermicity being $\Delta E_{\mathrm{exthm}} = 6.25$ eV. The reaction weighting function $\omega(r) = \chi(r)$;

(iv) (Exothermicity) we choose $c = 0.5$.

The intial density of the system is $\rho = 1.06$ g/cm^3, and the initial temperature $\bar{T} = 300$ K. The time-step used is $\Delta t = 2 \times 10^{-15}$ s. Figure 5.19 presents velocity profiles averaged in thin slices of the material in the direction of the skock, for a compression time $t_{\mathrm{p}} = 2$ ps at a velocity $u_{\mathrm{p}} = 5000$ m/s. We tested the independence of the resulting profiles for the initial loadings $(t_{\mathrm{p}}, u_{\mathrm{p}}) = (1\ \mathrm{ps}, 6000\ \mathrm{m/s})$, $(t_{\mathrm{p}}, u_{\mathrm{p}}) = (2\ \mathrm{ps}, 6000\ \mathrm{m/s})$, $(t_{\mathrm{p}}, u_{\mathrm{p}}) = (3\ \mathrm{ps}, 6000\ \mathrm{m/s})$ and $(t_{\mathrm{p}}, u_{\mathrm{p}}) = (3\ \mathrm{ps}, 5000\ \mathrm{m/s})$.

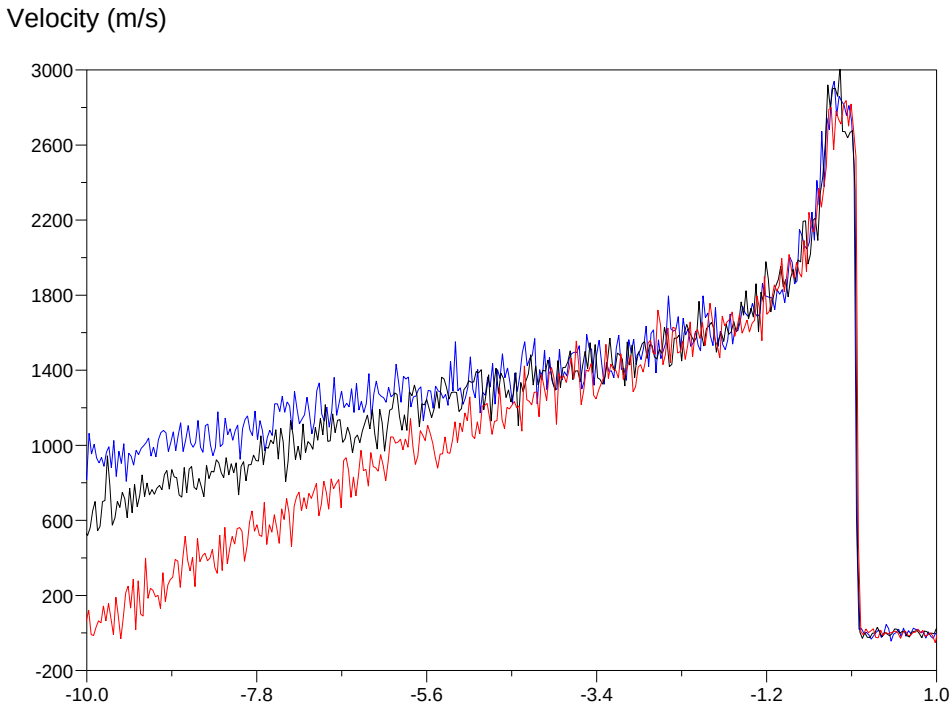


Fig. 5.19. Velocity profiles in the material as a function of the distance to the shock front (in μm) at different times (lower curve (red): $t = 1.2 \times 10^{-10}$ s; middle curve (black): $t = 1.6 \times 10^{-10}$ s; upper curve (blue): $t = 2 \times 10^{-10}$ s).

The velocity of the shock front is constant, and approximately equal to $u_{\mathrm{s}} = 3060$ m/s. Notice that the wave can be divided into three regions: the upstream region is unperturbed; the region around the shock front where chemical reactions happen is of constant width (approximately 300-400 Å, which is consistent with all-atoms studies, see for instance [154]); the downstream region is an autosimilar rarefaction wave. This profile is therefore reminiscent from ZND profiles [103] encountered in hydrodynamic simulations of detonation waves.

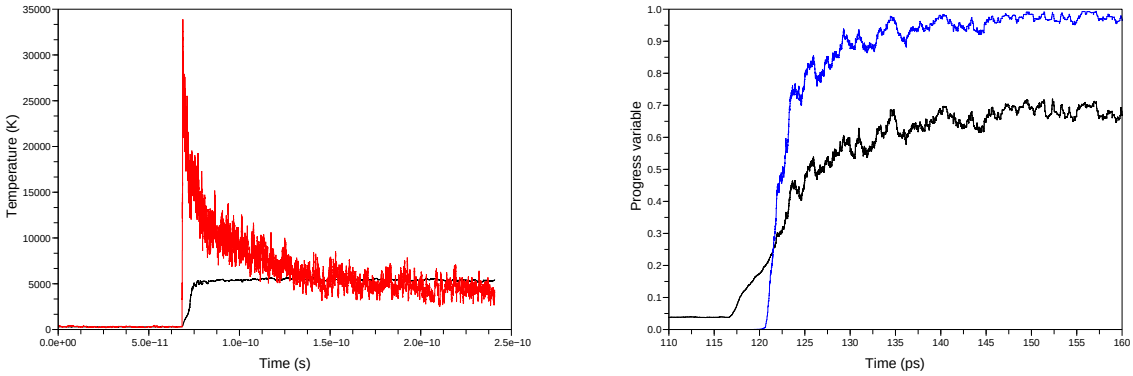


Fig. 5.20. Left: variations of internal (lower curve, black) and kinetic (upper curve, red) temperatures in the direction of the shock, as a function of time in a slice of material. Right: evolution of the progress variable averaged in a slice of material as a function of time (upper curve, blue). For comparison, a rescaled internal temperature profile is also presented (lower curve, black).

Figure 5.20 presents the evolution of internal and kinetic temperatures averaged in a slice of material in the direction of the shock as a function of time (Left), as well as the evolution of the average progress variables (Right). In particular, the reaction does not start immedialely at the shock front: the ignition asks first for a sufficient heating of the material (through an increasing internal energy), since the reaction constant are too low at temperatures lower than a few thousands Kelvins with the values chosen here.

Part IV

Mathematical Study of some Quantum Models

6

Variational Monte-Carlo

6.1 Description of the algorithms **229**
6.1.1 Random walks in the configuration space 229
6.1.2 Random walks in the phase space 231
6.2 Numerical experiments and applications **234**
6.2.1 Measuring the efficiency 234
6.2.2 Numerical results 236
6.2.3 Discussion of the results 238

Most quantities of interest in quantum physics and chemistry are expectation values of the form

$$\frac{\langle \psi, \hat{O}\psi \rangle}{\langle \psi, \psi \rangle} \tag{6.1}$$

where $\hat{O}$ is the self-adjoint operator (the observable) associated with a physical quantity O and Ψ a given wave function. For N-body systems in the position representation, ψ is a function of $3N$ real variables and

$$\frac{\langle \psi, \hat{O}\psi \rangle}{\langle \psi, \psi \rangle} = \frac{\displaystyle\int_{\mathbb{R}^{3N}} [\hat{O}\psi](x)\, \overline{\psi}(x)\, dx}{\displaystyle\int_{\mathbb{R}^{3N}} |\psi(x)|^2\, dx}. \tag{6.2}$$

High-dimensional integrals are very difficult to evaluate numerically by standard integration rules. For specific operators $\hat{O}$ and specific wave functions ψ, *e.g.* for electronic Hamiltonians and Slater determinants built from Gaussian atomic orbitals, the above integrals can be calculated analytically. In some other special cases, (6.2) can be rewritten in terms of integrals on lower-dimensional spaces (typically $\mathbb{R}^3$ or $\mathbb{R}^6$).

In the general case however, the only possible way to evaluate (6.2) is to resort to stochastic techniques. The VMC method [40] consists in remarking that

$$\frac{\langle \psi, \hat{O}\psi \rangle}{\langle \psi, \psi \rangle} = \frac{\displaystyle\int_{\mathbb{R}^{3N}} O_L(x)\, |\psi(x)|^2\, dx}{\displaystyle\int_{\mathbb{R}^{3N}} |\psi(x)|^2\, dx} \tag{6.3}$$

with $O_L(x) = [\hat{O}\psi](x)/\psi(x)$. The above expectation value is reminiscent of expectations values computed in Chapter 3, for the measure

$$d\pi(x) = \frac{|\psi(x)|^2}{\displaystyle\int_{\mathbb{R}^{3N}} |\psi|^2}\, dx. \tag{6.4}$$

This measure can be formally interpreted as a Boltzmann measure $Z^{-1}\,\mathrm{e}^{-\beta V(x)}\,dx$ with the choice $\beta = 1$ and

$$V(x) = -\ln\left(|\psi(x)|^2\right). \tag{6.5}$$

Hence, sampling configurations $(x^n)_{n\geq 1} \in \mathbb{R}^{3N}$ from the measure (6.4), the expectation value (6.3) can be approximated as

$$\frac{\langle \psi, \hat{O}\psi\rangle}{\langle \psi, \psi\rangle} \simeq \frac{1}{L}\sum_{n=1}^{L} O_L(x^n). \tag{6.6}$$

The VMC algorithms described below are generic, in the sense that they can be used to compute the expectation value of any observable, for any N-body system. In the numerical example, we will however focus on the important case of the calculation of electronic energies of molecular systems. In this particular case, the expectation value to be computed reads

$$\frac{\langle \psi, \hat{H}\psi\rangle}{\langle \psi, \psi\rangle} = \frac{\displaystyle\int_{\mathbb{R}^{3N}} E_L(x)\,|\psi(x)|^2\,dx}{\displaystyle\int_{\mathbb{R}^{3N}} |\psi(x)|^2\,dx} \tag{6.7}$$

where the scalar field $E_L(x) = [\hat{H}\psi](x)/\psi(x)$ is called the *local energy*. Remark that if ψ is an eigenfunction of $\hat{H}$ associated with the eigenvalue E, $E_L(x) = E$ for all x. Most often, VMC calculations are performed with trial wave functions ψ that are good approximations of some ground state wave function ψ_0. These trial wavefunctions are sums of single determinantal wave functions built upon Slater-type atomic orbitals, multiplied by a Jastrow factor. More precisely, for a system of N electrons (omitting spin variables and electron-nucleus correlations, see *e.g.* [105] for more general expressions), a typical expression of the wavefunction is

$$\psi(x_1,\ldots,x_N) = \left[\sum_{n=1}^{N_{\mathrm{det}}} a_n \mathrm{Det}(\phi_1^n,\ldots,\phi_N^n)(x_1,\ldots,x_N)\right]\cdot \prod_{1\leq i<j\leq N} \exp\left(\frac{b|x_i - x_j|}{1 + c|x_i - x_j|}\right), \tag{6.8}$$

where the functions ϕ_i^n are atomic-like orbitals

$$\phi_i^n(x) = Z^{-1}_{\alpha_i^n,\xi_i^n,l_i^n,m_i^n}\,|x|^{\alpha_i^n}\,\mathrm{e}^{-\xi_i^n|x|}\,Y_{l_i^n,m_i^n}\left(\frac{x}{|x|}\right).$$

In this last expression, the notation $x/|x|$ is a formal notation for the angles (θ,φ) associated with $x \in \mathbb{R}^3$ in spherical coordinates, and the functions $Y_{l,m}$ are spherical harmonics.

Since the trial wave functions are good approximations of some ground state wave function, $E_L(x)$ usually is a function of low variance (with respect to the probability density $\pi(x)$). This is the reason why, in practice, the approximation formula

$$\frac{\langle \psi, \hat{H}\psi\rangle}{\langle \psi, \psi\rangle} \simeq \frac{1}{L}\sum_{n=1}^{L} E_L(x^n) \tag{6.9}$$

is fairly accurate, even for relatively small values of L (in practical applications on realistic molecular systems L ranges typically between 10^6 and 10^9).

Of course, the quality of the above approximation formula depends on the way the points $(x^n)_{n\geq 1}$ are generated. In Section 6.1.1, we describe the standard sampling method currently used for VMC calculations. It consists in a biased random walk (overdamped Langevin dynamics) in the confi-

guration space $\mathbb{R}^{3N}$ corrected by a Metropolis-Hastings acceptance/rejection procedure. However, the numerical results of Chapter 3 suggest that Langevin dynamics have better sampling properties than overdamped Langevin dynamics. Therefore, in Section 6.1.2, we introduce fictitious masses, and consider a sampling scheme in which the points $(x^n)_{n\geq 1}$ are the projections on the configuration space of one realization of some Markov chain on the phase space $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$. This Markov chain is obtained by a modified Langevin dynamics, corrected by a Metropolis-Hastings acceptance/rejection procedure.

Another advantage of such a dynamics on an extended configuration space is a better behavior close to singularities of the formal potential V (as given by (6.5). Those singularities arise at those points where $\psi(x) = 0$. The set $\psi^{-1}(0)$ is called the nodal surface, and has its origin in the antisymmetric property of the wavefunction. Recall indeed that

$$\psi(x_1, x_2, x_3, \ldots, x_N) = -\psi(x_2, x_1, x_3, \ldots, x_N),$$

so that $\psi(x) = 0$ whenever $x_1 = x_2$ for example. A specific problem encountered in VMC calculations on fermionic systems is that the standard discretization of the biased random walk (Euler scheme) does not behave properly close to the nodal surface of the trial wave function ψ. This is due to the fact that the drift term blows up as the inverse of the distance to the nodal surface: if a random walker gets close to the nodal surface, the drift term repulses it far apart in a single time step. In some studies [47, 352], this difficulty is partially circumvented by resorting to more clever discretization schemes. Using here a Langevin dynamics, the walkers have a mass (hence some inertia) and the singular drift does not directly act on the position variables (as it is the case for the biased random walk), but indirectly *via* the momentum variables. The undesirable effects of the singularities are thus expected to be damped down.

Numerical results were performed by Anthony Scemama when he was a post-doc at CERMICS. These results, presented in Section 6.2, confirm these intuitions and demonstrate on a bench of representative examples that the algorithm based on the modified Langevin dynamics is the most efficient one of the algorithms studied here (the mathematical criteria for measuring the efficiency will be made precise below).

6.1 Description of the algorithms

6.1.1 Random walks in the configuration space

In this section, the state space is the configuration space $\mathbb{R}^{3N}$, so that the Metropolis-Hastings algorithm actually samples the probability density $\pi(x)$ (see Section 3.1.3 for a general presentation of the Metropolis-Hastings algorithm). Recall that the Metropolis-Hastings algorithm has a transition kernel given by

$$P(x, dy) = r(x, y)\mathcal{P}(x, y)\ dy + \left(1 - \int r(x, y')\mathcal{P}(x, y')\ dy'\right)\delta_x,$$

where the density $r(x, \cdot)$ is given by

$$r(x, y) = \min\left(1, \frac{\pi(y)\mathcal{P}(y, x)}{\pi(x)\mathcal{P}(x, y)}\right).$$

The function $\mathcal{P}$ is the proposal function. In words, the configuration y is proposed with probability $\mathcal{P}(x, y)$ from x, and accepted with probability $r(x, y)$, rejected otherwise.

Simple random walk

In the original paper [238] of Metropolis *et al.*, the Markov chain is a simple random walk:

$$\widetilde{x}^{n+1} = x^n + \delta\, U^n,$$

where δ is the step size and U^n are independent and identically distributed (i.i.d.) random vectors drawn uniformly in the $3N$-dimensional cube $K = [-1,1]^{3N}$. The corresponding transition density is

$$\mathcal{P}(x,y) = (2\delta)^{-3N}\, \chi_K\left(\frac{x-y}{\delta}\right),$$

where χ_K is the characteristic function of the cube K. Notice that in this particular case, $\mathcal{P}(x,y) = \mathcal{P}(y,x)$ so that the acceptance rate $r(x,y)$ only depends on the ratio $\pi(y)/\pi(x)$.

Biased random walk

The simple random walk is far from being the optimal choice: it induces a high rejection rate, hence a large variance. A variance reduction technique consists in considering the overdamped Langevin dynamics [58]:

$$dx_t = \nabla[\ln|\psi|](x_t)dt + dW_t, \tag{6.10}$$

where W_t is a $3N$-dimensional Wiener process. Note that $|\psi|^2$ is an invariant measure of the Markov process (6.10), and, better, that the dynamics (6.10) is in fact ergodic (see the results in Chapter 3) and satisfies a detailed balance property:

$$|\psi(x)|^2\, \mathcal{P}_{\Delta t}(x,y) \;=\; |\psi(y)|^2\, \mathcal{P}_{\Delta t}(y,x)$$

for any $\Delta t > 0$, where $\mathcal{P}_{\Delta t}(x,y)$ is the probability density that the Markov process (6.10) is at y at time $t + \Delta t$ starting from x at time t. These above results are classical for regular, positive functions ψ, and have been recently proven for fermionic wave functions [50] (in the latter case, the dynamics is ergodic in each nodal pocket of the wave function ψ).

Notice that if one uses the Markov chain of density $\mathcal{P}_{\Delta t}(x,y)$ in the Metropolis-Hastings algorithm, the acceptance/rejection step is useless, since (thanks to the detailed balance property) the acceptance rate always equals one. The exact value of $\mathcal{P}_{\Delta t}(x,y)$ is however unknown, so that a discretization of equation (6.10) with a simple Euler-Maruyama scheme is generally used

$$x^{n+1} = x^n + \Delta t\, \nabla[\ln|\psi|](x^n) + \Delta W^n \tag{6.11}$$

where ΔW^n are i.i.d. Gaussian random vectors with zero mean and covariance matrix $\Delta t\, I_{3N}$ (I_{3N} is the identity matrix). The Euler scheme leads to the approximated transition density

$$\mathcal{P}^{\text{Euler}}_{\Delta t}(x,y) = \frac{1}{(2\pi\Delta t)^{3N/2}}\, \exp\left(-\frac{|\; y - x - \Delta t\, \nabla[\ln|\psi|](x)\;|^2}{2\Delta t}\right).$$

The time discretization introduces the so-called *time-step error*, whose consequence is that (6.11) samples $d\pi$ *only approximately*. This error is however corrected by the Metropolis-Hastings acceptance/rejection procedure, which ensures that $d\pi$ is exactly sampled.

This sampling method is much more efficient than the Metropolis-Hastings algorithm based on the simple random walk, since the Markov chain (6.11) does a large part of the work (it samples a short time-step approximation of $d\pi$), which is clearly not the case for the simple random walk. The standard method in VMC computations currently is the Metropolis-Hastings algorithm based on the Markov chain defined by (6.11) (for refinements of this method, see [41, 332, 350]).

6.1.2 Random walks in the phase space

In this section, the state space is the phase space $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$. Let us emphasize that the introduction of momentum variables is nothing but a numerical artifice. The phase space trajectories that will be dealt with in this section do not have any physical meaning.

Langevin dynamics

We consider here the following Langevin dynamics of a system of N particles of mass m evolving in an external potential V:

$$\begin{cases} dx_t = \dfrac{p_t}{m}\, dt, \\ dp_t = -\nabla V(x_t)\, dt - \gamma p_t\, dt + \sigma dW_t. \end{cases} \tag{6.12}$$

The magnitudes σ and γ of the random forces σW_t and of the drag term $-\gamma p_t\, dt$ are related here through the fluctuation-dissipation formula

$$\sigma^2 = \frac{2m\gamma}{\beta}, \tag{6.13}$$

with $\beta = 1$ in the VMC framework. Since, for regular potentials, the canonical distribution

$$d\Pi(x,p) = Z^{-1} \exp\left[-\beta\left(V(x) + \frac{|p|^2}{2m}\right)\right] dx\, dp \tag{6.14}$$

is an invariant probability measure for the system (Z being a normalization constant), the projection on the position space of the Langevin dynamics samples $d\pi$. On the other hand, the Langevin dynamics does not satisfy the detailed balance property. We will come back to this important point in the forthcoming section.

In this context, the parameters m and γ (σ being then obtained through (6.13)) should be seen as numerical parameters to be optimized to get the best sampling. We now describe how to discretize and apply a Metropolis-Hastings algorithm to the Langevin dynamics (6.12), in the context of VMC.

Time discretization of the Langevin dynamics

Many discretization schemes exist for Langevin dynamics (see Section 3.2.4). In order to choose which algorithm is best for VMC, we have tested four different schemes available in the literature [4, 45, 183, 280], with parameters $\beta = 1$, $\gamma = 1$ and $m = 1$. The benchmark system is a Lithium atom, and ψ is a single determinantal wave function built upon Slater-type atomic orbitals, multiplied by a Jastrow factor[1]. We turn off the acceptance/rejection step in these preliminary tests, since our purpose is to compare the time-step errors for the various algorithms. From the results displayed in Table 6.1, one can see that the Ricci-Ciccotti algorithm [280] is the method which generates the smallest time-step error. This algorithm reads

$$\begin{cases} x^{n+1} = x^n + \Delta t \dfrac{p^n}{m} \mathrm{e}^{-\gamma \Delta t/2} + \dfrac{\Delta t}{2m}\left[-\nabla V(x^n)\Delta t + U^n\right] \mathrm{e}^{-\gamma \Delta t/4}, \\ p^{n+1} = p^n \mathrm{e}^{-\gamma \Delta t} - \dfrac{\Delta t}{2}\left[\nabla V(x^n) + \nabla V(x^{n+1})\right] \mathrm{e}^{-\gamma \Delta t/2} + U^n \mathrm{e}^{-\gamma \Delta t/2}, \end{cases} \tag{6.15}$$

[1] For all the numerical computations presented in this chapter, the interested reader should ask Anthony Scemama for details of the computations, in particular the values of the parameters for ψ given by (6.8).

where U^n are i.i.d. Gaussian random vectors with zero mean and variance $\sigma^2 I_{3N}$ with $\sigma^2 = \frac{2\gamma m}{\beta}\Delta t$. It can be seen from Table 6.1 that the Ricci-Ciccotti algorithm also outperforms the biased random walk (6.11), as far as sampling issues are concerned. In the following, we shall therefore use the Ricci-Ciccotti algorithm.

Table 6.1. Comparison of the energies computed with different discretization schemes for Langevin dynamics. The reference energy is -7.47198(4) a.u.

Δt	BRW	BBK [45]	Force interpolation [4]	Splitting [183]	Ricci & Ciccotti [280]
0.05	-7.3758(316)	-7.4395(246)	-7.4386(188)	-7.4467(137)	-7.4576(07)
0.005	-7.4644(069)	-7.4698(015)	-7.4723(015)	-7.4723(015)	-7.4701(20)
0.001	-7.4740(007)	-7.4728(013)	-7.4708(017)	-7.4708(017)	-7.4696(17)
0.0005	-7.4732(010)	-7.4700(023)	-7.4709(022)	-7.4708(022)	-7.4755(26)

Metropolized Langevin dynamics

The discretized Langevin dynamics does not exactly sample the target distribution Π, but rather some approximation $\Pi_{\Delta t}$ of Π. It is therefore tempting to introduce a Metropolis-Hastings acceptance/rejection step to further improve the quality of the sampling. Unfortunately, this idea cannot be straightforwardly implemented for two reasons:

(i) first, this is not technically feasible, since the Markov chain defined by (6.15) does not have a transition density. Indeed, as the same Gaussian random vectors U^n are used to update both the positions and the momenta, the conditional measure $\mathrm{p}((x^n, p^n), \cdot)$ is supported on a $3N$-dimensional submanifold of the phase space $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$;

(ii) second, leaving apart the above mentioned technical difficulty, which is specific to the Ricci-Ciccotti scheme, the Langevin dynamics is *a priori* not an efficient Markov chain for the Metropolis-Hastings algorithm because it does not satisfy the detailed balance property.

Let us now explain how to tackle these two issues, starting with the first one. To make it compatible with the Metropolis-Hastings framework, one needs to slightly modify the Ricci-Ciccotti algorithm. Following [4, 62] (see also the derivation in Section 3.2.4), we thus introduce i.i.d. *correlated* Gaussian vectors $(G^n_{1,i}, G^n_{2,i})$ $(1 \le i \le 3N)$ such that:

$$\begin{cases} \langle (G^n_{1,i})^2 \rangle &= \sigma_1^2 = \dfrac{\Delta t}{\beta m \gamma}\left(2 - \dfrac{3 - 4\mathrm{e}^{-\gamma\Delta t} + \mathrm{e}^{-2\gamma\Delta t}}{\gamma \Delta t}\right), \\ \langle (G^n_{2,i})^2 \rangle &= \sigma_2^2 = \dfrac{m}{\beta}\left(1 - \mathrm{e}^{-2\gamma\Delta t}\right), \\ \dfrac{\langle G^n_{1,i} G^n_{2,i} \rangle}{\sigma_1 \sigma_2} &= c_{12} = \dfrac{(1 - \mathrm{e}^{-\gamma\Delta t})^2}{\beta\gamma\sigma_1\sigma_2}. \end{cases}$$

Setting $G^n_1 = (G^n_{1,i})_{1\le i\le 3N}$ and $G^n_2 = (G^n_{2,i})_{1\le i\le 3N}$, the modified Ricci-Ciccotti algorithm reads

$$\begin{cases} x^{n+1} = x^n + \dfrac{\Delta t}{m} p^n \mathrm{e}^{-\gamma\Delta t/2} - \dfrac{\Delta t^2}{2m}\nabla V(x^n)\mathrm{e}^{-\gamma\Delta t/4} + G^n_1, \\ p^{n+1} = p^n \mathrm{e}^{-\gamma\Delta t} - \dfrac{\Delta t}{2}\left[\nabla V(x^n) + \nabla V(x^{n+1})\right]\mathrm{e}^{-\gamma\Delta t/2} + G^n_2. \end{cases} \tag{6.16}$$

The above scheme is a consistent discretization of (6.12) and the corresponding Markov chain does have a transition density, which reads (see Section 4.3.1 for example)

$$\mathcal{P}_{\Delta t}^{\mathrm{MRC}}((x^n,p^n),(x^{n+1},p^{n+1})) = Z^{-1}\exp\left[-\frac{1}{2(1-c_{12}^2)}\left(\left(\frac{|d_1|}{\sigma_1}\right)^2+\left(\frac{|d_2|}{\sigma_2}\right)^2-2c_{12}\frac{d_1}{\sigma_1}\cdot\frac{d_2}{\sigma_2}\right)\right], \tag{6.17}$$

with

$$d_1 = x^{n+1}-x^n-\Delta t\frac{p^n}{m}\mathrm{e}^{-\gamma\Delta t/2}+\frac{\Delta t^2}{2m}\nabla V(x^n)\mathrm{e}^{-\gamma\Delta t/4},$$

$$d_2 = p^{n+1}-p^n\mathrm{e}^{-\gamma\Delta t}+\frac{1}{2}\Delta t\left[\nabla V(x^n)+\nabla V(x^{n+1})\right]\mathrm{e}^{-\gamma\Delta t/2}.$$

Unfortunately, inserting directly the transition density (6.17) in the Metropolis-Hastings algorithm leads to a high rejection rate. Indeed, if (x^n,p^n) and (x^{n+1},p^{n+1}) are related through (6.16), $\mathcal{P}_{\Delta t}^{\mathrm{MRC}}((x^n,p^n),(x^{n+1},p^{n+1}))$ usually is much greater than $\mathcal{P}_{\Delta t}^{\mathrm{MRC}}((x^{n+1},p^{n+1}),(x^n,p^n))$, since the probability that the random forces are strong enough to make the particle go back in one step from where it comes, is very low in general. This is related to the fact that the Langevin dynamics does not satisfy the detailed balance relation.

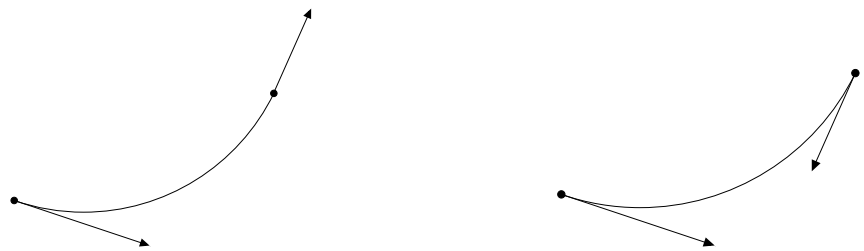


Fig. 6.1. Left: Usual Langevin dynamics; in this case, it is very unlikely to re-obtain the initial configuration starting from the final one. Right: Momentum reversal after integration time Δt; in this case, the dynamics is reversible.

It is however possible to further modify the overall algorithm by ensuring some microscopic reversibility, in order to finally obtain low rejection rates. For this purpose, we introduce momentum reversions. Such a procedure was already considered for Hybrid Monte Carlo algorithms (see for instance [2]). Denoting by $\mathcal{P}_{\Delta t}^{\mathrm{Langevin}}$ the transition density of the Markov chain obtained by integrating (6.12) *exactly* on the time interval $[t,t+\Delta t]$, it is indeed not difficult to check (under convenient assumptions on $V=-\ln|\psi|^2$), that the Markov chain defined by the transition density

$$\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{Langevin}}((x,p),(x',p')) = \mathcal{P}_{\Delta t}^{\mathrm{Langevin}}((x,p),(x',-p')) \tag{6.18}$$

is ergodic with respect to Π and satisfies the detailed balance property (see Figure 6.1)

$$\Pi(x,p)\,\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{Langevin}}((x,p),(x',p')) = \Pi(x',p')\,\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{Langevin}}((x',p'),(x,p)). \tag{6.19}$$

Replacing the exact transition density $\mathcal{P}_{\Delta t}^{\mathrm{Langevin}}$ by the approximation $\mathcal{P}_{\Delta t}^{\mathrm{MRC}}$, we now consider the transition density

$$\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{MRC}}((x,p),(x',p')) = \mathcal{P}_{\Delta t}^{\mathrm{MRC}}((x,p),(x',-p')). \tag{6.20}$$

These considerations are summarized in Algorithm 6.1. Note that a momentum reversion is systematically performed just after the Metropolis-Hastings step. As the invariant measure Π is left unchanged by this operation, the global algorithm (Metropolis-Hastings step based on the transition density $\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{MRC}}$ plus momentum reversion) actually samples Π. The role of the final momentum reversion is to preserve the underlying Langevin dynamics: while the proposals are accepted, the above algorithm generates Langevin trajectories, that are known to efficiently sample

an approximation of the target density Π. Numerical tests seem to show that, in addition, the momentum reversion also plays a role when the proposal is rejected: it seems to increase the acceptance rate of the next step, preventing the walkers from being trapped in the vicinity of the nodal surface $\psi^{-1}(0)$.

As the points (x^n, p^n) of the phase space generated by the above algorithm form a sampling of Π, the positions (x^n) sample $d\pi$ and can therefore be used for VMC calculations.

LANGEVIN METROPOLIZED VMC ALGORITHM

Algorithm 6.1. Starting from some initial configuration (x^0, p^0),

(1) Propose a move from (x^n, p^n) to $(\widetilde{x}^{n+1}, \widetilde{p}^{n+1})$ using the transition density $\widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{MRC}}$. In other words, perform one step of the modified Ricci-Ciccotti algorithm (6.16)

$$\begin{cases} x_*^{n+1} = x^n + \dfrac{\Delta t}{m} p^n \mathrm{e}^{-\gamma \Delta t/2} - \dfrac{\Delta t^2}{2m} \nabla V(x^n) + \mathrm{e}^{-\gamma \Delta t/4} + G_1^n, \\ p_*^{n+1} = p^n \mathrm{e}^{-\gamma \Delta t} - \dfrac{\Delta t}{2} \left[\nabla V(x^n) + \nabla V(x^{n+1}) \right] \mathrm{e}^{-\gamma \Delta t/2} + G_2^n, \end{cases}$$

and set $(\widetilde{x}^{n+1}, \widetilde{p}^{n+1}) = (x_*^{n+1}, -p_*^{n+1})$;

(2) Compute the acceptance rate

$$\alpha^n = \min \left(\frac{\Pi(\tilde{x}^{n+1}, \tilde{p}^{n+1})\, \widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{MRC}}((\widetilde{x}^{n+1}, \widetilde{p}^{n+1}), (x^n, p^n))}{\Pi(x^n, p^n)\, \widetilde{\mathcal{P}}_{\Delta t}^{\mathrm{MRC}}((x^n, p^n), (\widetilde{x}^{n+1}, \widetilde{p}^{n+1}))}, 1 \right);$$

(3) Draw a random variable $U^n \sim \mathcal{U}(0,1)$:
– if $U^n \leq \alpha^n$, accept the proposal and set $(\overline{x}^{n+1}, \overline{p}^{n+1}) = (\widetilde{x}^{n+1}, \widetilde{p}^{n+1})$;
– if $U^n > \alpha^n$, reject the proposal, and set $(\overline{x}^{n+1}, \overline{p}^{n+1}) = (x^n, p^n)$;

(4) Reverse the momenta: $(x^{n+1}, p^{n+1}) = (\overline{x}^{n+1}, -\overline{p}^{n+1})$.

A Hybrid Monte Carlo VMC algorithm

Generalized Hybrid Monte Carlo (HMC) algorithms could also be used (see Section 3.2.2 for more precisions on the HMC algorithm), relying in particular on the idea of using correlated momenta from one HMC step to the other [173]. For i.i.d. standard Gaussian random vectors G^n, the momenta may be updated as

$$p^{n+1} = \sqrt{1 - 2\gamma \Delta t}\, p^n + \sqrt{2\gamma \Delta t}\, G^n \simeq (1 - \gamma \Delta t)\, p^n + \sqrt{2\gamma \Delta t}\, G^n$$

when $\gamma \Delta \ll 1$. Therefore, using a very strong correlation from one step to another, and combining this momentum update in a HMC algorithm results in an approximation of Langevin dynamics. The interesting point in HMC algorithms is that the integration scheme to be used is a discretization of the Hamiltonian dynamics, and often the Störmer-Verlet algorithm is the most convenient scheme to use. Only some tuning of the parameters γ, m, Δt (and possibly the number of HMC steps before the acceptance/rejection step) has to be made.

6.2 Numerical experiments and applications

6.2.1 Measuring the efficiency

A major drawback of samplers based on Markov processes is that they generate sequentially correlated data. For a trajectory of L steps, the effective number of independent observations is

in fact $L_{\text{eff}} = L/N_{\text{corr}}$, where N_{corr} is the *correlation length*, namely the number of successive correlated moves. In the following applications, we provide estimators for the correlation length N_{corr} and for the so-called inefficiency η (see below), which are relevant indicators of the quality of the sampling. In this section, following Stedman *et al.* [322], we describe the way these quantities are defined and computed.

The sequence of samples is split into N_B blocks of L_B steps, where the number L_B is chosen such that it is a few orders of magnitude higher than N_{corr}. The mean energy is $\langle E_{\text{L}}\rangle_{|\psi|^2}$ and the variance is $\sigma^2 = \langle \big(E_{\text{L}} - \langle E_{\text{L}}\rangle_{|\psi|^2}\big)^2\rangle_{|\psi|^2}$. These quantities are defined independently on the VMC algorithm used. The empirical mean of the local energy reads

$$\langle E_L\rangle_{|\psi|^2}^{N_B,L_B} = \frac{1}{N_B L_B}\sum_{i=1}^{N_B L_B} E_L(x^i). \tag{6.21}$$

The empirical variance over all the individual steps is given by

$$[\sigma^{N_B,L_B}]^2 = \frac{1}{N_B L_B}\sum_{i=1}^{N_B L_B}\Big(E_L(x^i) - \langle E_L\rangle_{|\psi|^2}^{N_B,L_B}\Big)^2 \tag{6.22}$$

and the empirical variance over the blocks by

$$[\sigma_B^{N_B,L_B}]^2 = \frac{1}{N_B}\sum_{i=1}^{N_B}\Big(E_{B,i} - \langle E_L\rangle_{|\psi|^2}^{N_B,L_B}\Big)^2, \tag{6.23}$$

where $E_{B,i}$ is the average energy over block i:

$$E_{B,i} = \frac{1}{L_B}\sum_{j=(i-1)L_B+1}^{iL_B} E_L(x^j). \tag{6.24}$$

Following [322], we define the correlation length as

$$N_{\text{corr}} = \lim_{N_B\to\infty}\lim_{L_B\to\infty} L_B\frac{[\sigma_B^{N_B,L_B}]^2}{[\sigma^{N_B,L_B}]^2}, \tag{6.25}$$

and the inefficiency η of the run as:

$$\eta = \lim_{N_B\to\infty}\lim_{L_B\to\infty} L_B[\sigma_B^{N_B,L_B}]^2. \tag{6.26}$$

On the numerical examples presented below, the relative fluctuations of the quantities $L_B\frac{[\sigma_B^{N_B,L_B}]^2}{[\sigma^{N_B,L_B}]^2}$ and $L_B[\sigma_B^{N_B,L_B}]^2$ become small for $L_B > 50$ and $N_B > 50$.

The definition of these two quantities can be understood as follows. Since $L_B \gg N_{\text{corr}}$ and only L_B/N_{corr} are independent samples among the samples in the block, the central limit theorem yields

$$E_{B,i} \simeq \langle E_L\rangle_{|\psi|^2} + \frac{\sigma\, G^i}{\sqrt{L_B/N_{\text{corr}}}}$$

where G^i are i.i.d. normal random variables. Thus, in the limit $N_B \to \infty$ and $L_B \to \infty$, we obtain

$$(\sigma_B^{N_B,L_B})^2 = \frac{\sigma^2}{L_B/N_{\text{corr}}}.$$

Since $\lim_{N_B\to\infty}\lim_{L_B\to\infty}[\sigma^{N_B,L_B}]^2 = \sigma^2$, we obtain (6.25). The inefficiency η is thus equal to $N_{\text{corr}}\sigma^2$ and is large if the variance is large, or if the number of correlated steps is large.

Using this measure of efficiency, we can now compare the sampling algorithms (the simple random walk, the biased random walk and the Langevin algorithm) for various systems. In any case, a Metropolis-Hastings acceptance/rejection step is used. We found empirically from several tests that convenient values for the parameters of the Langevin algorithm are $\gamma = 1$ and $m = Z^{3/2}$ where Z is the highest nuclear charge among all the nuclei. For each algorithm, we compare the efficiency for various values of the step length, namely the increment δ in the case of the simple random walk, and the time-step Δt for the other two schemes. For a given algorithm, simple arguments corroborated by numerical tests show that there exists an optimal value of this increment: for smaller (resp. for larger increments), the correlation between two successive positions increases since the displacement of the particle is small (resp. since many moves are rejected), and this increases the number of correlated steps N_{corr}.

One can notice on the results (see tables 6.2, 6.3, 6.4, 6.5) that a large error bar corresponds to large values for N_{corr} and η. The quantities N_{corr} and η are a way to refine the measure of efficiency, since the same length of error bar may be obtained for different values of the numerical parameters.

6.2.2 Numerical results

Some numerical tests based on the above estimators of (in)efficiency are presented in this section. We compare the algorithms and parameters at a fixed computational cost. The reference values are obtained by ten times longer VMC simulations. The error bars given in parenthesis are 60% confidence intervals. We also provide the acceptance rate (denoted by A in the tables) and, when it is relevant, the mean of the length of the increment $x^{n+1} - x^n$ over one time-step (denoted by $\langle|\Delta x|\rangle$ in the tables) for the biased random walk and the Langevin dynamics. These tests were performed by Anthony Scemama using the `QMC=Chem` program[2].

Lithium.

The Lithium atom was chosen as a first simple example. The wave function is the same as for the benchmark system used for the comparison of the various Langevin schemes, namely a single Slater determinant of Slater-type basis functions improved by a Jastrow factor to take account of the electron correlation. The reference energy associated with this wave function is $-7.47198(4)$ a.u., and the comparison of the algorithms is given in Table 6.2. The runs were made of 100 random walks composed of 50 blocks of 1000 steps. For the simple random walk, the lowest values of the correlation length and of the inefficiency are respectively 11.4 and 1.40. The biased random walk is much more efficient, since the optimal correlation length and inefficiency are more than twice smaller, i.e. 4.74 and 0.55. The proposed algorithm is even more efficient: the optimal correlation length is 3.75 and the optimal inefficiency is 0.44.

Fluorine.

The Fluorine atom was chosen for its relatively "high" nuclear charge ($Z = 9$), leading to a timescale separation of the core and valence electrons. The wave function is a Slater-determinant with Gaussian-type basis functions where the $1s$ orbital was substituted by a Slater-type orbital, with a reference energy of $-99.397(2)$ a.u. The runs were made of 100 random walks composed of 100 blocks of 100 steps. The results are given in Table 6.3. For the simple random walk, the lowest values of the correlation length and of the inefficiency are respectively 15.6 and 282. The biased random walk, for which the optimal correlation length and inefficiency are 7.4 and 137, is again

[2] `Chem` is a Quantum Monte Carlo program written by M. Caffarel, IRSAMC, Université Paul Sabatier – CNRS, Toulouse, France. The wave functions are available upon request.

Table 6.2. The Lithium atom: Comparison of the Simple random walk, the Biased random walk and the proposed Langevin algorithm. The runs were carried out with 100 walkers, each realizing 50 blocks of 1000 steps. The reference energy is -7.47198(4) a.u., and A is the average acceptance rate.

ΔR	$\langle E_L \rangle$	N_{corr}	η	A	
Simple random walk					
0.05	-7.47126(183)	94.5 ± 3.3	11.72(42)	0.91	
0.10	-7.47239(97)	35.2 ± 1.2	4.08(14)	0.82	
0.15	-7.47189(75)	20.5(5)	2.30(06)	0.74	
0.20	-7.47157(56)	14.3(4)	1.62(04)	0.66	
0.25	-7.47182(56)	12.1(3)	1.40(05)	0.59	
0.30	-7.47189(56)	11.4(3)	1.57(17)	0.52	
0.35	-7.47275(59)	12.4(3)	1.57(17)	0.46	
0.40	-7.47130(63)	14.4(5)	1.93(22)	0.40	
Δt	$\langle E_L \rangle$	N_{corr}	η	$\langle \lvert\Delta x\rvert \rangle$	A
Biased random walk					
0.01	-7.47198(53)	10.31(29)	1.23(3)	0.284(09)	0.98
0.03	-7.47156(39)	5.26(14)	0.73(7)	0.444(21)	0.92
0.04	-7.47195(35)	4.82(12)	0.57(3)	0.486(26)	0.88
0.05	-7.47219(32)	4.74(11)	0.55(2)	0.514(31)	0.85
0.06	-7.47204(38)	4.95(11)	0.58(3)	0.533(36)	0.81
0.07	-7.47251(32)	5.39(14)	0.61(3)	0.546(40)	0.78
0.10	-7.47249(42)	7.56(25)	0.87(5)	0.555(50)	0.68
Langevin					
0.20	-7.47233(34)	5.07(10)	0.60(1)	0.236(08)	0.97
0.30	-7.47207(34)	4.14(09)	0.47(1)	0.328(15)	0.93
0.35	-7.47180(31)	3.96(08)	0.45(1)	0.366(18)	0.91
0.40	-7.47185(29)	3.75(08)	0.44(2)	0.399(22)	0.89
0.45	-7.47264(29)	3.88(08)	0.45(2)	0.426(25)	0.86
0.50	-7.47191(29)	4.07(14)	0.46(2)	0.426(25)	0.84
0.60	-7.47258(32)	4.78(16)	0.52(2)	0.481(36)	0.78

twice more efficient than the simple random walk. The Langevin algorithm is more efficient than the biased random walk: the optimal correlation length is 5.3 and the optimal inefficiency is 102.

Copper.

We can go even further in the timescale separation and take the Copper atom ($Z = 29$) as an example. The wave function is a Slater determinant with a basis of Slater-type atomic orbitals, improved by a Jastrow factor to take account of the electron correlation. The reference energy is $-1639.2539(24)$. The runs were made of 40 random walks composed of 500 blocks of 500 steps. From Table 6.4, one can remark that the Langevin algorithm is again more efficient than the biased random walk, since the optimal correlation length and inefficiency are respectively 28.7 and 4027, whereas using the biased random walk, these values are 51.0 and 5953.

The phenol molecule.

The Phenol molecule was chosen to test the proposed algorithm because it contains three different types of atoms (H, C and O). The wave function here is a single Slater determinant with Gaussian-type basis functions. The core molecular orbitals of the Oxygen and Carbon atoms were substituted by the corresponding atomic $1s$ orbitals. The comparison of the biased random walk with the Langevin algorithm is given in Table 6.5. The optimal correlation length using the biased

Table 6.3. The Fluorine atom : Comparison of the Simple random walk, the Biased random walk and the proposed Langevin algorithm. The runs were carried out with 100 walkers, each realizing 100 blocks of 100 steps. The reference energy is -99.397(2) a.u.

ΔR	$\langle E_L \rangle$	$N_{\rm corr}$	η	A	
Simple random walk					
0.02	-99.398(72)	38.9(7)	823(31)	0.87	
0.05	-99.426(39)	20.3(4)	405(11)	0.69	
0.08	-99.406(28)	15.6(4)	326(17)	0.53	
0.10	-99.437(23)	15.8(3)	282(07)	0.44	
0.12	-99.402(24)	16.6(4)	341(24)	0.36	
0.15	-99.398(25)	19.4(5)	412(41)	0.27	
Δt	$\langle E_L \rangle$	$N_{\rm corr}$	η	$\langle \lvert\Delta x\rvert \rangle$	A
Biased random walk					
0.002	-99.411(21)	9.9(2)	206(04)	0.211(08)	0.94
0.003	-99.424(17)	8.8(2)	173(04)	0.242(11)	0.90
0.004	-99.430(15)	7.6(2)	147(03)	0.263(16)	0.86
0.005	-99.399(14)	7.3(2)	142(03)	0.275(17)	0.82
0.006	-99.406(14)	7.4(1)	137(03)	0.282(19)	0.79
0.007	-99.430(14)	7.4(2)	142(08)	0.286(21)	0.75
0.008	-99.421(13)	7.6(2)	141(05)	0.287(23)	0.71
0.009	-99.406(13)	7.8(2)	177(19)	0.285(25)	0.67
0.010	-99.419(15)	7.8(2)	162(10)	0.281(27)	0.64
0.011	-99.416(14)	8.3(2)	147(05)	0.276(28)	0.60
0.012	-99.420(15)	9.1(3)	205(34)	0.270(29)	0.57
0.013	-99.425(17)	10.2(4)	224(38)	0.263(30)	0.54
Langevin					
0.10	-99.402(16)	8.9(2)	199(04)	0.095(02)	0.98
0.20	-99.403(12)	6.0(1)	123(02)	0.174(06)	0.94
0.25	-99.402(12)	5.4(1)	108(02)	0.204(09)	0.91
0.30	-99.395(11)	5.3(1)	104(02)	0.228(10)	0.87
0.35	-99.409(12)	5.4(1)	108(06)	0.245(15)	0.83
0.40	-99.402(11)	5.5(1)	102(03)	0.256(18)	0.78
0.45	-99.406(11)	5.9(1)	114(06)	0.261(21)	0.73
0.50	-99.408(12)	6.6(2)	124(07)	0.262(24)	0.68
0.55	-99.407(14)	7.9(4)	149(10)	0.257(26)	0.62
0.60	-99.405(15)	9.2(4)	178(13)	0.250(42)	0.56

random walk is 10.17, whereas it is 8.23 with our Langevin algorithm. The optimal inefficiency is again lower with the Langevin algorithm ($\eta = 544$) than with the biased random walk ($\eta = 653$).

6.2.3 Discussion of the results

In conclusion, the numerical tests show that the Langevin dynamics is always more efficient than the biased random walk. Indeed,

(i) The error bar (or $N_{\rm corr}$, or η) obtained with the Langevin dynamics for an optimal set of numerical parameters is always smaller than the error bar obtained with other algorithms (for which we also optimize the numerical parameters);

(ii) The size of the error bar does not seem to be as sensitive to the choice of the numerical parameters as for other methods. In particular, we observe on our numerical tests that the

Table 6.4. The Copper atom: Comparison of the Biased random walk with the proposed Langevin algorithm. The runs were carried out with 40 walkers, each realizing 500 blocks of 500 steps. The reference energy is -1639.2539(24) a.u.

Δt	$\langle E_L \rangle$	N_{corr}	η	$\langle \lvert \Delta x \rvert \rangle$	A
Biased random walk					
0.0003	-1639.2679(78)	79.1 ± 2.7	10682(420)	0.1311(108)	0.86
0.0004	-1639.2681(98)	70.4 ± 1.3	8682(204)	0.1385(137)	0.81
0.0005	-1639.2499(96)	61.3 ± 2.5	7770(297)	0.1414(162)	0.75
0.0006	-1639.2629(96)	56.0 ± 1.2	6834(88)	0.1414(183)	0.70
0.0007	-1639.2575(73)	53.8 ± 0.8	6420(81)	0.1393(201)	0.65
0.00075	-1639.2518(85)	53.1 ± 0.9	6330(91)	0.1377(209)	0.62
0.0008	-1639.2370(86)	55.7 ± 3.6	6612(405)	0.1357(216)	0.60
0.00105	-1639.2694(85)	51.0 ± 0.8	5953(90)	0.1228(241)	0.48
0.0011	-1639.2563(110)	54.3 ± 1.8	6513(221)	0.1198(245)	0.46
0.0012	-1639.2523(72)	59.9 ± 5.5	7266(658)	0.1136(251)	0.43
Langevin					
0.05	-1639.2553(92)	61.3 ± 1.7	8256(89)	0.0371(1)	0.99
0.10	-1639.2583(76)	40.6 ± 3.1	5319(383)	0.0705(30)	0.97
0.15	-1639.2496(65)	30.1 ± 0.8	4042(103)	0.0978(60)	0.93
0.20	-1639.2521(71)	28.7 ± 0.9	4027(403)	0.1173(96)	0.87
0.30	-1639.2510(67)	35.2 ± 2.5	4157(291)	0.1326(170)	0.71
0.40	-1639.2524(78)	50.5 ± 3.7	5922(455)	0.1210(225)	0.52

Table 6.5. The Phenol molecule : Comparison of the Biased random walk with the proposed Langevin algorithm. The runs were carried out with 100 walkers, each realizing 100 blocks of 100 steps. The reference energy is -305.647(2) a.u.

Δt	$\langle E_L \rangle$	N_{corr}	η	$\langle \lvert \Delta x \rvert \rangle$	A
Biased random walk					
0.003	-305.6308(83)	18.71(24)	1368(12)	0.522(29)	0.85
0.004	-305.6471(78)	16.00(28)	1193(30)	0.547(36)	0.80
0.005	-305.6457(65)	15.29(20)	1077(14)	0.555(43)	0.74
0.006	-305.6412(79)	15.00(17)	1018(11)	0.552(48)	0.69
0.007	-305.6391(67)	14.52(26)	1051(53)	0.540(52)	0.63
0.008	-305.6530(65)	14.72(19)	980(10)	0.523(56)	0.58
0.009	-305.6555(82)	15.28(28)	1272(163)	0.502(59)	0.54
Langevin					
0.05	-305.6417(101)	23.13(41)	1932(41)	0.126(02)	0.99
0.1	-305.6416(68)	13.97(22)	1189(23)	0.240(06)	0.97
0.2	-305.6496(57)	9.70(13)	812(12)	0.408(20)	0.89
0.3	-305.6493(56)	9.36(16)	817(36)	0.487(36)	0.78
0.4	-305.6473(58)	12.21(22)	834(20)	0.485(50)	0.61
0.5	-305.6497(80)	17.51(44)	1237(52)	0.425(58)	0.43

value $\Delta t = 0.2$ seems to be convenient to obtain good results with the Langevin dynamics, whatever the atom or molecule.

7

Second-order reduced density matrices

7.1 The electronic structure problem in terms of second order reduced density matrices **242**

7.1.1 The ensemble of N-repsentable second-order density matrices 242

7.1.2 The energy minimization problem in terms of second order reduced-density matrices 243

7.2 The N-representability problem **244**

7.2.1 Some necessary N-representability conditions for 2-RDMs 244

7.2.2 An explicit (counter)example 246

7.3 A dual formulation of the optimization problem **247**

7.3.1 Dual Formulation of the RDM Minimization Problem 247

7.3.2 Algorithm for solving the dual problem 248

7.3.3 Numerical results 250

As early as in 1951, it was noticed by Coleman that the electronic N-body ground-state energy could be obtained by minimizing over the set of N-representable two-body reduced density matrices (2-RDM), and Mayer definitely opened the field in 1955 with his pioneering article [232]. At a conference in 1959, Coulson then proposed to completely eliminate wavefunctions from Quantum Chemisty, since all the electronic ground-state properties of molecular systems can be computed from the 2-RDM [72,220,232]. Unfortunately, the set of N-representable 2-RDM is not known explicitly. Some mathematical characterizations were provided [70,71,197] but they could not be used to derive a numerical method with a complexity of a lower order than the usual N-body problem. In practice, only *approximate* RDM minimization problems, in which only a few necessary N-representability conditions are imposed (for example the so-called P,Q,G conditions [69,121]), can be considered. The first numerical studies relying on this strategy gave encouraging results [120].

Recently a new interest in the Reduced Density Matrix (RDM) approach arose. Very good numerical results have been obtained by two different strategies issued from semidefinite programming: primal-dual interior point methods [118, 233, 253, 376] on the one hand, augmented Lagrangian formulations using matrix factorizations of the 2-RDM [234–236] on the other hand. These results use a small number of known *necessary conditions* of N-representability. Yet, the so-obtained ground-state energies are as accurate as the ones obtained with coupled-cluster methods, see e.g. [234,235]. In addition, these energies provide lower bounds of the Full CI energies, whereas the variational post Hartree-Fock methods, such as CI or MCSCF, all provide upper bounds.

Since the RDM method is a linear minimization problem over a convex set of complicated structure, it is natural to use the concept of duality to mathematically characterize and numerically compute the minimum. Duality is an underlying issue in all the RDM studies [70,71,92,93,121,197], but surprisingly, the specific form of the dual formulation of the RDM problem has not yet been

used to derive an efficient algorithm. The current methods (see, e.g. [118, 234, 235, 253, 376]) all use general duality considerations in their algorithms, but none of them solves directly (and only) the dual RDM problem. As will be shown below, the associated dual optimization problem boils down to the search of the zero of a one-dimensional convex function.

This chapter is organized as follows. We first present the reformulation of the electronic problem in terms of 2-RDMs in Section 7.1, and recall the N-representability problem in Section 7.2. We then propose a dual formulation of the electronic problem in Section 7.3, and illustrate this approach with some numerical results.

7.1 The electronic structure problem in terms of second order reduced density matrices

7.1.1 The ensemble of N-repsentable second-order density matrices

We denote by $x = (\mathbf{x}, \sigma)$ the vector containing both the space variable $\mathbf{x} \in \mathbb{R}^3$ and the spin variable $\sigma \in \{|\uparrow\rangle, |\downarrow\rangle\}$. The summation on the spin variable will sometimes be denoted as an integral to simplify notations. For an antisymmetric N-body wavefunctions $\psi(x_1, ..., x_N) \in \bigwedge_{n=1}^{N} \mathfrak{h}$, the second-order reduced density matrix Γ is

$$\Gamma(x_1, x_2; y_1, y_2) = N(N-1) \int_{(\mathbb{R}^3 \times \{\pm 1\})^{N-2}} \overline{\psi}(x_1, x_2, x_3, \ldots, x_N) \psi(y_1, y_2, x_3, \ldots, x_N)\, dx_3 \ldots dx_N, \tag{7.1}$$

while the first-order reduced density matrix γ is

$$\begin{aligned} \gamma(x, y) &= \frac{1}{N-1} \int_{\mathbb{R}^3 \times \{\pm 1\}} \Gamma(x, z; y, z)\, dz \\ &= N \int_{(\mathbb{R}^3 \times \{\pm 1\})^{N-1}} \overline{\psi}(x, x_2, x_3, \ldots, x_N) \psi(y, x_2, x_3, \ldots, x_N)\, dx_2 \ldots dx_N. \end{aligned}$$

For a basis $(\phi_i)_{i \in \mathbb{N}^*}$ of the space $L^2(\mathbb{R}^3 \times \{|\uparrow\rangle, |\downarrow\rangle\}, \mathbb{C})$,

$$\Gamma(x_1, x_2; y_1, y_2) = \sum_{i_1, i_2, j_1, j_2 \in \mathbb{N}^*} \Gamma_{i_1, i_2}^{j_1, j_2} \overline{\phi}_{i_1}(x_1) \overline{\phi}_{i_2}(x_2) \phi_{j_1}(y_1) \phi_{j_2}(y_2), \quad \gamma(x, y) = \sum_{i,j} \gamma_i^j \overline{\phi}_i(x) \phi_j(y).$$

In the case of fermions, the matrix $\Gamma_{i_1, i_2}^{j_1, j_2}$ is antisymmetric, which means that $\Gamma_{i_1, i_2}^{j_1, j_2} = -\Gamma_{i_2, i_1}^{j_1, j_2} = \Gamma_{i_1, i_2}^{j_2, j_1}$. This ensures that $\Gamma(x_1, x_2; y_1, y_2) = -\Gamma(x_2, x_1; y_1, y_2)$ for instance.

For any vector space X, we denote by $\mathcal{S}(X)$ the space of self-adjoint matrices acting on X, and by $\mathcal{P}(X) \subset \mathcal{S}(X)$ the cone of positive semi-definite matrices. We also use the simplified notation $\mathcal{P}_N := \mathcal{P}\left(\bigwedge_1^N \mathfrak{h}\right)$ and $\mathcal{S}_N := \mathcal{S}\left(\bigwedge_1^N \mathfrak{h}\right)$. The cone of ensemble representable N-order density matrices is the convex envelope

$$\mathcal{P}_N = \left\{ \sum_{i=1}^{+\infty} n_i \,|\, \psi_i \,\rangle \langle\, \psi_i \,|\,, \quad \psi_i \in \bigwedge_{n=1}^{N} \mathfrak{h} \right\},$$

where $|\, \psi_i \,\rangle \langle\, \psi_i \,|$ is the projector onto span(ψ_i) :

$$|\, \psi_i \,\rangle \langle\, \psi_i \,|\, \psi \,\rangle = \left(\int_{(\mathbb{R}^3 \times \{\pm 1\})^N} \overline{\psi_i}(x) \psi(x)\, dx \right) \psi_i$$

Therefore, the cone of 2-RDM arising from an ensemble representable N-order density matrix is

$$\mathcal{C}_N = L_N^2(\mathcal{P}_N) \subset \mathcal{C}_2.$$

In this expression, the Kummer contraction operator L_N^2 [71,197] is the linear operator $|\,\psi\,\rangle\langle\,\psi\,| \mapsto \Gamma$ defined by (7.1). The corresponding $\Gamma \in \mathcal{C}_N$ are said to be *N-representable*. Of course the 2-RDMs of physical interest are the elements $\Gamma \in \mathcal{C}_N$ which arise from a normalized N-body density matrix $\Upsilon \in \mathcal{P}_N$ (satisfying $\mathrm{Tr}(\Upsilon) = 1$), so that $\Gamma = L_N^2(\Upsilon)$ satisfies $\mathrm{Tr}(\Gamma) = N(N-1)$.

7.1.2 The energy minimization problem in terms of second order reduced-density matrices

The electronic Hamiltonian H_N acting on the N-body fermionic space $\bigwedge_{n=1}^N \mathfrak{h}$ of antisymmetric N-body wavefunctions $\psi(x_1, ..., x_N)$ is formally defined as

$$H_N = \sum_{i=1}^N h_{x_i} + \sum_{1\le i<j\le N} \frac{1}{|\mathbf{x}_i - \mathbf{x}_j|},$$

where $h = -\Delta/2 + V$ and V is the external Coulomb potential generated by the nuclei. It holds

$$E = \inf_{\substack{\Psi\in\bigwedge_{n=1}^N \mathfrak{h},\\ \|\Psi\|=1}} \langle \Psi, H_N\Psi\rangle = \inf_{\substack{\Upsilon\in\mathcal{P}_N,\\ \mathrm{Tr}(\Upsilon)=1}} \mathrm{Tr}(H_N\Upsilon). \tag{7.2}$$

The second equality holds true since the minimum of a linear function over a convex set is attained on an extremal point of the convex set (on a point $\Gamma = |\,\psi_0\,\rangle\langle\,\psi_0\,|$, which is a rank 1 projector on $\mathrm{Span}\{\psi_0\}$)). The physical interpretation is that the infimum of the energy over the set of mixed states coincides with the infimum of the energy over the set of pure states.

Since the Hamiltonian H_N only contains two-body interactions, the energy of the system can be expressed in terms of the two-body density matrix Γ only (see, e.g. [71, 233]). By linearity, this property has to be shown only for extremal points $\Gamma_\psi = |\,\psi\,\rangle\langle\,\psi\,|$. Let us then show that

$$\left\langle \psi \,\middle|\, \hat{H} \,\middle|\, \psi \right\rangle = \mathrm{Tr}(K\Gamma),$$

where the two-body operator K is defined as

$$K = \frac{1}{2(N-1)}(h_{x_1} + h_{x_2}) + \frac{1}{2|x_1 - x_2|}.$$

It holds:

$$\begin{aligned}
\left\langle \psi \,\middle|\, \hat{H} \,\middle|\, \psi \right\rangle &= \sum_{i=1}^N \int_{(\mathbb{R}^3\times\{\pm1\})^N} \overline{\psi}((x_1,\sigma_1),\ldots,(x_N,\sigma_N))\,[h(\mathbf{x}_i)\cdot\psi((x_1,\sigma_1),\ldots,(x_N,\sigma_N))\,] \\
&\quad + \sum_{1\le i<j\le N} \int_{(\mathbb{R}^3\times\{\pm1\})^N} \frac{|\psi((x_1,\sigma_1),\ldots,(x_N,\sigma_N))|^2}{|\mathbf{x}_i - \mathbf{x}_j|}, \\
&= \sum_{\sigma_1\in\{\pm1\}} \int_{\mathbb{R}^3} h(x_1)\cdot\gamma((x_1,\sigma_1),(x_1',\sigma_1))|_{x_1'=x_1}\,dx_1 \\
&\quad + \frac{1}{2}\sum_{(\sigma_1,\sigma_2)\in\{\pm1\}^2} \int_{\mathbb{R}^6} \frac{\Gamma((x_1,\sigma_1),(x_2,\sigma_2)\,;\,(x_1,\sigma_1),(x_2,\sigma_2))}{|\mathbf{x}_1 - \mathbf{x}_2|}\,dx_1\,dx_2.
\end{aligned} \tag{7.3}$$

Therefore,

$$\boxed{E = \inf_{\substack{\Gamma \in \mathcal{C}_N, \\ \mathrm{Tr}(\Gamma)=N(N-1)}} \mathrm{Tr}(K\Gamma).} \tag{7.4}$$

Notice that we did not impose any constraint on the spin state in (7.4), but such constraints can be easily taken into account.

The Galerkin approximation

In practice, finite-dimensional spaces are used:

$$\mathfrak{h} := \mathrm{span}(\chi_i,\ i = 1, ..., r),$$

where $(\chi_i)_{i\geq 1}$ is a Hilbert basis of the one-body space $L^2(\mathbb{R}^3 \times \{|\uparrow\rangle, |\downarrow\rangle\}, \mathbb{C})$. The 2-RDM Γ associated with an N-body density matrix $\Upsilon \in \mathcal{P}_N$ is still defined by means of Kummer's contraction operator L_N^2 as

$$\Gamma_{i_1,i_2}^{j_1,j_2} = L_N^2(\Upsilon)_{i_1,i_2}^{j_1,j_2} = N(N-1) \sum_{k_3,\ldots,k_N=1}^{r} \Upsilon_{i_1 i_2 k_3 \ldots k_N}^{j_1 j_2 k_3 \ldots k_N}. \tag{7.5}$$

The 2-RDM Γ is now completely characterized by the matrix $(\Gamma_{i_1,i_2}^{j_1,j_2})_{i_1<i_2,\ j_1<j_2}$.

7.2 The N-representability problem

The electronic ground state problem reformulated as (7.4) is not tractable since the set $\mathcal{C}_N = \mathrm{L}_N^2(\mathcal{P}^N)$ over which the minimization is performed is unknown. This set is the set of 2-RDM obtained from a wavefunction (or an ensemble of wavefunctions) through the Kummer contraction. Characterizing this set is the so-called N-representability problem. No necessary and sufficient conditions of N-representability are known for 2-RDM (or higher order RDMs). This is in contrast with first-order reduced density matrices [69], which are N-representable as soon as $0 \leq \gamma \leq 1$ (as an operator) and $\mathrm{Tr}(\gamma) = N$.

Only necessary conditions are known for 2-RDM. The most famous ones are the so-called P, Q, G conditions [69, 121], and we will focus on them in the sequel. Additional conditions T_1 et T_2 [92] can also be considered. Imposing only this necessary conditions results in minimizing the energy on too large a variational space. Therefore, only lower bounds to the true energy are found this way.

7.2.1 Some necessary N-representability conditions for 2-RDMs

Origin of the P, Q, G conditions

An operator $\Gamma \in \mathcal{S}(\mathfrak{h} \wedge \mathfrak{h})$ is non-negative if and only if, for any $g \in \mathfrak{h} \wedge \mathfrak{h}$, $\langle g, \Gamma g\rangle \geq 0$. The P, Q, G conditions are obtained by requiring

$$\left\langle \psi \,\middle|\, A^\dagger A \,\middle|\, \psi \right\rangle \geq 0,$$

for certain operators A. In the formalism of second quantization (see [71] for more precisions), the P condition correspond to the positivity of the matrix $\left\langle \psi \,\middle|\, a_{i_1}^\dagger a_{i_2}^\dagger a_{j_1} a_{j_2} \,\middle|\, \psi \right\rangle$, the condition Q to the positivity of $\left\langle \psi \,\middle|\, a_{j_1} a_{j_2} a_{i_1}^\dagger a_{i_2}^\dagger \,\middle|\, \psi \right\rangle$, and G to the positivity of $\left\langle \psi \,\middle|\, a_{i_1}^\dagger a_{j_2} a_{i_2}^\dagger a_{j_1} \,\middle|\, \psi \right\rangle$.

Explicit formulation of the P,Q,G conditions

The P, Q, G conditions are linear equalities of the form

$$\mathcal{L}_P(\Gamma) \geq 0, \quad \mathcal{L}_Q(\Gamma) \geq 0, \quad \mathcal{L}_G(\Gamma) \geq 0.$$

The above operators are

$$\mathcal{L}_P(\Gamma) = \Gamma, \tag{7.6}$$

$$\mathcal{L}_Q(\Gamma)^{j_1,j_2}_{i_1,i_2} = \Gamma^{j_1,j_2}_{i_1,i_2} - \delta^{j_1}_{i_1}\gamma^{j_2}_{i_2} - \delta^{j_2}_{i_2}\gamma^{j_1}_{i_1} + \delta^{j_2}_{i_1}\gamma^{j_1}_{i_2} + \delta^{j_1}_{i_2}\gamma^{j_2}_{i_1} + (\delta^{j_1}_{i_1}\delta^{j_2}_{i_2} - \delta^{j_2}_{i_1}\delta^{j_1}_{i_2})\,\mathrm{Tr}(\Gamma), \tag{7.7}$$

$$\mathcal{L}_G(\Gamma)^{j_1,j_2}_{i_1,i_2} = -\Gamma^{j_1,i_2}_{i_1,j_2} + \delta^{j_1}_{i_1}\gamma^{j_2}_{i_2}. \tag{7.8}$$

The first order reduced density matrix is still obtained by means of the Kummer contraction

$$\gamma^j_i = \frac{1}{N-1}\sum_{k=1}^{N} \Gamma^{j,k}_{i,k}.$$

Notice that the operators $\mathcal{L}_P$ and $\mathcal{L}_Q$ defined on $\mathcal{S}(\mathfrak{h}\wedge\mathfrak{h})$ have values in $\mathcal{S}(\mathfrak{h}\wedge\mathfrak{h})$, so that $\mathcal{L}_P^* = \mathcal{L}_P$, $\mathcal{L}_Q^* = \mathcal{L}_Q$ (where the notation * refers to the adjoint operator). Therefore, the constraints $\mathcal{L}_P(\Gamma), \mathcal{L}_Q(\Gamma) \geq 0$ must be understood as

$$\forall B \in \mathcal{S}(\mathfrak{h}\wedge\mathfrak{h}), \quad \mathrm{Tr}(B\mathcal{L}_P(\Gamma)) \geq 0, \quad \mathrm{Tr}(B\mathcal{L}_Q(\Gamma)) \geq 0.$$

The operator $\mathcal{L}_G$ is also defined on $\mathcal{S}(\mathfrak{h}\wedge\mathfrak{h})$ but has values in a space larger than $\mathcal{S}(\mathfrak{h}\wedge\mathfrak{h})$, *a priori* the whole set $\mathcal{S}(\mathfrak{h}\otimes\mathfrak{h})$. Therefore, $\mathcal{L}_G(\Gamma) \geq 0$ means

$$\forall B \in \mathcal{S}(\mathfrak{h}\otimes\mathfrak{h}), \quad \mathrm{Tr}(B\mathcal{L}_G(\Gamma)) \geq 0.$$

Relationship with the N-representability of the first-order RDM

We verify here that the necessary N-representability conditions for the 2-RDM imply the N-representability of the first-order RDM. It is straightforward that the P condition ensures $\gamma \geq 0$. It then remains to check $\gamma \leq 1$ [69]. The proof we present here is suited for finite-dimensional spaces (which is the case of interest in practice), with r spatial basis functions ($2r$ basis functions when considering the spin variable).

Up to an orthogonal transformation, the first-order reduced density matrix can be chosen diagonal. It is then enough to show that $\gamma^i_i \leq 1$ for any $1 \leq i \leq 2r$. Since the diagonal elements of $\mathcal{L}_Q(\Gamma^N)$ are positive, if follows

$$\Gamma^{i_1,i_2}_{i_1,i_2} - \gamma^{i_1}_{i_1} - \gamma^{i_2}_{i_2} + 1 \geq 0.$$

Summing over $i_2 \neq i_1$ and dividing by $N-1$,

$$\frac{1}{N-1}\sum_{i_2\neq i_1} \Gamma^{i_1,i_2}_{i_1,i_2} - \frac{2r-1}{N-1}\gamma^{i_1}_{i_1} - \frac{1}{N-1}(\mathrm{Tr}(\gamma) - \gamma^{i_1}_{i_1}) + \frac{2r-1}{N-1} \geq 0,$$

since $\sum_{i_2\neq i_1}\gamma^{i_2}_{i_2} = \mathrm{Tr}(\gamma) - \gamma^{i_1}_{i_1}$. The first term of the above inequality being $\gamma^{i_1}_{i_1}$ (by contraction of the 2-RDM) and using $\mathrm{Tr}(\gamma) = N$, it finally holds

$$\gamma^{i_1}_{i_1}\left(1 - \frac{2r}{N-1}\right) + \frac{2r-1-N}{N-1} \geq 0,$$

so that, when $2r-1-N > 0$ (as in the case in practice), $\gamma^{i_1}_{i_1} \leq 1$.

7.2.2 An explicit (counter)example

The aim of this section is to show on an example that the set of N-representable 2-RDM has a very complicated topology. In particular, there exist N-representable 2-RDM that are no longer N-representable after an arbitrary small perturbation.

Consider $N = 3$ electrons, and an orthonoral system $(\phi_1, \ldots, \phi_5)$ in $L^2(\mathbb{R}^3)$. We denote Υ_ψ the density matrix of order $N = 3$ associated with the wavefunction ψ, and Γ_ψ the 2-RDM obtained from Υ_ψ through the Kummer operator $L \equiv L_3^2$. A basis of the 3-body space $\mathcal{H}^3 \subset \mathfrak{h}^3$ is given by the Slater determinants $\{\,|\,\phi_i\phi_j\phi_k\,\rangle\}_{1\le i<j<k\le 5}$, where

$$|\phi_i\phi_j\phi_k\rangle(x,y,z) = \frac{1}{\sqrt{3!}}\begin{vmatrix} \phi_i(x) & \phi_i(y) & \phi_i(z) \\ \phi_j(x) & \phi_j(y) & \phi_j(z) \\ \phi_k(x) & \phi_k(y) & \phi_k(z) \end{vmatrix}.$$

The space $\mathcal{H}_3$ is of dimension $\begin{pmatrix} 5 \\ 3 \end{pmatrix} = 10$. We will use in the sequel the short-hand notations

$$\psi_1 = |\phi_1\phi_2\phi_3\rangle, \quad \psi_2 = |\phi_1\phi_4\phi_5\rangle, \quad \psi_3 = |\phi_2\phi_4\phi_5\rangle, \quad \psi_4 = |\phi_3\phi_4\phi_5\rangle, \quad \psi_5 = |\phi_2\phi_3\phi_5\rangle.$$

The remaining basis functions $\psi_6, \cdots, \psi_{10}$ are chosen arbitrarily among the remaining Slater determinants, so that $\mathcal{B}^3 = (\Psi_1, ..., \Psi_{10})$ is a basis of $\mathcal{H}^3$. The space of 2-body functions $\mathcal{H}_2$ is also of dimension 10. A basis of this space is given by the Slater determinants $\{|\phi_i\phi_j\rangle\}_{1\le i<j\le 5}$, where for example

$$|\phi_1\phi_2\rangle(x,y) = \frac{1}{\sqrt{2}}\begin{vmatrix} \phi_1(x) & \phi_1(y) \\ \phi_2(x) & \phi_2(y) \end{vmatrix}.$$

This basis is ordered as

$$\mathcal{B}^2 := \{|\phi_1\,\phi_2\rangle, |\phi_1\,\phi_3\rangle, |\phi_1\,\phi_4\rangle, |\phi_1\,\phi_5\rangle, |\phi_2\,\phi_3\rangle, |\phi_2\,\phi_4\rangle, |\phi_2\,\phi_5\rangle, |\phi_3\,\phi_4\rangle, |\phi_3\,\phi_5\rangle, |\phi_4\,\phi_5\rangle\}.$$

Let us first compute the matrices τ_i associated with the 2-RDM Γ_{ψ_i} in the basis $\mathcal{B}^2$. For example,

$$L(\Upsilon_{\psi_1}) = \frac{1}{3}(|\phi_1\phi_2\rangle\langle\phi_1\phi_2| + |\phi_1\phi_3\rangle\langle\phi_1\phi_3| + |\phi_2\phi_3\rangle\langle\phi_2\phi_3|),$$

so that, in the ordered basis $\mathcal{B}^2$,

$$\tau_1 = \frac{1}{3}\mathrm{Diag}(1,1,0,0,1,0,0,0,0,0).$$

Analogously,

$$\tau_2 = \frac{1}{3}\mathrm{Diag}(0,0,1,1,0,0,0,0,0,1),$$

$$\tau_3 = \frac{1}{3}\mathrm{Diag}(0,0,0,0,0,1,1,0,0,1),$$

$$\tau_4 = \frac{1}{3}\mathrm{Diag}(0,0,0,0,0,0,0,1,1,1).$$

The 3-order density matrix

$$\Upsilon = \frac{1}{4}(\Gamma_{\Psi_1} + \Gamma_{\Psi_2} + \Gamma_{\Psi_3} + \Gamma_{\Psi_4}) \tag{7.9}$$

is therefore in $\mathcal{P}^3$ since it is a convex combination of elements of $\mathcal{P}^3$. The matrix τ associated with the corresponding 2-RDM is

$$\tau = \frac{1}{3}\mathrm{Diag}\left(\frac{1}{4}, \ldots, \frac{1}{4}, \frac{3}{4}\right).$$

The 2-RDM $\Gamma = L(\Upsilon)$ is then such that $\Gamma > 0$, and Υ defined by (7.9) is in fact the unique element in $\mathcal{B}^3$ such that $\Gamma = L(\Upsilon)$ (because L is one-to-one in the specific case we consider). Notice that Υ is non-negative but not positive, since its kernel is of dimension 6.

Consider now an arbitrary small perturbation of Γ of the form

$$\Gamma_\epsilon(x, y\,;\, x', y') = \Gamma(x, y\,;\, x', y') + \frac{\epsilon}{2}\left\{|\phi_1\phi_4\rangle(x, y)\,\langle\phi_2\phi_3|(x', y') + |\phi_2\phi_3\rangle(x, y)\,\langle\phi_1\phi_4|(x', y')\right\}$$

The matrix τ_ϵ corresponding to Γ_ϵ reads in the $\mathcal{B}^2$ basis

$$\tau_\epsilon = \tau + \frac{\epsilon}{2}(\delta_{3,5} + \delta_{5,3}).$$

Therefore, for ϵ small enough, the symmetric matrix τ_ϵ still verifies $\tau_\epsilon > 0$ and $\mathrm{tr}(\tau_\epsilon) = 3$. However, τ_ϵ is not 3-representable! Indeed, since L is one-to-one, τ_ϵ is obtained by contraction of

$$\begin{aligned}\Upsilon_\epsilon &= \Upsilon + \frac{\epsilon}{2}\{|\phi_1\phi_4\phi_5\rangle\langle\phi_2\phi_3\phi_5| + |\phi_2\phi_3\phi_5\rangle\langle\phi_1\phi_4\phi_5|\}\\ &= \Gamma + \frac{\epsilon}{2}\{|\Psi_5\rangle\langle\Psi_2| + |\Psi_2\rangle\langle\Psi_5|\}.\end{aligned}$$

In the basis $\{\psi_i\}_{i=1,\ldots,M}$, the matrix $\mathcal{T}_\epsilon$ corresponding to Υ_ϵ is

$$\mathcal{T}_\epsilon = \mathrm{Diag}\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0\right) + \frac{\epsilon}{2}(\delta_{2,5} + \delta_{5,2}),$$

which has a negative eigenvalue $-\epsilon$, so that the operator Γ_ϵ is not positive semi-definite.

7.3 A dual formulation of the optimization problem

7.3.1 Dual Formulation of the RDM Minimization Problem

We now present the dual formulation of the minimization (7.4). We recall that the polar cone $\mathcal{C}^*$ of a cone $\mathcal{C}$ in any Hermitian space is defined as $\mathcal{C}^* = \{x \mid \forall y \in \mathcal{C}, \langle x, y\rangle \geq 0\}$, where $\langle\cdot,\cdot\rangle$ denotes the considered scalar product (here, the Frobenius scalar product). The dual method then consists in formulating (7.4) in terms of $(\mathcal{C}_N)^*$ instead of $\mathcal{C}_N$:

$$\boxed{E = N(N-1)\sup\{\mu \mid K - \mu \in (\mathcal{C}_N)^*\}.} \tag{7.10}$$

Formula (7.10) can be easily derived from (7.4). Introducing the Lagrangian

$$\mathcal{L}(\Gamma, B, \mu) = \mathrm{Tr}(K\Gamma) - \mathrm{Tr}(B\Gamma) - \mu\{\mathrm{Tr}(\Gamma) - N(N-1)\},$$

it follows

$$E = \inf_{\Gamma\in\mathcal{S}_2}\ \sup_{B\in(\mathcal{C}_N)^*,\ \mu\in\mathbb{R}} \mathcal{L}(\Gamma, B, \mu). \tag{7.11}$$

As usual when using Lagrangian, the constraints are not stated explicitely, but penalized using some Lagrange parameter: μ is used to ensure that $\mathrm{Tr}(\Gamma) = N(N-1)$, and $B \in (\mathcal{C}_N)^*$ ensures that $\Gamma \in \mathcal{C}_N$. It then suffices to exchange the inf and the sup in (7.11) to obtain (7.10).

We therefore obtain an optimization problem in dimension 1 over $\mu \in \mathbb{R}$ which is the variable dual to the constraint $\mathrm{Tr}(\Gamma) = N(N-1)$. Of course characterizing the polar cone $(\mathcal{C}_N)^*$ is as difficult as characterizing $\mathcal{C}_N$, this issue is called the *N-representability problem*. Indeed $\mathcal{C}_N = (\mathcal{C}_N)^{**}$. Even if the dual formulation (7.10) does not simplify the theoretical N-representability problem, it turns out to be more convenient for numerical purposes.

Since both $(\mathcal{C}_N)^*$ and $\mathcal{C}_N$ are unknown and difficult to characterize, it is necessary to approximate (7.10) by a variational problem that can be carried out numerically. To this end, some necessary conditions for N-representability are selected. We consider L conditions of the following general form

$$\forall \ell = 1...L, \quad \mathcal{L}_\ell(\Gamma) \geq 0 \tag{7.12}$$

where for any ℓ, $\mathcal{L}_\ell : \mathcal{S}_2 \to \mathcal{S}(X_\ell)$ is a linear map and X_ℓ is some vector space. Here, we restrict ourselves to the P, Q, G conditions, with associated operators $\mathcal{L}_P$, $\mathcal{L}_P$ and $\mathcal{L}_G$ given respectively by (7.6), (7.7) an (7.8), and associated vector spaces $X_P = X_Q = \mathfrak{h} \wedge \mathfrak{h}$ and $X_G = \mathfrak{h} \otimes \mathfrak{h}$.

Imposing only the necessary conditions (7.12) means that $\mathcal{C}_N$ is replaced by the approximate cone $\mathcal{C}_{\text{app}} \supset \mathcal{C}_N$ defined as

$$\mathcal{C}_{\text{app}} := \{\Gamma \in \mathcal{S}_2 \mid \forall \ell = 1...L,\ \mathcal{L}_\ell(\Gamma) \geq 0\}.$$

Its polar cone can easily be shown to be

$$(\mathcal{C}_{\text{app}})^* := \left\{ \sum_{\ell=1}^{L} (\mathcal{L}_\ell)^* B_\ell \mid B_\ell \in \mathcal{S}(X_\ell),\ B_\ell \geq 0 \right\}, \tag{7.13}$$

and the associated approximate energy is then, in view of (7.10),

$$E_{\text{app}} = \inf_{\substack{\Gamma \in \mathcal{C}_{\text{app}}, \\ \text{Tr}(\Gamma) = N(N-1)}} \text{Tr}(K\Gamma) \tag{7.14}$$

$$= N(N-1) \sup\{\mu \mid K - \mu \in (\mathcal{C}_{\text{app}})^*\}. \tag{7.15}$$

Let us emphasize again that, since $\mathcal{C}_{\text{app}} \supset \mathcal{C}_N$, the energy E_{app} is a *lower bound* to the full CI energy in the chosen basis, $E_{\text{app}} \leq E$. We present in Section 7.3.2 an algorithm for solving problem (7.15). Notice that we obtain only the ground-state energy (and not the ground state density matrix), but, resorting to first order perturbation theory, any observable including at most two-body interaction terms can be obtained by a finite difference of energies.

7.3.2 Algorithm for solving the dual problem

Let us introduce the distance to the dual cone $(\mathcal{C}_{\text{app}})^*$

$$\delta(\mu) = \text{dist}\left(K - \mu, (\mathcal{C}_{\text{app}})^*\right).$$

Denoting $\mu^*_{\text{app}} = E_{\text{app}}/(N(N-1))$, the function δ satisfies the following properties:

(1) $\delta \equiv 0$ on $(-\infty, \mu^*_{\text{app}}]$ and is increasing on $[\mu^*_{\text{app}}, \infty)$;

(2) δ is convex on $\mathbb{R}$;

(3) δ^2 is continuously differentiable on $\mathbb{R}$, thus δ is continuously differentiable on $\mathbb{R} \setminus \{\mu^*_{\text{app}}\}$ and

$$\forall \mu > \mu^*_{\text{app}}, \quad \delta'(\mu) = -\frac{\text{Tr}(K - \mu - A_\mu)}{\|K - \mu - A_\mu\|} \tag{7.16}$$

where A_μ denotes the projection of $K - \mu$ onto the polar cone $(\mathcal{C}_{\text{app}})^*$.

Proofs for $(ii) - (iii)$ can be found in [249]. To prove (i), one notices that when $\mu \leq \mu^*_{\text{app}}$, $K - \mu = K - \mu^* + (\mu^* - \mu)$ belongs to $(\mathcal{C}_{\text{app}})^*$ since $\mu^* - \mu \in \mathcal{P}_2 \subset (\mathcal{C}_{\text{app}})^*$. To illustrate the above properties, we provide a plot of $\delta(\mu)$ for N_2 in a STO-6G basis set (see Figure 7.1).

In order to compute μ^*_{app}, we use a Newton-like scheme that strongly exploits the above mentioned properties in a natural way: starting from an initial energy above μ^*_{app} (such as the Hartree-Fock energy for instance) and using the convexity of the function δ, the Newton algorithm

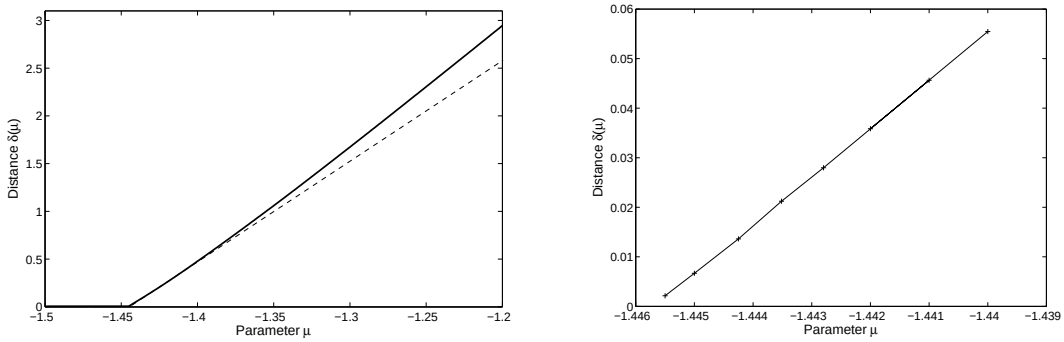


Fig. 7.1. Left: Distance $\delta(\mu)$ of $K - \mu$ to the cone $(\mathcal{C}_{\text{app}})^*$ as a function of μ for N_2 in a STO-6G basis set. The tangent at the estimated value for μ^*_{app} is also displayed (dashed line). Right: Zoom near the FCI reference value. The Hartree-Fock value is $\mu_{\text{HF}} = -1.4435153$ while the reference FCI value is $\mu_{\text{CI}} = -1.4453909$.

ensures that the energy μ decreases at each step of the optimization process and converges to μ^*_{app}. The right derivative of δ at μ^*_{app} being always positive, the convergence rate is guaranteed to be at least superlinear.

Of course, the most difficult part of the algorithm is the computation of the distance $\delta(\mu)$ to the cone, and of the projection A_μ of $K - \mu$. To this end, we chose to minimize, for a given μ, the objective function

$$J_\mu(B_1, \ldots, B_L) = \frac{1}{2} \left\| K - \mu - \sum_{\ell=1}^{L} (\mathcal{L}_\ell)^* B_\ell \right\|^2 ,$$

under the constraints $B_\ell \geq 0$ ($\ell = 1...L$), according to the definition (7.13) of the polar cone $(\mathcal{C}_{\text{app}})^*$. The above minimization is performed using a classical limited-memory BFGS algorithm [36], keeping the last $m = 3$ descent directions. The positivity constraints were parametrized by $B_\ell = (C_\ell)^2$ with C_ℓ symmetric, as suggested by Mazziotti in [234, 235].

Computing $\delta(\mu)$ with sufficient accuracy when μ is close to μ^*_{app} can be difficult because the minimization of $J_\mu(B)$ then is ill-conditioned. We therefore consider a "truncated" version of the Newton algorithm where μ is updated by a fraction $0 < a \leq 1$ of the Newton step. We then use the linearity of δ for values close to μ^*_{app} to devise a stopping criterion limiting the number of iterations. The algorithm is as follows:

Dual RDM optimization

Algorithm 7.1. Consider an initial value μ^0 (for example the Hartree-Fock value μ_{HF}), and $0 < a \leq 1$. Compute the projection A_{μ^0} of $K - \mu^0$ on $(\mathcal{C}_{\text{app}})^*$ and the distance $d^0 = \delta(\mu^0)$, and consider $\mu^1 = \mu^0 - \frac{\delta(\mu^0)}{\delta'(\mu^0)}$. For $n \geq 1$, and $\epsilon > 0$ small,

(1) Compute the projection $A_{\mu^n} = \sum_{\ell=1}^{L} (\mathcal{L}_\ell)^* \left[(C_\ell^n)^2 \right]$ of $K - \mu^n$ on $(\mathcal{C}_{\text{app}})^*$, the associated distance $d^n = \delta(\mu^n) = ||K - \mu^n - A_{\mu^n}||$ and the derivative $\delta'(\mu^n)$;

(2) Compute the interpolation slope $p^n = \frac{d^{n-1} - d^n}{\mu^{n-1} - \mu^n}$;

(3) If $p^n \leq (1 + \epsilon)\delta'(\mu^n)$, then the linear assumption is satisfied and the final value is extrapolated from the current position as $\mu^* = \mu^n - \frac{\delta(\mu^n)}{\delta'\mu^n)}$;

(4) Otherwise, set $\mu^{n+1} = \mu^n - a\frac{\delta(\mu^n)}{\delta'(\mu^n)}$ and start again from (1) using as initial guess $C_\ell^{n+1} = C_\ell^n$ for any $\ell = 1...L$.

In practice, the above algorithm converges in a few iterations. The only time consuming step is the projection performed in Step (1). As described above, this projection is done iteratively by minimizing the objective function J_μ by a limited-memory BFGS algorithm. The cost of one BFGS iteration scales as $O(r^6)$. We did not observe a clear scaling of the number of BFGS iterations with respect to the basis set size. The memory requirements scale as $O(r^4)$. Both computational time and memory requirements are comparable to those of [234].

7.3.3 Numerical results

We have tested the method on several molecules at equilibrium geometries using data from the EMSL Computational Results DataBase,[1] for STO-6G and 6-31G basis sets. The results are reported in Table 7.1 and 7.2 respectively.

Table 7.1. Correlation energies in a STO-6G basis set.

System	FCI energy	Correlation energy	Dual RDM energy (% of the correlation energy)
Be	-14.556086	-0.0527274	-14.556123 (100.07)
LiH	-7.972557	-0.0190867	-7.9727078 (100.79)
BH	-25.058806	-0.0569044	-25.061771 (105.21)
Li_2	-14.837571	-0.0286889	-14.839066 (105.21)
BeH_2	-15.759498	-0.0335151	-15.761284 (105.33)
H_2O	-75.735839	-0.0546392	-75.738582 (105.02)
NH_3	-56.0586005	-0.0693410	-56.074805 (123.37)

Table 7.2. Correlation energies in a 6-31G basis set.

System	FCI energy	Correlation energy	Dual RDM energy (% of the correlation energy)
Be	-14.613545	-0.0467812	-14.613653 (100.23)
LiH	-7.995678	-0.0185565	-7.9959693 (101.57)
BH	-25.171730	-0.0630461	-25.176736 (107.94)
Li_2	-14.893607	-0.0277581	-14.895389 (106.42)
BeH_2	-15.798440	-0.0402691	-15.801066 (106.52)
H_2O	-76.120220	-0.1401501	-76.142125 (115.63)
NH_3	-56.291315	-0.1336141	-56.318065 (120.02)

The reference Full CI (FCI) energies have been computed using GAMESS [300]. The correlation energies are recovered with a good accuracy. This is consistent with previous results already obtained with different RDM methods [118, 234, 235, 253, 376].

In general, we have observed that the function δ is almost linear in quite large a right neighborhood of μ^*_{app} (see Figure 7.1). Usually, only 3 or 4 Newton iterations are necessary to achieve convergence. Therefore, the only limiting step of the method is the computation of the distance $\delta(\mu)$ and of the projection A_μ of $K - \mu$ on the polar cone. The method is very robust with respect to initial choices of the energy μ^0 and the matrices C_k^0. However, we have observed that the computational time needed for finding the projection A_μ highly depends on the quality of the initial guess. The choice of genuine initial conditions is not obvious since we are manipulating abstract objects (dual elements of 2-RDM). Some CPU times are reported in Table 7.3 for very crude initial conditions $C_k^0 = \text{Id}$ and $\mu^0 \simeq 0.9\mu_{\text{HF}}$.

[1] See the web site `http://www.emsl.pnl.gov/proj/crdb/`

Table 7.3. CPU time (s) in a STO-6G basis using very crude initial guesses ($C_l = I$).

System	Spatial basis size r	CPU time (s)	Newton iterations
Be	5	25.7	2
LiH	6	240.9	3
H_2O	7	958.8	4
BeH_2	7	1143.3	3

We would like to underline that our projection algorithm is far from being optimal. There is clearly much room for improvement here. Let us also mention that the curve $\mu \mapsto \delta(\mu)$ can be easily sampled using parallel computing (one value of μ per processor).

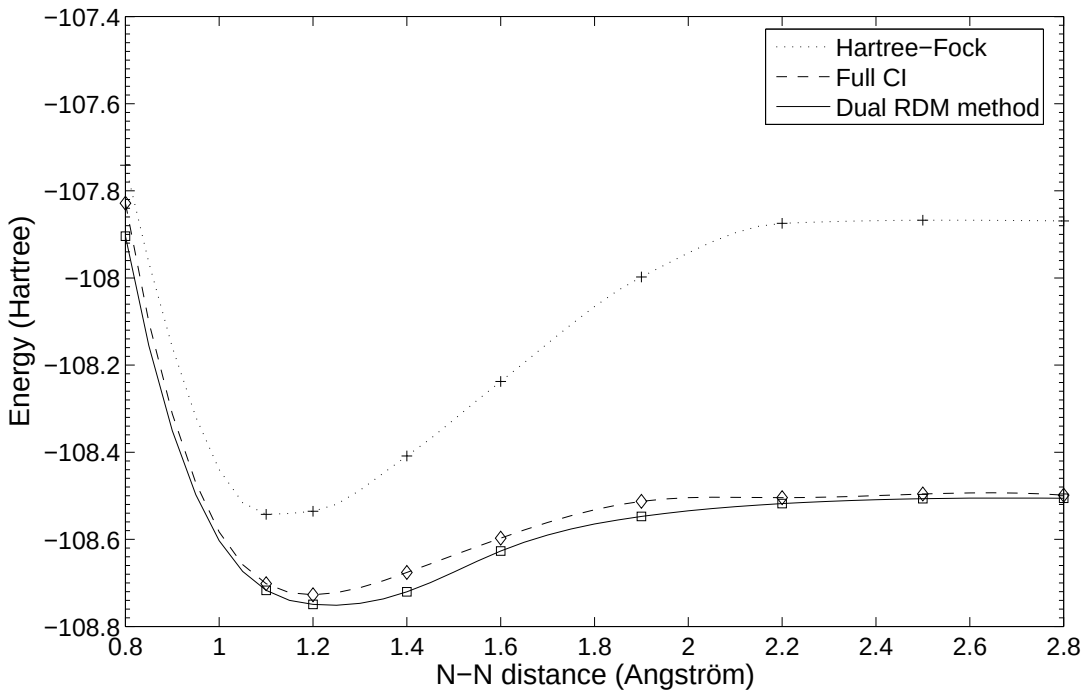


Fig. 7.2. Dissociation curve for N_2 in a STO-6G basis set.

We finally present in Figure 7.2 dissociation curves for N_2 in a STO-6G basis set. This example was already studied in several works [124,188,252]. The agreement of our results with the reference Full CI is excellent, and the dissociation energy is therefore recovered with a very good accuracy.

8

Local Exchange Potentials and Optimized Effective Potentials

8.1 The Slater exchange potential **255**
8.2 The Optimized Effective Potential problem **257**
8.2.1 Usual formulation of the OEP problem 257
8.2.2 A well-posed reformulation of the OEP problem 258
8.3 The effective local potential minimization problem **260**
8.4 Mathematical proofs **261**
8.4.1 Some useful preliminary results 261
8.4.2 Proofs for the Slater potential 262
8.4.3 Proof of Proposition 8.4 267

This chapter presents a work on progress with E. Cancès, E. Davidson, A. Izmaylov, G. Scuseria and V. Staroverov, on the mathematical understanding of the optimized effective potential (OEP) and other local potentials mathematically motivated by some minimization procedure. We seek here a local potential accounting for the exchange part of the electronic interactions (of course, eletronic correlations should ultimately be handled as well), and reproducing as accurately as possible the Hartree-Fock exchange, also called 'exact exchange' in the physics and chemistry literature.

The Hartree-Fock method, presented in Section 2.1.4, is a variational wavefunction method restricting the variational space to single Slater determinants:

$$\psi(x_1, \ldots, x_N) = \frac{1}{\sqrt{N!}} \operatorname{Det}(\phi_i(x_j)), \tag{8.1}$$

with $\phi_i \in \mathrm{H}^1(\mathbb{R}^3)$, $\int_{\mathbb{R}^3} \phi_i(x)\phi_j(x)\,dx = \delta_{ij}$. In the sequel,

$$\mathcal{X}_N = \left\{ \Phi = (\phi_i)_{1\le i\le N} \in (H^1(\mathbb{R}^3))^N \;\middle|\; \int_{\mathbb{R}^3} \phi_i\phi_j = \delta_{ij} \right\}.$$

The Hartree-Fock energy functional of a system of N spin-less electrons reads

$$E^{\mathrm{HF}}(\Phi) = \frac{1}{2}\sum_{i=1}^{N}\int_{\mathbb{R}^3} |\nabla\phi_i|^2 + \int_{\mathbb{R}^3} V_{\mathrm{nuc}}\rho_\Phi + \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{\rho_\Phi(x)\rho_\Phi(y)}{|x-y|}\,dx\,dy - \frac{1}{2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\,dx\,dy, \tag{8.2}$$

where the density ρ_Φ and the density matrix γ_Φ are defined respectively by

$$\rho_\Phi(x) = \sum_{i=1}^{N} |\phi_i(x)|^2, \quad \gamma_\Phi(x,y) = \sum_{i=1}^{N} \phi_i(x)\phi_i(y). \tag{8.3}$$

The potential created by the nuclei is, for a molecule with K atoms of charge z_k at positions $\bar{x}_k$,

$$V_{\mathrm{nuc}}(x) = -\sum_{k=1}^{K} \frac{z_k}{|x - \bar{x}_k|}.$$

For simplicity, we will consider in the sequel the Coulombic atomic potential

$$V_{\mathrm{nuc}}(x) = -\frac{Z}{|x|}$$

for $Z \geq 0$. A minimizer of (8.2) satisfies the Hatree-Fock equations, which are the Euler-Lagrange equations associated with (8.2) (up to a unitary transformation):

$$\mathcal{F}_{\Phi}\phi_i = -\frac{1}{2}\Delta\phi_i + V_{\mathrm{nuc}}\phi_i + \left(\rho_{\Phi} \star \frac{1}{|x|}\right)\phi_i + K_{\Phi}\phi_i = \epsilon_i\phi_i. \tag{8.4}$$

In this expression, the exchange operator K_{Φ} is defined as

$$K_{\Phi}\varphi(x) = -\int_{\mathbb{R}^3} \frac{\gamma_{\Phi}(x,y)}{|x-y|}\varphi(y)\,dy. \tag{8.5}$$

It is therefore a non-local operator depending on the orbitals $\Phi = \{\phi_i\}_{i=1,\ldots,N}$.

Mathematical setting

We consider here a given N-tuple $\Phi = \{\phi_i\}_{1\leq i\leq N}$ of functions defined on $\mathbb{R}^3$, orthogonal for the $\mathrm{L}^2(\mathbb{R}^3)$ inner product and belonging to the Sobolev space $\mathrm{H}^2(\mathbb{R}^3)$ (notice that the latter two conditions are automatically satisfied for any solution of the Hartree-Fock or Kohn-Sham equations). The corresponding density and density matrix are defined as in (8.3). As the $\{\phi_i\}_{1\leq i\leq N}$ are assumed to be in $\mathrm{H}^2(\mathbb{R}^3)$, it follows from Sobolev embedding theorems that the density ρ_{Φ} is a continuous function going to zero at infinity. We also assume that ρ_{Φ} does not vanish on $\mathbb{R}^3$ (this condition is automatically satisfied if the $\{\phi_i\}_{1\leq i\leq N}$ are the lowest N eigenfunctions of a Kohn-Sham operator).

The exchange operator (8.5) associated with the N-tuple $\{\phi_i\}_{1\leq i\leq N}$ is the Hilbert-Schmidt operator on $\mathrm{L}^2(\mathbb{R}^3)$ defined for all $\varphi \in \mathrm{L}^2(\mathbb{R}^3)$ as

$$(K_{\Phi}\varphi)(x) = -\int_{\mathbb{R}^3} \frac{\gamma_{\Phi}(x,y)}{|x-y|}\varphi(y)\,dy.$$

Note that the right hand side of the above definition actually makes sense as a $\mathrm{L}^2(\mathbb{R}^3)$ function. This is a consequence of Cauchy-Schwarz and Hardy inequalities (for the Hardy inequality, see *e.g.* [52, Theorem 2.12]), since, for fixed $x \in \mathbb{R}^3$,

$$\begin{aligned}\left|\int_{\mathbb{R}^3} \frac{\gamma_{\Phi}(x,y)}{|x-y|}\varphi(y)\,dy\right| &\leq \sum_{i=1}^{N} |\phi_i(x)|\,\|\varphi\|_{\mathrm{L}^2(\mathbb{R}^3)} \left\|\frac{\phi_i}{|\cdot - x|}\right\|_{\mathrm{L}^2(\mathbb{R}^3)} \\ &\leq 2\sum_{i=1}^{N} |\phi_i(x)|\,\|\varphi\|_{\mathrm{L}^2(\mathbb{R}^3)}\|\nabla\phi_i\|_{\mathrm{L}^2(\mathbb{R}^3)}.\end{aligned} \tag{8.6}$$

Recall that a Hilbert-Schmidt operator on $\mathrm{L}^2(\mathbb{R}^3)$ is a linear operator on $\mathrm{L}^2(\mathbb{R}^3)$ for which there exists $g \in \mathrm{L}^2(\mathbb{R}^3 \times \mathbb{R}^3)$ such that

$$\forall f \in \mathrm{L}^2(\mathbb{R}^3), \quad (Gf)(x) = \int_{\mathbb{R}^3} g(x,y)f(y)\,dy.$$

The function g (which is unique) is called the kernel of G. The set of Hilbert-Schmidt operators on $\mathrm{L}^2(\mathbb{R}^3)$ is denoted by $\sigma_2(\mathrm{L}^2(\mathbb{R}^3))$. Endoved with the inner product defined by

$$\langle G, H\rangle_{\mathrm{HS}} = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} g(x,y)h(x,y)\,dx\,dy$$

(where g and h are the kernels of G and H respectively), $\sigma_2(\mathrm{L}^2(\mathbb{R}^3))$ is a Hilbert space. The corresponding norm is thus defined by

$$\|G\|_{\mathrm{HS}} = \left(\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} |g(x,y)|^2\,dx\,dy\right)^{1/2}.$$

Here, the kernel k_Φ of K_Φ reads

$$k_\Phi(x,y) = -\frac{\gamma_\Phi(x,y)}{|x-y|},$$

and, making use once again of Cauchy-Schwarz and Hardy inequalities,

$$\begin{aligned}\|K_\Phi\|_{\mathrm{HS}}^2 = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|^2}\,dx\,dy &\le \sum_{i=1}^N \int_{\mathbb{R}^3} |\phi_i(x)|^2\,dx \sum_{j=1}^N \left\|\frac{\phi_j}{|\cdot - x|}\right\|_{\mathrm{L}^2(\mathbb{R}^3)}^2 \\ &\le 4N\sum_{j=1}^N \|\nabla\phi_j\|_{\mathrm{L}^2(\mathbb{R}^3)}^2 < +\infty.\end{aligned}$$

The one-body density matrix γ_Φ is also the kernel of a Hilbert-Schmidt operator on $\mathrm{L}^2(\mathbb{R}^3)$, denoted by γ_Φ (abusing notations) and defined as

$$\forall f \in \mathrm{L}^2(\mathbb{R}^3), \quad (\gamma_\Phi f)(x) = \int_{\mathbb{R}^3} \gamma_\Phi(x,y)f(y)\,dy = \sum_{i=1}^N \phi_i(x)\int_{\mathbb{R}^3}\phi_i(y)f(y)\,dy.$$

8.1 The Slater exchange potential

The exchange operator (8.5) is not a local operator (see Section 8.2.1 for a tentative definition of local operators). In order to reduce the complexity of the Hartree-Fock equations, Slater proposed to replace the non-local exchange operator by some local operator [312]. This local operator is obtained by some averaging procedure (but can also be defined in terms of some variational procedure, see Remark 8.3), and can be expressed in terms of the density matrix of the system as

$$v_{\mathrm{x},S}^{\Phi}(x) = -\frac{1}{\rho_\Phi(x)}\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\,dy. \tag{8.7}$$

Nowadays, the complexity of the Hartree-Fock equations is no more an obstacle for ground-state computations. However, it is still very interesting to find approximate local exchange operators for the purpose of interpretation, or to improve the exchange part of exchange-correlation functionals in Density Functional Theory. The local exchange potentials can also be used as an input in other approaches, especially time-dependent methods.

The existence of a radial solution to the self-consistent Kohn-Sham equations with the Slater exchange potential as an exchange-correlation potential is given by the following theorem. Recall that a function ϕ is said to be radial if there exists a function φ such that $\phi(x) = \varphi(|x|)$. We will denote by $L_r^2(\mathbb{R}^3)$ (resp. $H_r^1(\mathbb{R}^3)$) the set of radial $L^2(\mathbb{R}^3)$ (resp. radial $H^1(\mathbb{R}^3)$) functions, and set

$$\mathcal{X}_N^r = \left\{\Phi = (\phi_i)_{1\le i\le N} \in (H_r^1(\mathbb{R}^3))^N \;\middle|\; \int_{\mathbb{R}^3}\phi_i\phi_j = \delta_{ij}\right\}.$$

Theorem 8.1. *In the case of a single nucleus of charge $Z \geq N$, the nonlinear eigenvalue problem*

$$1 \leq i \leq N, \qquad \left(-\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_\Phi \star \frac{1}{|x|} - \frac{1}{\rho_\Phi(x)} \int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\, dy\right) \phi_i = \epsilon_i \phi_i, \tag{8.8}$$

with $\epsilon_1 < \cdots \leq \epsilon_N \leq 0$ and ρ_Φ, γ_Φ defined as in (8.3), has a solution[1] $\Phi = (\phi_i) \in \mathcal{X}_N^r$ and the corresponding exchange potential $v_{\mathrm{x},S}^\Phi$ is globally Lipschitz in $\mathbb{R}^3$, C^∞ away from the nucleus, and satisfies, for all $\eta > 0$,

$$v_{\mathrm{x},S}^\Phi(x) = -\frac{1}{|x|} + \mathrm{o}\left(\mathrm{e}^{-(2\sqrt{-2\epsilon_N}-\eta)|x|}\right).$$

Besides, the minimum of the Hartree-Fock energy over the set of the radial solutions to (8.8) is attained.

The proof of Theorem 8.1 can be read in Section 8.4.

Remark 8.1 (Practical computation through an iterative procedure). *To compute in practice a solution (8.8), it is possible to consider the following iterative procedure:*

Algorithm 8.1. *Starting from some set of N orbitals $\Phi^0 = \{\phi_1^0, \ldots, \phi_N^0\}$,*

(1) compute the local Slater exchange potential $v_{\mathrm{x},S}^{\Phi^n}$ given by (8.7) using the orbitals $\Phi^n = \{\phi_i^n\}_{i=1,\ldots,N}$;

(2) compute the first N eigenvectors of the operator

$$\left(-\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^n} \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Phi^n}\right) \phi_i^{n+1} = \epsilon_i^{n+1} \phi_i^{n+1}. \tag{8.9}$$

When there are degeneracies in the highest energy levels, some arbitrary choice is made;

(3) replace n by $n+1$ and go back to Step 1.

In some case, we will restrict ourselves to radial eigenvectors. Recall that, when the orbitals are radial, the eigenvalues of the operators appearing in Algorithm 8.1 are non-degenerate, and the radial i-th eigenvector ϕ_i has exactly $i-1$ nodal spheres.

The well-posedness of this iterative procedure is ensured provided the operator in (8.9) has at least N negative eigenvalues, its essential spectrum still being $[0, +\infty)$. This is easier to check when the orbitals are radial, or when the nuclear charge satisfies $Z > N$. In the general case, some exponential decay of the initial orbitals has to be assumed. The well-posedness of the iterative procedure is precised in the following propositions:

Proposition 8.1. *Assume $Z > N-1$. For initial radial orbitals $(\phi_1^0, \ldots, \phi_N^0) \in [\mathrm{H}^2(\mathbb{R}^3)]^N$, and when $(\phi_1^{n+1}, \ldots, \phi_N^{n+1})$ are the first N radial orbitals in the diagonalization (8.9), the iterative procedure of Algorithm 8.1 is well-defined.*

Proposition 8.2. *Assume $Z > N$. For initial orbitals $(\phi_1^0, \ldots, \phi_N^0) \in [\mathrm{H}^2(\mathbb{R}^3)]^N$, the iterative procedure of Algorithm 8.1 is well-defined.*

Proposition 8.3. *Assume $Z = N$. For initial orbitals $(\phi_1^0, \ldots, \phi_N^0) \in [\mathrm{H}^2(\mathbb{R}^3)]^N$ exponentially decreasing, i.e. such that there exists $C^0, \gamma^0, R^0 > 0$ with*

$$\forall 1 \leq i \leq N, \quad \forall |x| \geq R^0, \qquad |\phi_i^0(x)| \leq C^0 \mathrm{e}^{-\gamma^0|x|},$$

the iterative procedure of Algorithm 8.1 is well-defined.

However, we were not able to show that this numerical procedure indeed converges to a solution of the self-consistent Kohn-Sham equations with Slater exchange potential.

[1] In the *Aufbau* condition ($\epsilon_1 \leq \cdots \leq \epsilon_N$ are the lowest N eigenvalues of $\left(-\frac{1}{2}\Delta + V_{\mathrm{nuc}} + \rho_\Phi \star \frac{1}{|x|} + v_{\mathrm{x},S}^\Phi\right)$), the mean-field Hamiltonian is here considered as an operator on $L_r^2(\mathbb{R}^3)$.

8.2 The Optimized Effective Potential problem

8.2.1 Usual formulation of the OEP problem

In order to generalize and improve Slater's approach, Sharp and Horton [308] proposed a systematic way to obtain local potentials approximating the non local exchange operator. They suggest to minimize the energy of the Slater determinant constructed with the lowest N eigenfunctions of some one-electron Schrödinger operator $-\frac{1}{2}\Delta + W$, W being a 'local potential'. This track was further explored by Talman and Shadwick [338]. Note that what is precisely meant by 'local potential' is not clear.

Leaving this issue aside until next section, we introduce the set of admissible 'local potentials'

$$\mathcal{W} = \left\{ W \text{ 'local potential'} \;\middle|\; H_W = -\frac{1}{2}\Delta + W \text{ is a self-adjoint operator on } L^2(\mathbb{R}^3), \right.$$
$$\left. \text{bounded from below, with at least } N \text{ eigenvalues below its essential spectrum} \right\},$$

and the OEP minimization set

$$\mathcal{X} = \left\{ \Phi = \{\phi_i\}_{1\le i\le N} \;\middle|\; \phi_i \in H^1(\mathbb{R}^3),\ (8.11) \text{ and } (8.12) \text{ hold for some } W \in \mathcal{W} \right\}, \tag{8.10}$$

where conditions (8.11) and (8.12) are defined as

$$\left(-\frac{1}{2}\Delta + W\right)\phi_i = \epsilon_i \phi_i, \qquad \int_{\mathbb{R}^3} \phi_i \phi_j = \delta_{ij}, \tag{8.11}$$

and

$$\epsilon_1 \le \cdots \le \epsilon_N \text{ are the lowest } N \text{ eigenvalues of } H_W = -\frac{1}{2}\Delta + W. \tag{8.12}$$

The optimized effective potential problem then reads

$$\inf_{\Phi\in\mathcal{X}} E^{\mathrm{HF}}(\Phi). \tag{8.13}$$

Denoting by Φ^{OEP} a minimizer to (8.13), an optimal effective potential is a 'local potential' $W^{\mathrm{OEP}} \in \mathcal{W}$ which allows to generate Φ^{OEP} through (8.11)-(8.12). It is convenient to decompose W^{OEP} as

$$W^{\mathrm{OEP}}(x) = V_{\mathrm{nuc}}(x) + \int_{\mathbb{R}^3} \frac{\rho_{\Phi^{\mathrm{OEP}}}(y)}{|x-y|}\, dy + v_{\mathrm{x,OEP}}(x).$$

In order to emphasize the mathematical issues arising from the above formulation of the OEP problem, it is worth recalling the general method for proving existence of solutions to a minimization problem such as (8.13). The first step consists in considering a so-called minimizing sequence, that is a sequence $(\Phi^n)_{n\in\mathbb{N}}$ of elements of $\mathcal{X}$ such that

$$\lim_{n\to+\infty} E^{\mathrm{HF}}(\Phi^n) = \inf_{\Phi\in\mathcal{X}} E^{\mathrm{HF}}(\Phi).$$

It is easy to check that the sequence $(\Phi^n)_{n\in\mathbb{N}}$ is bounded in $(H^1(\mathbb{R}^3))^N$, hence weakly converges, up to extraction, toward some $\Phi^\infty \in (H^1(\mathbb{R}^3))^N$. It is then standard to show (see [211] for instance) that

$$E^{\mathrm{HF}}(\Phi^\infty) \le \inf_{\Phi\in\mathcal{X}} E^{\mathrm{HF}}(\Phi). \tag{8.14}$$

The difficult step of the proof is to show that $\Phi^\infty \in \mathcal{X}$ (if $\Phi^\infty \in \mathcal{X}$, we can immediately conclude, using (8.14), that Φ^∞ is a solution to (8.13)). For this purpose, we need to introduce a sequence $(W_n)_{n\in\mathbb{N}}$ of 'local potentials' such that Φ^n can be generated by W_n *via* (8.11)-(8.12). If $(W_n)_{n\in\mathbb{N}}$

was bounded in some convenient functional space $\mathcal{Y}$, $(W_n)_{n\in\mathbb{N}}$ would converge (up to extraction and in some weak sense) to some potential $W_\infty \in \mathcal{Y}$. We could then try to pass to the limit in the system

$$\begin{cases} -\frac{1}{2}\Delta\phi_i^n + W_n\phi_i^n = \epsilon_i^n\phi_i^n, \\ \int_{\mathbb{R}^3} \phi_i^n\phi_j^n = \delta_{ij}, \\ \epsilon_1^n \leq \cdots \leq \epsilon_N^n \text{ are the lowest } N \text{ eigenvalues of } H_{W_n} = -\frac{1}{2}\Delta + W_n, \end{cases}$$

using more or less sophisticated functional analysis arguments, in order to prove that Φ^∞ satisfies

$$\begin{cases} -\frac{1}{2}\Delta\phi_i^\infty + W_\infty\phi_i^\infty = \epsilon_i^\infty\phi_i^\infty, \\ \int_{\mathbb{R}^3} \phi_i^\infty\phi_j^\infty = \delta_{ij}, \\ \epsilon_1^\infty \leq \cdots \leq \epsilon_N^\infty \text{ are the lowest } N \text{ eigenvalues of } H_{W_\infty} = \frac{1}{2}\Delta + W_\infty, \end{cases}$$

hence belongs to $\mathcal{X}$.

To make this strategy of proof work, we therefore need to find a functional space $\mathcal{Y}$ in which the sequence $(W_n)_{n\in\mathbb{N}}$ is bounded. This will allow us in addition to clarify the notion of local potential in this setting (a local potential will be defined as an element of $\mathcal{Y}$). Unfortunately, we have not been able to find any non trivial[2] functional space $\mathcal{W}$ satisfying the above request. This mathematical difficulty has well-known numerical counterparts [321]:

(i) it is easy to construct dramatic modifications of the (computed) optimized effective potential that are "almost solutions" of the OEP problem;

(ii) variational approximations of the OEP problem in which the molecular orbitals and the trial effective potentials are discretized in independent basis sets lead to unphysical results.

8.2.2 A well-posed reformulation of the OEP problem

A way to circumvent the issue raised in the above discussion is to replace (8.11)-(8.12) with formally equivalent conditions that do not explicitly refer to a 'local potential' W [25].

Let us first deal with (8.11). Consider some operator W such that $(W\phi)\psi = \phi(W\psi)$ for all $(\phi,\psi) \in H^1(\mathbb{R}^3) \times H^1(\mathbb{R}^3)$ (which is the least we can demand to a 'local potential'). It is then clear that if $\Phi = \{\phi_1, \ldots, \phi_N\} \in (H^1(\mathbb{R}^3))^N$ satisfies (8.11), we also have

$$\begin{cases} \operatorname{div}(\phi_i\nabla\phi_1 - \phi_1\nabla\phi_i) = c_i\phi_1\phi_i, \\ \int_{\mathbb{R}^3} \phi_i\phi_j = \delta_{ij}, \end{cases} \tag{8.15}$$

with $c_i = 2(\epsilon_i - \epsilon_1)$. Conversely, if $\Phi = \{\phi_i\} \in (H^1(\mathbb{R}^3))^N$ satisfies (8.15), then *at least formally*, Φ satisfies (8.11) with, for instance,

$$W = \frac{\sum_{i=1}^N \phi_i\Delta\phi_i + \sum_{i=2}^N c_i\phi_i^2}{2\rho_\Phi}, \tag{8.16}$$

$\epsilon_1 = 0$, and $\epsilon_i = c_i/2$ for $2 \leq i \leq N$. We are therefore now in position to rigorously define a set of admissible local potentials

[2] It is of course possible to construct finite dimensional functional spaces $\mathcal{W}$ for which (8.13), with $\mathcal{X}$ defined by (8.10), has a solution. Reducing artificially the class of admissible potentials is however not a very satisfactory way to tackle the OEP problem.

$$\mathcal{W} = \Bigg\{ W \text{ operator on } L^2(\mathbb{R}^3) \;\Bigg|\; H_W = -\frac{1}{2}\Delta + W \text{ is a self-adjoint operator on } L^2(\mathbb{R}^3)$$
$$\text{with domain } D(W) \subset H^1_{\mathrm{loc}}(\mathbb{R}^3),$$
$$\text{bounded from below with at least } N \text{ eigenvalues below its essential spectrum,}$$
$$\text{and such that} \quad \forall (\phi,\psi) \in D(W)\times D(W),\ (H_W\phi)\psi - (H_W\psi)\phi = \frac{1}{2}\mathrm{div}\,(\phi\nabla\psi - \psi\nabla\phi)\Bigg\}.$$

In order to account for condition (8.12), we remark that for any $\Phi \in \mathcal{X}$, it holds for all $1 \le i \le N$,

$$\forall \psi \in C_0^\infty(\mathbb{R}^3), \qquad \frac{1}{2}\int_{\mathbb{R}^3} \phi_i^2\,|\nabla\psi|^2 = \langle \psi\phi_i, (H_W - \epsilon_i)\,\psi\phi_i\rangle.$$

It follows from the above equality (see [25] for details) that conditions (8.11)-(8.12) are rigorously equivalent to

$$\begin{cases} \left(-\frac{1}{2}\Delta + W\right)\phi_i = \epsilon_i\phi_i, \qquad \int_{\mathbb{R}^3}\phi_i\phi_j = \delta_{ij}, \\ \forall\psi \in C_0^\infty(\mathbb{R}^3),\ \forall 1 \le i \le N-1, \\ \qquad \int_{\mathbb{R}^3}\phi_i^2\,|\nabla\psi|^2 \ge 2\sum_{j=1}^{i}(\epsilon_j - \epsilon_1)\left(\int_{\mathbb{R}^3}\psi\phi_i\phi_j\right)^2 + 2(\epsilon_{i+1}-\epsilon_1)\left(\int_{\mathbb{R}^3}\psi^2\,\phi_i^2 - \sum_{j=1}^{i}\left(\int_{\mathbb{R}^3}\psi\phi_i\phi_j\right)^2\right). \end{cases}$$

Combining the above result with the formal equivalence between (8.11) and (8.15) with $c_i = 2(\epsilon_i - \epsilon_1)$, it is natural to introduce the optimization problem

$$\inf_{\Phi\in\widetilde{\mathcal{X}}} E^{\mathrm{HF}}(\Phi). \tag{8.17}$$

where

$$\widetilde{\mathcal{X}} = \Bigg\{ \Phi = \{\phi_i\}_{1\le i\le N} \;\Bigg|\; \phi_i \in H^1(\mathbb{R}^3), \quad \int_{\mathbb{R}^3}\phi_i\phi_j = \delta_{ij},$$
$$\exists 0 = c_1 \le c_2 \le \cdots \le c_N < \infty,\ \forall 2 \le i \le N, \quad \mathrm{div}\,(\phi_i\nabla\phi_1 - \phi_1\nabla\phi_i) = c_i\phi_1\phi_i,$$
$$\forall 1 \le i \le N-1,\ \forall\psi \in C_0^\infty(\mathbb{R}^3),$$
$$\int_{\mathbb{R}^3}\phi_i^2\,|\nabla\psi|^2 \ge \sum_{j=1}^{i} c_j\left(\int_{\mathbb{R}^3}\psi\phi_i\phi_j\right)^2 + c_{i+1}\left(\int_{\mathbb{R}^3}\psi^2\,\phi_i^2 - \sum_{j=1}^{i}\left(\int_{\mathbb{R}^3}\psi\phi_i\phi_j\right)^2\right)\Bigg\}.$$

We have therefore eliminated any explicit reference to a 'local potential'. The connection between the original OEP problem (8.13) and its reformulation (8.17) can be stated as follows:

(i) if Φ^{OEP} is solution to (8.13), then Φ^{OEP} is solution to (8.17);

(ii) if $\widetilde{\Phi}^{\mathrm{OEP}} = \left\{\widetilde{\phi}_i^{\mathrm{OEP}}\right\}_{1\le i\le N}$ is solution to (8.17), and if the reconstructed potential

$$W^{\mathrm{OEP}} = \frac{\sum_{i=1}^{N}\widetilde{\phi}_i^{\mathrm{OEP}}\Delta\widetilde{\phi}_i^{\mathrm{OEP}} + \sum_{i=2}^{N} c_i|\widetilde{\phi}_i^{\mathrm{OEP}}|^2}{2\rho_{\widetilde{\Phi}^{\mathrm{OEP}}}} \tag{8.18}$$

defines an element of $\mathcal{W}$, then $\widetilde{\Phi}^{\mathrm{OEP}}$ is solution to (8.13) and W^{OEP} is an optimized effective potential.

It is proved in [25] that for a neutral or positively charged two electron system, problem (8.17) has at least one global minimizer $\widetilde{\Phi}^{\mathrm{OEP}}$. Unfortunately, we have not been able to establish whether or not the reconstructed potential formally defined by (8.18) is in $\mathcal{W}$.

8.3 The effective local potential minimization problem

As shown in Section 8.2, the OEP problem in its original formulation is not well posed. We consider here an alternative way of obtaining an effective local potential (ELP), relying on some variance minimization. We show that the corresponding minimization problem is well-posed in the sense that the ELP is uniquely defined up to a uniform constant. We also provide an explicit analytical expression for it.

The effective local potential associated with a given $\Phi \in \mathcal{X}_N$ was originally defined as the local potential minimizing the function [185]

$$v \mapsto S_\Phi(v) = \sum_{i=1}^{N} \sum_{a=N+1}^{+\infty} |\langle \phi_i | (v - K_\Phi) | \phi_a \rangle|^2 ,$$

$(\phi_a)_{a \geq N+1}$ being a Hilbert basis of the orthogonal of the vector space generated by $(\phi_i)_{1 \leq i \leq N}$. A simple calculation shows that $S_\Phi(v) = J_\Phi^{\mathrm{ELP}}(v)$ where

$$J_\Phi^{\mathrm{ELP}}(v) = \frac{1}{2} \| [v - K_\Phi, \gamma_\Phi] \|_{\mathrm{HS}}^2,$$

$[A, B] = AB - BA$ denoting the commutator of the operators A and B. An intrinsic formulation of the ELP problem therefore reads

$$\inf \{ J_\Phi^{\mathrm{ELP}}(v),\ v \in \mathrm{L}^3(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3) \}. \tag{8.19}$$

Proposition 8.4. *Let $\Phi = (\phi_i)_{1 \leq i \leq N} \in \mathcal{X}_N$. Any solution $v_{x,\mathrm{ELP}}^\Phi$ to (8.19) satisfies*

$$\rho_\Phi(x) v_{x,\mathrm{ELP}}^\Phi(x) = -\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\, dy + \sum_{i,j=1}^{N} \left(\langle \phi_i | v_{x,\mathrm{ELP}}^\Phi | \phi_j \rangle - \langle \phi_i | K_\Phi | \phi_j \rangle \right) \phi_i(x)\, \phi_j(x) \tag{8.20}$$

and the symmetric matrix $M^\Phi = [\langle \phi_i | v_{x,\mathrm{ELP}}^\Phi | \phi_j \rangle]$ is solution to the linear system

$$(I - A^\Phi) M^\Phi = G^\Phi \tag{8.21}$$

with

$$A_{kl,ij}^\Phi = \int_{\mathbb{R}^3} \frac{\phi_i\, \phi_j\, \phi_k\, \phi_l}{\rho_\Phi}, \qquad G_{kl}^\Phi = \int_{\mathbb{R}^3} v_{x,S}^\Phi \phi_k \phi_l - \sum_{i,j=1}^{N} A_{kl,ij}^\Phi \langle \phi_i | K_\Phi | \phi_j \rangle.$$

Besides, if the orbitals ϕ_i are continuous and if the open set $\mathbb{R}^3 \setminus \rho_\Phi^{-1}(0)$ is connected, then the solutions to (8.21) form a one-dimensional affine set of the form

$$\bar{M} + \mathbb{R} I_N,$$

so that $v_{x,\mathrm{ELP}}^\Phi$ is uniquely defined, up to an additive constant, on the set where $\rho_\Phi > 0$, and can be given arbitrary values on the set where $\rho_\Phi = 0$.

Remark 8.2 (ELP for systems with spin states). *We denote the spin variables by α, β, and the number of electrons of spin σ by N_σ. The Euler-Lagrange equations associated with the Unrestricted Hartree-Fock problem read*

$$\begin{cases} -\dfrac{1}{2}\Delta \phi_i^\alpha + V_{\mathrm{nuc}} \phi_i^\alpha + \left(\rho_\Phi \star \dfrac{1}{|x|} \right) \phi_i^\alpha + K_{\Phi^\alpha} \phi_i^\alpha = \epsilon_i^\alpha \phi_i^\alpha, \\ -\dfrac{1}{2}\Delta \phi_i^\beta + V_{\mathrm{nuc}} \phi_i^\beta + \left(\rho_\Phi \star \dfrac{1}{|x|} \right) \phi_i^\beta + K_{\Phi^\beta} \phi_i^\beta = \epsilon_i^\beta \phi_i^\beta, \end{cases}$$

where ρ_Φ is the total density $\rho_\Phi(x) = \rho_{\Phi^\alpha}(x) + \rho_{\Phi^\beta}(x)$, with $\rho_{\Phi^\sigma}(x) = \sum_{i=1}^{N_\sigma} |\phi_i^\sigma(x)|^2$, and K_{Φ^α} and K_{Φ^β} the exchange operators defined by

$$(K_{\Phi^\alpha}\varphi)(x) = -\int_{\mathbb{R}^3} \frac{\gamma_{\Phi^\alpha}(x,y)}{|x-y|}\,\varphi(y)\,dy, \qquad (K_{\Phi^\beta}\varphi)(x) = -\int_{\mathbb{R}^3} \frac{\gamma_{\Phi^\beta}(x,y)}{|x-y|}\,\varphi(y)\,dy,$$

with $\gamma_{\Phi^\sigma}(x,y) = \sum_{i=1}^{N^\sigma} \phi_i^\sigma(x)\phi_i^\sigma(y)$. The effective local potentials v^α and v^β are then obtained by solving

$$\inf\ \{J_{\Phi^\sigma}(v^\sigma), \quad v^\sigma \in \mathrm{L}^3(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3)\ \}, \tag{8.22}$$

where $J_{\Phi^\sigma} :\ \mathrm{L}^3(\mathbb{R}^3) + \mathrm{L}^\infty(\mathbb{R}^3) \to \mathbb{R}^3$ is defined as

$$J_{\Phi^\sigma}(v^\sigma) = \frac{1}{2}\|[v^\sigma - K_{\Phi^\sigma}, \gamma_{\Phi^\sigma}]\|_{\mathrm{HS}}^2.$$

The results obtained in the spinless case straightforwardly apply.

Remark 8.3 (Variational definition of the Slater potential). *There is also an alternative definition of the Slater potential in terms of some minimization procedure in Hilbert-Schmidt norm, namely*

$$v_{\mathrm{x},S}^{\Phi} = \underset{v\in \mathrm{L}^3(\mathbb{R}^3)+\mathrm{L}^\infty(\mathbb{R}^3)}{\operatorname{arginf}} \|v\gamma_\Phi - K_\Phi\|_{\mathrm{HS}}^2.$$

This variational characterization is reminiscent of the definition of the effective local potential (ELP) through the minimization (8.19). Actually, as will be seen below, the ELP can be decomposed as a Slater part, plus correction terms. The Slater potentiel is believed to represent the long-range part of the exchange potential (decaying as $-1/|x|$ when $|x| \to +\infty$), whereas the remaining terms are believed to be exponentially decreasing.

8.4 Mathematical proofs

8.4.1 Some useful preliminary results

Recall that the set $\mathrm{L}^{3/2}(\mathbb{R}^3) + (\mathrm{L}^\infty(\mathbb{R}^3))_\epsilon$ is the set of all function ϕ which can be written, for all $\epsilon > 0$, as a sum $\phi = \phi_{3/2} + \phi_\infty$ with $\phi_{3/2} \in \mathrm{L}^{3/2}(\mathbb{R}^3)$ and $\|\phi_\infty\|_{\mathrm{L}^\infty(\mathbb{R}^3)} \le \epsilon$. When $W \in \mathrm{L}^{3/2}(\mathbb{R}^3) + (L^\infty(\mathbb{R}^3))_\epsilon$, the essential spectrum of the operator $-\frac{1}{2}\Delta + W$ is still $[0, +\infty)$ [52, 277].

Lemma 8.1 (Exponential decay of the orbitals). *Consider an orbital $\phi \in \mathrm{H}^2(\mathbb{R}^3)$ satisfying an equation of the form*

$$-\frac{1}{2}\Delta\phi + W\phi = -\mu\phi, \tag{8.23}$$

where the potential $W \in \mathrm{L}^{3/2}(\mathbb{R}^3) + (L^\infty(\mathbb{R}^3))_\epsilon$ is such that $W(x) \to 0$ when $|x| \to +\infty$, and $\mu > 0$. Then, for any $\eta > 0$, there exists $M_\eta > 0$ and $R_\eta > 0$ such that

$$\forall |x| \ge R_\eta, \qquad |\phi(x)| \le M_\eta \mathrm{e}^{-\sqrt{2\mu-\eta}|x|}. \tag{8.24}$$

Proof of Lemma 8.1. Kato's inequality $-\Delta|\phi| \le -\mathrm{sgn}(\phi)\,\Delta\phi$ implies

$$-\Delta|\phi| \le 2(\mu + W)\phi\ \mathrm{sgn}(\phi) = -2(\mu + W)|\phi|.$$

For $0 < \eta < 2\mu$, there exists $R_\eta > 0$ such that $2|W(x)| \le \eta$ when $|x| \ge R_\eta$. Then,

$$-\Delta|\phi| + (2\mu - \eta)|\phi| \le (2W - \eta)|\phi|.$$

Using the elementary solution of $-\Delta + (2\mu - \eta)$, namely $u(x) = (4\pi|x|)^{-1} \exp(-\sqrt{2\mu - \eta}|x|)$, it follows

$$|\phi(x)| = \int_{\mathbb{R}^3} u(x-y)(-\eta + 2W(y))|\phi(y)|\, dy \leq \int_{|y|\leq R_\eta} u(x-y)(-\eta + 2W(y))|\phi(y)|\, dy$$

since $-\eta + 2W(y) \leq 0$ when $|x| \geq R_\eta$ and $|\phi| \geq 0$. Finally, the last integral in the above equality can be bounded by $M_\eta \exp(-\sqrt{2\mu - \eta}|x|)$ for some $M_\eta > 0$ and for $|x| \geq R_\eta > 0$, so that (8.24) follows. □

Lemma 8.2 (Asymptotic behavior of the Slater potential for exponentially decreasing orbitals). *Consider $\Phi^\star = (\phi_1^\star, \ldots, \phi_N^\star) \in [\mathrm{H}^2(\mathbb{R}^3)]^N$ and assume that there exists $R_\star > 0$ such that, for $|x| \geq R_\star$,*

$$\forall 1 \leq i \leq N, \qquad |\phi_i^\star(x)| \leq C_\star \exp\left(-\gamma_\star |x|\right), \tag{8.25}$$

for some $\gamma_\star, C_\star > 0$. Then the Slater potential $v_{\mathrm{x},S}^{\Phi^\star}$ defined by (8.7) is such that

$$v_{\mathrm{x},S}^{\Phi^\star}(x) \sim -\frac{1}{|x|}$$

when $|x| \to +\infty$.

Proof of Lemma 8.2. First, for any $R > R_\star$,

$$\int_{|y|\geq R} \frac{\phi_i^\star \phi_j^\star(y)}{|x-y|}\, dy \leq \left(\int_{|y|\geq R} \frac{|\phi_i^\star(y)|^2}{|x-y|^2}\right)^{1/2} \left(\int_{|y|\geq R} |\phi_i^\star(y)|^2\right)^{1/2} \leq CR^2\, \mathrm{e}^{-\gamma_\star R} \tag{8.26}$$

for some $C > 0$, using Hardy's inequality to bound the first term on the right hand-side, and the exponential fall-off of the j-th orbital for the second term. Second,

$$\left|\int_{|y|\leq R} \frac{\phi_i^\star \phi_j^\star(y)}{|x-y|}\, dy - \frac{\int_{|y|\leq R} \phi_i^\star \phi_j^\star(y)\, dy}{|x|}\right| \leq \int_{|y|\leq R} |\phi_i^\star(y)\phi_j^\star(y)| \left|\frac{|y-x| - |x|}{|x| \cdot |y-x|}\right| dy,$$

so that

$$\left|\int_{|y|\leq R} \frac{\phi_i^\star \phi_j^\star(y)}{|x-y|}\, dy - \frac{\int_{|y|\leq R} \phi_i^\star \phi_j^\star(y)\, dy}{|x|}\right| \leq \frac{1}{|x|} \int_{|y|\leq R} |y|\, |\phi_i^\star(y)\phi_j^\star(y)| \frac{1}{|y-x|}\, dy.$$

Using a Hölder inequality,

$$\int_{|y|\leq R} |y|\, |\phi_i^\star(y)\phi_j^\star(y)| \frac{1}{|y-x|}\, dy \to 0$$

when $|x| \to +\infty$, which concludes the proof. □

8.4.2 Proofs for the Slater potential

Proof of Theorem 8.1. The strategy of proof is based on a fixed-point argument. Notice that variational methods cannot be used since (8.8) seems to have no variational interpretation.

For all $\eta \geq 0$, we consider the problem

$$\begin{cases} \left(-\frac{1}{2}\Delta - \frac{Z+\eta}{|x|} + \rho_{\Phi^\eta} \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Phi^\eta,\eta}\right)\phi_i^\eta = \epsilon_i^\eta \phi_i^\eta, \\ \int_{\mathbb{R}^3} \phi_i^\eta \phi_j^\eta = \delta_{ij}, \\ \epsilon_1^\eta \leq \cdots \leq \epsilon_N^\eta \text{ are the lowest } N \text{ eigenvalues of } \left(-\frac{1}{2}\Delta - \frac{Z+\eta}{|x|} + \rho_{\Phi^\eta} \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Phi^\eta,\eta}\right) \text{ (on } L^2_r(\mathbb{R}^3)) \end{cases} \tag{8.27}$$

where

$$v_{\mathrm{x},S}^{\Phi,\eta}(x) = -\frac{1}{\rho_\Phi(x)+\eta}\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\,dy.$$

The proof of existence of a solution to (8.27) for $\eta = 0$ follows the lines of the proof of Theorem III.3 in [214]. We first construct, for $\eta > 0$, a continuous application T^η whose fixed points are solutions to (8.27) in $\mathcal{X}_N^r$. We then prove the existence of a fixed point of T^η using Schauder Theorem. The existence of a solution to (8.27) in the case when $\eta = 0$ is finally obtained using some limiting procedure. Note that we have introduced the parameter η both in the nucleus-electron interaction and in the Slater potential. In the former term, η plays the same role as in [214] (i.e. it enables us to control the decay of the orbitals at infinity). The role of η in the latter term is to ensure the continuity of the nonlinear application T^η for $\eta > 0$.

First step. Construction of the application T^η.

Let $\eta > 0$ and

$$K = \left\{ \Psi = (\psi_i)_{1\leq i\leq N} \in (H_r^1(\mathbb{R}^3))^N \;\middle|\; \left[\int_{\mathbb{R}^3} \phi_i\phi_j\right] \leq I_N \right\},$$

I_N denoting the identity matrix of rank N. The semidefinite constraint $\left[\int_{\mathbb{R}^3} \phi_i\phi_j\right] \leq I_N$ means

$$\forall x \in \mathbb{R}^N, \quad \sum_{i,j=1}^N \left(\int_{\mathbb{R}^3} \phi_i\phi_j\right) x_i x_j \leq |x|^2.$$

It is easy to see that K is a nonempty, closed, bounded, convex subset of the Hilbert space $(H_r^1(\mathbb{R}^3))^N$, containing $\mathcal{X}_N^r$. For $\Psi \in K$, we denote by $\gamma_\Psi(x,y) = \sum_{i=1}^N \psi_i(x)\psi_i(y)$, $\rho_\Psi(x) = \gamma_\Psi(x,x)$ and

$$\widetilde{F}_\Psi^\eta = -\frac{1}{2}\Delta - \frac{Z+\eta}{|x|} + \rho_\Psi \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Psi,\eta}.$$

As the potential $V_\Psi^\eta = -\frac{Z+\eta}{|x|} + \rho_\Psi \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Psi,\eta}$ belongs to

$$L^2(\mathbb{R}^3) + L^\infty_\epsilon(\mathbb{R}^3) = \left\{ W \mid \forall \epsilon > 0,\ \exists (W_2, W_\infty) \in L^2(\mathbb{R}^3) \times L^\infty(\mathbb{R}^3),\ \|W_\infty\|_{L^\infty} \leq \epsilon,\ W = W_2 + W_\infty \right\},$$

it is a compact perturbation of the kinetic energy operator. By Weyl Theorem [277], $\sigma_{\mathrm{ess}}(\widetilde{F}_\Psi^\eta) = \sigma_{\mathrm{ess}}(-\frac{1}{2}\Delta) = [0,\infty)$. Besides, using Gauss theorem and the inequalities $-\frac{N}{|\cdot|} \leq -\rho_\Psi \star \frac{1}{|x|} \leq v_{\mathrm{x},S}^{\Psi,\eta} \leq 0$, one has $-\frac{Z+\eta}{|x|} \leq V_\Psi^\eta \leq -\frac{\eta}{|x|}$. Hence,

$$\mathcal{G}^{Z+\eta} := -\frac{1}{2}\Delta - \frac{Z+\eta}{|x|} \leq \widetilde{F}_\Psi^\eta \leq \mathcal{G}^\eta := -\frac{1}{2}\Delta - \frac{\eta}{|x|}. \tag{8.28}$$

As the hydrogen-like Hamiltonian $\mathcal{G}^\eta$, considered as an operator on $L^2_r(\mathbb{R}^3)$, has infinitely many negative eigenvalues, so does $\widetilde{F}_\Psi$ (this is a straightforward consequence of Courant-Fischer min-max principle). Besides, the eigenvalues of the radial Schrödinger operator $\widetilde{F}_\Psi^\eta$ being simple, the spectral problem

$$\begin{cases} \widetilde{F}^{\eta}_{\Psi}\phi_i = \epsilon_i\phi_i, \\ \displaystyle\int_{\mathbb{R}^3} \phi_i\phi_j = \delta_{ij}, \\ \epsilon_1 \leq \cdots \leq \epsilon_N \text{ are the lowest } N \text{ eigenvalues of } \widetilde{F}^{\eta}_{\Psi} \text{ (on } L^2_r(\mathbb{R}^3)), \end{cases}$$

has a unique solution $\Phi = (\phi_i)$ in $\mathcal{X}^r_N \subset K$ up to the signs of the orbitals ϕ_i. We can therefore define a nonlinear application T^η from K to K which associates with any $\Psi \in K$ the unique solution $\Phi = (\phi_i) \in \mathcal{X}^r_N \subset K$ to (8.27), for which $\phi_i \geq 0$ in a neighborhood of $x = 0$, for all $1 \leq i \leq N$ (by the strong maximum principle, ϕ_i cannot vanish on an open set of $\mathbb{R}^3$).

Second step. Existence of a solution to (8.27) for $\eta > 0$.

By standard perturbation theory, it is not difficult to prove that T^η is continuous (for the H^1 norm topology). Let us prove that T^η is compact. Let (Ψ^n) be a bounded sequence in K, and let $\Phi^n = T^\eta\Psi^n$. There is no restriction in assuming that (Ψ^n) converges to some $\Psi^\eta \in (H^1(\mathbb{R}^3))^N$, weakly in $(H^1(\mathbb{R}^3))^N$, strongly in $(L^2_{\text{loc}}(\mathbb{R}^3))^N$ and almost everywhere. This implies in particular that the sequence $(\rho_{\Psi^n} \star \frac{1}{|x|} + v^{\Psi^n,\eta}_{\text{x},S})$ is bounded in L^∞ and converges almost everywhere to $\rho_{\Psi^\eta} \star \frac{1}{|x|} + v^{\Psi^\eta,\eta}_{\text{x},S}$ when n goes to infinity. Using again (8.28) and denoting by ϵ^n_i the i-th eigenvalue of $F^\eta_{\Psi^n}$, one obtains

$$\frac{1}{2}\sum_{i=1}^N \left(\|\nabla\phi^n_i\|_{L^2} - 2(Z+\eta)\right)^2 - 2(Z+\eta)^2 \leq \sum_{i=1}^N \frac{1}{2}\int_{\mathbb{R}^3} |\nabla\phi^n_i|^2 - \int_{\mathbb{R}^3} \frac{Z+\eta}{|x|}\rho_{\Phi^n} \leq \sum_{i=1}^N \epsilon^n_i < 0.$$

Thus, for all $1 \leq i \leq N$, the sequence $(\phi^n_i)_{n\in\mathbb{N}^*}$ is uniformly bounded in $H^1(\mathbb{R}^3)$ (independently of (Ψ^n)), and therefore converges, up to extraction, to some $\phi^\eta_i \in H^1_r(\mathbb{R}^3)$, weakly in $H^1(\mathbb{R}^3)$, strongly in $L^2_{\text{loc}}(\mathbb{R}^3)$ and almost everywhere. Besides, using (8.28) and Courant-Fischer formula, one obtains

$$-\frac{(Z+\eta)^2}{2i^2} \leq \epsilon^n_i \leq -\frac{\eta^2}{2i^2}.$$

Up to extraction, (ϵ^n_i) therefore converges to some $\epsilon^\eta_i \in [-\frac{(Z+\eta)^2}{2i^2}, -\frac{\eta^2}{2i^2}]$. Next, by Kato inequality [277],

$$\begin{aligned} -\Delta|\phi^n_i| &\leq -\text{sgn}(\phi^n_i)\Delta\phi^n_i = 2(\epsilon^n_i - V^\eta_{\Psi^n})|\phi^n_i| \\ &\leq 2\left(\frac{Z+\eta}{|x|} - \frac{\eta^2}{i^2}\right)|\phi^n_i|. \end{aligned} \tag{8.29}$$

As, moreover, (Ψ^n) and (Φ^n) are bounded for the H^1 norm topology, $(V^\eta_{\Psi^n}\phi^n_i)$ is bounded in $L^2(\mathbb{R}^3)$, so that (ϕ^n_i) is bounded in $H^2(\mathbb{R}^3)$, hence in $L^\infty(\mathbb{R}^3)$. Consequently, it follows from (8.29) and the maximum principle that there exists $\delta > 0$ small enough and $M \geq 0$ independent of i and n, such that

$$|\phi^n_i(x)| \leq M\,\mathrm{e}^{-\left(\frac{\sqrt{2}\,\eta}{N} - \delta\right)|x|}.$$

This implies that $(\phi^n_i)_{n\in\mathbb{N}^*}$ converges (up to extraction) to ϕ^η_i strongly in $L^2(\mathbb{R}^3)$. In particular, $\Phi^\eta = (\phi^\eta_i) \in \mathcal{X}^r_N$. It is therefore possible to check, using the convergence of (Ψ^n) to Ψ^η and the convergence - up to extraction - of (Φ^n) to Φ^η and of (ϵ^n_i) to ϵ^η_i, that

$$-\frac{1}{2}\Delta\phi^\eta_i + V^\eta_{\Psi^\eta}\phi^\eta_i = \epsilon^\eta_i\phi^\eta_i$$

and then, using the positivity of $\rho_{\Psi^n} \star \frac{1}{|x|} + v^{\Psi^n,\eta}_{\text{x},S}$ and Fatou lemma on the one hand, and the lower semi-continuity of the functional $\phi \mapsto \int_{\mathbb{R}^3} |\nabla\phi|^2$ on the other hand, that

$$\liminf_{n\to+\infty} -\int_{\mathbb{R}^3} |\nabla\phi_i^n|^2 = \liminf_{n\to+\infty} 2\int_{\mathbb{R}^3} (V_{\Psi^n}^\eta - \epsilon_i^n)|\phi_i^n|^2$$
$$\geq 2\int_{\mathbb{R}^3} (V_{\Psi^\eta}^\eta - \epsilon_i^\eta)|\phi_i^\eta|^2 = -\int_{\mathbb{R}^3} |\nabla\phi_i^\eta|^2.$$

As on the other hand,

$$\int_{\mathbb{R}^3} |\nabla\phi_i^\eta|^2 \leq \liminf_{n\to+\infty} \int_{\mathbb{R}^3} |\nabla\phi_i^n|^2,$$

(Ψ^n) converges to Ψ^η strongly in $(H^1(\mathbb{R}^3))^N$, which proves that T^η is compact. It then follows from Schauder fixed point theorem [375] that T^η has a fixed point $\Phi^\eta \in \mathcal{X}_N^r$, which is solution to (8.27).

Third step. Existence of a solution to (8.27) for $\eta = 0$.

Let (η_n) be a sequence of positive real numbers converging to zero. As the sequence of corresponding fixed points (Φ^{η_n}) is uniformly bounded in $(H^1(\mathbb{R}^3))^N$ and as $-\frac{(Z+\eta_n)^2}{2i^2} \leq \epsilon_i^{\eta_n} \leq 0$, there is no restriction in assuming that (Φ^{η_n}) converges to some $\Phi^\star \in (H^1(\mathbb{R}^3))^N$, weakly in $(H^1(\mathbb{R}^3))^N$, strongly in $(L^2_{\mathrm{loc}}(\mathbb{R}^3))^N$ and almost everywhere, and that $(\epsilon_i^{\eta_n})$ converges to $\epsilon_i^* \leq 0$. Besides, the sequence (Φ^{η_n}) is bounded in $(H^2(\mathbb{R}^3))^N$, hence in $(L^\infty(\mathbb{R}^3))^N$.

Passing to the limit in the equation $\widetilde{\mathcal{F}}_{\Phi^{\eta_n}}^{\eta_n}\phi_i^{\eta_n} = \epsilon_i^{\eta_n}\phi_i^{\eta_n}$ yields

$$-\frac{1}{2}\Delta\phi_i^\star - \frac{Z}{|x|}\phi_i^\star + \left(\rho_{\Phi^\star} \star \frac{1}{|x|}\right)\phi_i^\star + v_{\mathrm{x},S}^{\Phi^\star}\phi_i^\star = \epsilon_i^\star\phi_i^\star.$$

Assume that $\int_{\mathbb{R}^3} \rho_{\Phi^\star} < N$. As

$$\widetilde{\mathcal{F}}_{\Phi^{\eta_n}} \leq -\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^{\eta_n}} \star \frac{1}{|x|},$$

one has, using Courant-Fischer formula, and denoting by $\lambda_i(A)$ the i-th eigenvalue of A,

$$\begin{aligned}
\epsilon_i^\star &= \lim_{n\to+\infty} \epsilon_i^{\eta_n} \\
&= \lim_{n\to+\infty} \lambda_i\left(\widetilde{\mathcal{F}}_{\Phi^{\eta_n}}\right) \\
&\leq \lim_{n\to+\infty} \lambda_i\left(-\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^{\eta_n}} \star \frac{1}{|x|}\right) \\
&= \lambda_i\left(-\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^\star} \star \frac{1}{|x|}\right) \\
&\leq \lambda_i\left(-\frac{1}{2}\Delta - \frac{N - \int_{\mathbb{R}^3}\rho_{\Phi^\star}}{|x|}\right) \\
&= -\frac{(N - \int_{\mathbb{R}^3}\rho_{\Phi^\star})^2}{2i^2} < 0.
\end{aligned}$$

It follows that for n large enough, the sequence $(\epsilon_i^{\eta_n})$ is isolated from zero. As (Φ^{η_n}) is bounded in $(L^\infty(\mathbb{R}^3))^N$, we conclude, reasoning as above, that there exists $M \in \mathbb{R}_+$ and $\alpha > 0$ such that for n large enough

$$|\phi_i^{\eta_n}(x)| \leq M\,\mathrm{e}^{-\alpha|x|}.$$

This implies that (Φ^{η_n}) converges to $\Phi^\star \in (H^1(\mathbb{R}^3))^N$ strongly in $(L^2(\mathbb{R}^3))^N$, and consequently that $\int_{\mathbb{R}^3}\rho_{\Phi^\star} = N$. We reach a contradiction. This means that $\int_{\mathbb{R}^3}\rho_{\Phi^\star} = N$ and therefore that $\Phi^\star \in \mathcal{X}_N^r$.

This proves that $(\phi_i^\star)$ are orthonormal eigenvectors of $\widetilde{F}^0_{\Phi^\star}$. The fact that $\epsilon_1^\star < \cdots < \epsilon_N^\star$ are the lowest eigenvalues of $\widetilde{F}^0_{\Phi^\star}$ follows from Courant-Fischer formula.

In view of the proof of Proposition 8.1, the Slater potential $v_{x,S}^{\Phi^\star}$ is equivalent to $-\frac{1}{|x|}$ at infinity. This proves that $\epsilon_1^\star < \cdots < \epsilon_N^\star < 0$, from which it follows that the orbitals $\phi_i^\star$ enjoy exponential decay: For all $\eta > 0$, there exists $M \in \mathbb{R}^3$ such that

$$|\phi_i^\star(x)| \leq M\, \mathrm{e}^{-(\sqrt{-2\epsilon_N^\star}-\eta/3)|x|},$$

so that

$$v_{\mathrm{x},S}^{\Phi^\star}(x) = -\frac{1}{|x|} + \mathrm{o}\left(\mathrm{e}^{-(2\sqrt{-2\epsilon_N^\star}-\eta)|x|}\right).$$

Lastly, the same arguments as above can be used to prove that the minimum of the Hartree-Fock energy over the set of solutions to (8.8) is attained. □

Proof of Proposition 8.1. The well-posedness of the iterative procedure is granted provided the aufbau principle associated with the Hamiltonian

$$H_{\Phi^n} = -\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^n} \star \frac{1}{|x|} + v_{\mathrm{x},S}^{\Phi^n} \tag{8.30}$$

is well-posed for all $n \geq 0$. This in turn is guaranteed provided the lowest N negative eigenvalues of H_{Φ^n} can be computed unambiguously (Notice that the essential spectrum of H_{Φ^n} is still $[0, +\infty)$).

When the orbitals $\Phi = \{\phi_i\}_{i=1,\ldots,N}$ are radial, the asymptotic behavior of the Slater potential can be precised. Gauss's theorem shows that

$$\int_{\mathbb{R}^3} \frac{\phi_i\phi_j(y)}{|x-y|}\,dy = \int_{\mathbb{R}^3} \frac{\phi_i\phi_j(y)}{\max(|x|,|y|)}\,dy = \begin{cases} \dfrac{1}{|x|} + \mathrm{o}\left(\dfrac{1}{|x|}\right) & \text{when } i = j, \\ \mathrm{o}\left(\dfrac{1}{|x|}\right) & \text{when } i \neq j. \end{cases}$$

Indeed,

$$\int_{\mathbb{R}^3} \frac{\phi_i\phi_j(y)}{\max(|x|,|y|)}\,dy = \frac{1}{|x|}\left(\delta_{ij} - \int_{|y|\geq|x|} \phi_i\phi_j\right) + \int_{|y|\geq|x|} \frac{\phi_i\phi_j(y)}{|y|}\,dy.$$

The second integral on the right hand side converges to 0 when $|x| \to +\infty$, and the rate of convergence can be precised as

$$\left|\int_{|y|\geq|x|} \frac{\phi_i\phi_j(y)}{|y|}\,dy\right| \leq \frac{1}{|x|}\left(\int_{|y|\geq|x|} \phi_i^2\right)^{1/2}\left(\int_{|y|\geq|x|} \phi_j^2\right)^{1/2} = \mathrm{o}\left(\frac{1}{|x|}\right)$$

since the functions ϕ_i are in $\mathrm{L}^2(\mathbb{R}^3)$. The first term is handled in a similar manner. Finally,

$$v_{\mathrm{x},S}^{\Phi}(x) = -\sum_{i=1}^{N} \frac{\phi_i^2(x)}{\rho(x)}\,\frac{1}{|x|} + \mathrm{o}\left(\frac{1}{|x|}\right) = -\frac{1}{|x|} + \mathrm{o}\left(\frac{1}{|x|}\right)$$

when $|x| \to +\infty$.

A classical scaling argument (as for in proof of Lemma II.1 in [214] for instance) then shows that, for all $n \geq 0$, H_{Φ^n} has infinitely many single negative eigenvalues. Therefore, the new orbitals can be uniquely constructed. □

Proof of Proposition 8.2. Using a Cauchy-Schwarz inequality, the following bound is obtained:

$$-\frac{1}{2}\Delta - \frac{Z}{|x|} \leq H_{\Phi^n} \leq \tilde{H}_{\Phi^n} = -\frac{1}{2}\Delta - \frac{Z}{|x|} + \rho_{\Phi^n} \star \frac{1}{|x|}. \tag{8.31}$$

It is not sufficient to obtain the existence of infinitely many negtive eigenvalues when $Z = N$ and the orbitals are not required to be radial. This is however the case when $Z = N + \eta$ (for some $\eta > 0$), using again a scaling argument as in [214, Lemma II.1]. The proof of Proposition 8.2 is therefore analogous to the proof of Proposition 8.1, and we skip it. □

Proof of Proposition 8.3. When $Z = N$ and the orbitals are not radial but have an initial exponential decay, we show that

(i) the Hamiltonian H_{Φ^n} defined by (8.30) has infinitely many eigenvalues below 0;

(ii) the corresponding eigenvectors are still exponentially decreasing.

The proof of well-posedness of the iterative procedure is done using the following recurrence assumption:

Recurrence assumption 8.1. *There exists $R^n > 0$ such that, for $|x| \geq R^n$,*

$$\forall 1 \leq i \leq N, \qquad |\phi_i^n(x)| \leq C^n \exp\left(-\gamma^n |x|\right), \tag{8.32}$$

for some $\gamma^n, C^n > 0$.

This assumption is verified for $n = 0$. If it is verified for $n \geq 0$, then, by Lemma 8.2, the Slater potential behaves as $-1/|x|$ at infinity. A classical scaling argument then shows that there are infinitely many negatives eigenvalues. The exponential fall-off of the associated orbitals $\{\phi_i^{n+1}\}_{i=1,\ldots,N}$ can then be shown using Lemma 8.1, so that the recurrence assumption (8.1) is satisfied for $n+1$. □

8.4.3 Proof of Proposition 8.4

For all $v \in L^3(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$, the operator $B^\Phi v = [v, \gamma_\Phi]$ is Hilbert-Schmidt. One can therefore define on $L^3(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$ the functional

$$J_\Phi^{\mathrm{ELP}}(v) = \frac{1}{2}\|[v - K_\Phi, \gamma_\Phi]\|_{\mathrm{HS}}^2 = \frac{1}{2}\|B^\Phi v - [K_\Phi, \gamma_\Phi]\|_{\mathrm{HS}}^2.$$

For all v and h in $L^3(\mathbb{R}^3) + L^\infty(\mathbb{R}^3)$,

$$J_\Phi^{\mathrm{ELP}}(v + h) = J_\Phi^{\mathrm{ELP}}(v) + \langle B^\Phi v - [K_\Phi, \gamma_\Phi], B^\Phi h\rangle_{\mathrm{HS}} + \frac{1}{2}\|B^\Phi h\|_{\mathrm{HS}}^2$$

and

$$\begin{aligned}
&\langle B^\Phi v - [K_\Phi, \gamma_\Phi], B^\Phi h\rangle_{\mathrm{HS}} \\
&= \int_{\mathbb{R}^3} \left(\rho_\Phi(x) v + \int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\, dy + \sum_{i,j=1}^N \langle \phi_i \,|v - K_\Phi\,|\phi_j\rangle \phi_i(x)\phi_j(x) \right) h(x)\, dx.
\end{aligned}$$

The global minimizers v of (8.19) are therefore exactly the solutions to the equation

$$\rho_\Phi(x) v(x) = -\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\, dy + \sum_{i,j=1}^N \langle \phi_i | v - K_\Phi | \phi_j\rangle \phi_i(x)\phi_j(x). \tag{8.33}$$

Multiplying the above equation by $\frac{\phi_i\phi_j}{\rho}$ and integrating over $\mathbb{R}^3$, one then observes that a function v satisfying

$$\rho_\Phi(x)v(x) = -\int_{\mathbb{R}^3} \frac{|\gamma_\Phi(x,y)|^2}{|x-y|}\,dy + \sum_{i,j=1}^{N} (M_{ij} - \langle \phi_i | K_\Phi | \phi_j \rangle)\phi_i(x)\phi_j(x).$$

is solution to (8.33) if and only if the matrix M is solution to the linear system

$$(I - A^\Phi)M = G^\Phi. \tag{8.34}$$

Let us now prove that, if the orbitals ϕ_i are continuous and if $\mathbb{R}^3 \setminus \rho_\Phi^{-1}(0)$ is connected, then $\mathrm{Ker}(I - A^\Phi) = \mathbb{R} I_N$ and $G^\Phi \in \mathrm{Ran}(I - A^\Phi)$. For this purpose, let us consider a matrix $M \in \mathcal{M}_S(N)$ such that $\left(I - A^\Phi\right) M = 0$. As M is symmetric, it can be diagonalized in an orthonormal basis set as

$$M = U^T \,\mathrm{Diag}(\lambda_1, \cdots, \lambda_N)\, U$$

where U is a unitary matrix. Denoting by $(\psi_1, \ldots, \psi_N)^T = U(\phi_1, \ldots, \phi_N)^T$, a simple calculation leads to

$$0 = (\left(I - A^\Phi\right) M, M)_F = \sum_{i=1}^{N} \lambda_i^2 - \int_{\mathbb{R}^3} \left| \sum_{i=1}^{N} \lambda_i \psi_i(x)^2 \right|^2 \frac{dx}{\rho_\Phi(x)},$$

where $(\cdot,\cdot)_F$ is the Frobenius inner product on $\mathcal{M}_S(N)$. As U is a unitary transform, the ψ_i are orthonormal for the $\mathrm{L}^2(\mathbb{R}^3)$ inner product and $\sum_{i=1}^{N} \psi_i(x)^2 = \rho_\Phi(x)$. Therefore, using Cauchy-Schwarz inequality,

$$\left| \sum_{i=1}^{N} \lambda_i \psi_i(x)^2 \right|^2 \leq \left(\sum_{i=1}^{N} \psi_i(x)^2 \right) \left(\sum_{i=1}^{N} \lambda_i^2 \psi_i(x)^2 \right) = \rho_\Phi(x) \sum_{i=1}^{N} \lambda_i^2 \psi_i(x)^2,$$

with equality if and only if there exists $C(x)$ such that $\lambda_i \psi_i(x) = C(x)\psi_i(x)$ for all $1 \leq i \leq N$. Hence,

$$\sum_{i=1}^{N} \lambda_i^2 - \int_{\mathbb{R}^3} \left| \sum_{i=1}^{N} \lambda_i \psi_i(x)^2 \right|^2 \frac{dx}{\rho_\Phi(x)} \geq \sum_{i=1}^{N} \lambda_i^2 - \int_{\mathbb{R}^3} \sum_{i=1}^{N} \lambda_i^2 \psi_i^2 = 0,$$

with equality if and only if for almost all $x \in \mathbb{R}^3$, there exists $C(x)$ such that $\lambda_i \psi_i(x) = C(x)\psi_i(x)$ for all $1 \leq i \leq N$.

If the orbitals ϕ_i are continuous, so are the functions ψ_i. Let us consider the open sets $\Omega_i = \mathbb{R}^3 \setminus \psi_i^{-1}(0)$ and $\Omega = \cup_{i=1}^{N} \Omega_i = \mathbb{R}^3 \setminus \rho_\Phi^{-1}(0)$. On Ω_i, one has $C(x) = \lambda_i$. This implies that the function $C(x)$ is constant on each connected component of Ω. If Ω is connected, one therefore has $\lambda_1 = \lambda_2 = \cdots = \lambda_N$, i.e. M is proportional to the identity matrix.

In summary, under the assumptions that the orbitals ϕ_i are continuous and that $\mathbb{R}^3 \setminus \rho_\Phi^{-1}(0)$ is connected,

(1) the linear equation (8.34) has a solution if and only if $G^\Phi \in \mathrm{Ran}\left(I - A^\Phi\right)$. Note that $\mathrm{Ran}\left(I - A^\Phi\right) = \mathrm{Ker}\left(I - (A^\Phi)^*\right)^\perp = \mathrm{Ker}\left(I - A^\Phi\right)^\perp$, since A^Φ is self-adjoint for the Frobenius inner product. It then follows $\mathrm{Ran}\left(I - A^\Phi\right) = \mathrm{Span}(I_N)^\perp$. Since $(I_N, G^\Phi)_F = \mathrm{Tr}(G^\Phi) = 0$, $G^\Phi \in \mathrm{Ran}\left(I - A^\Phi\right)$ and (8.34) has at least one solution $M_\star^\Phi$;

(2) if $M_\star^\Phi$ is a solution to (8.34), then the set of the solutions of (8.34) is $\left\{M_\star^\Phi + \lambda I_{\mathbb{R}^N},\ \lambda \in \mathbb{R}\right\}$.

Note that replacing M^Φ with $M^\Phi + \lambda I_{\mathbb{R}^N}$ in (8.34) amounts to replacing $v_{\mathrm{x,ELP}}^\Phi$ with $v_{\mathrm{x,ELP}}^\Phi + \lambda$. $\square$

Partie V

Bibliographie

Bibliographie

[1] S. A. Adelman and J. D. Doll, Generalized langevin equation approach for atom-solid-surface scattering - general formulation for classical scattering off harmonic solids, *J. Chem. Phys.* **64**(6) (1976) 2375–2388.

[2] E. Akhmatskaya and S. Reich, The targetted shadowing hybrid Monte Carlo (TSHMC) method, In *New Algorithms for Macromolecular Simulation*, B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schuette, and R. Skeel (Eds.), volume 49 of *Lecture Notes in Computational Science and Engineering* (Springer Verlag, Berlin and New York, 2006), pp. 145–158.

[3] B. J. Alder and W. T. Wainwright, Molecular dynamics by electronic computers, In *Proc. of the Int. Symp. on Statistical Mechanical Theory of Transport Processes (Brussels, 1956)*, I. Progigine (Ed.) (Interscience, Wiley, New-York, 1956), pp. 97–131.

[4] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids* (Oxford University Press, 1987).

[5] R. J. Allen, D. Frenkel, and P. R. ten Wolde, Simulating rare events in equilibrium or nonequilibrium stochastic systems, *J. Chem. Phys.* **124**(2) (2006) 024102.

[6] S. A. Allison and J. A. McCammon, Transport-properties of rigid and flexible macromolecules by brownian dynamics simulation, *Biopolymers* **23**(1) (1984) 167–187.

[7] H. C. Andersen, Molecular-dynamics simulations at constant pressure and-or temperature, *J. Chem. Phys.* **72**(4) (1980) 2384–2393.

[8] H. C. Andersen, RATTLE - a velocity version of the SHAKE algorithm for molecular-dynamics calculations, *J. Comput. Phys.* **52**(1) (1983) 24–34.

[9] E. Arévalo, G. M. Mertens, Y. Gaididei, and A. R. Bishop, Thermal diffusion of supersonic solitons in anharmonic chain of atoms, *Phys. Rev. E* **67**(1) (2003) 016610.

[10] A. Arnold, P. A. Markowich, G. Toscani, and A. Unterreiter, On convex Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations, *Commun. Part. Diff. Eq.* **26**(1) (2001) 43–100.

[11] V. Arnol'd, *Mathematical methods of classical mechanics* (Springer, New York, 1989).

[12] W. T. Ashurst and W. G. Hoover, Dense-fluid shear viscosity *via* nonequilibrium molecular-dynamics, *Phys. Rev. A* **11**(2) (1975) 658–678.

[13] R. Assaraf, M. Caffarel, and A. Khelif, Diffusion Monte Carlo methods with a fixed number of walkers, *Phys. Rev. E* **61**(4) (2000) 4566–4575.

[14] Y. F. Atchade and J. S. Liu, The Wang-Landau algorithm for Monte Carlo computation in general state spaces, *Technical Report* (2004).

[15] J. B. Avalos and A. D. Mackie, Dissipative particle dynamics with energy conservation, *Europhys. Lett.* **40**(2) (1997) 141–146.

[16] J. B. Avalos and A. D. Mackie, Dynamic and transport properties of dissipative particle dynamics with energy conservation, *J. Chem. Phys.* **111**(11) (1999) 5267–5276.

[17] V. Bach, E. H. Lieb, M. Loss, and J. P. Solovej, There are no unfilled shells in unrestricted Hartree-Fock theory, *Phys. Rev. Lett.* **72**(19) (1994) 2981–2983.

[18] L. Baffico, S. Bernard, Y. Maday, G. Turinici, and G. Zerah, Parallel-in-time molecular-dynamics simulations, *Phys. Rev. E* **66**(5) (2002) 057701.

[19] D. Bakry and M. Emery, Diffusions hypercontractives, In *Séminaire de Probabilités XIX*, volume 1123 of *Lecture Notes in Mathematics* (Springer Verlag, 1985), pp. 177–206.

[20] E. Barth, B. J. Leimkuhler, and C. R. Sweet, Approach to thermal equilibrium in biomolecular simulation, In *New Algorithms for Macromolecular Simulation*, B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schuette, and R. Skeel (Eds.), volume 49 of *Lecture Notes in Computational Science and Engineering* (Springer Verlag, Berlin and New York, 2006), pp. 125–140.

[21] M. I. Baskes, Application of the embedded-atom method to covalent materials - a semiempirical potential for silicon, *Phys. Rev. Lett.* **59**(23) (1987) 2666–2669.

[22] M. I. Baskes, Modified embedded-atom potentials for cubic materials and impurities, *Phys. Rev. B* **46**(5) (1992) 2727–2742.

[23] M. I. Baskes, Atomistic model of plutonium, *Phys. Rev. B* **62**(23) (2000) 15532–15537.

[24] M. I. Baskes, J. S. Nelson, and A. F. Wright, Semiempirical modified embedded-atom potentials for silicon and germanium, *Phys. Rev. B* **40**(9) (1989) 6085–6100.

[25] A. Ben-Haj-Yedder, E. Cancès, and C. Le Bris, Mathematical remarks on the optimized effective potential problem, *Differential and Integral Equations* **17** (2004) 331–368.

[26] C. H. Bennett, Exact defect calculations in model substances, In *Diffusion in solids: Recent Developments*, A. S. Nowick and J. J. Burton (Eds.) (Academic Press, New York, 1977), pp. 73–113.

[27] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, and J. R. Haak, Molecular-dynamics with coupling to an external bath, *J. Chem. Phys.* **81**(8) (1984) 3684–3690.

[28] M. Berkowitz and J. A. McCammon, Molecular-dynamics with stochastic boundary-conditions, *Chem. Phys. Lett.* **90**(3) (1982) 215–217.

[29] A. Beskos, G. O. Roberts, A. M. Stuart, and J. Voss, An MCMC method for diffusion bridges, *Warwick preprint* **05/2006** (2006).

[30] B. M. Bibby and M. Sorensen, On estimation for discretely observed diffusions: a review, *Theory Stoch. Process.* **2** (1998) 49–56.

[31] J. J. Biesiadecki and R. D. Skeel, Dangers of multiple time step methods, *J. Comput. Phys.* **109** (1993) 318–328.

[32] X. Blanc, C. Le Bris, and F. Legoll, Analysis of a prototypical multiscale method coupling atomistic and continuum mechanics, *Math. Model. Numer. Anal.* **39**(4) (2005) 797–826.

[33] G. Blower and F. Bolley, Concentration of measure on product spaces with applications to Markov processes, *Studia Math.* **175** (2006) 47–72.

[34] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark, *Ann. Rev. Phys. Chem.* **53** (2002) 291–318.

[35] S. D. Bond, B. J. Leimkuhler, and B. B. Laird, The Nosé-Poincaré method for constant temperature molecular dynamics, *J. Comput. Phys.* **151**(1) (1999) 114–134.

[36] J.-F. Bonnans, J.-C. Gilbert, C. Lemaréchal, and C. Sagastizabal, *Numerical optimization. Theoretical and practical aspects* (Springer, 2003).

[37] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, New algorithm for Monte-Carlo simulation of Ising spin systems, *J. Comput. Phys.* **17**(1) (1975) 10–18.

[38] D. W. Brenner, D. H. Robertson, M. L. Elert, and C. T. White, Detonations at nanometer resolution using molecular-dynamics, *Phys. Rev. Lett.* **70**(14) (1993) 2174–2177.

[39] D. W. Brenner, D. H. Robertson, M. L. Elert, and C. T. White, Detonations at nanometer resolution using molecular dynamics (vol 70, pg 2174, 1993), *Phys. Rev. Lett.* **76**(12) (1996) 2202–2202.

[40] D. Bressanini and P. J. Reynolds, *Monte Carlo Methods in Chemical Physics*, volume 105 of *Advances in Chemical Physics* (Wiley New York, 1999).

[41] D. Bressanini and P. J. Reynolds, Spatial-partitioning-based acceleration for variational Monte Carlo, *J. Chem. Phys.* **111**(14) (1999) 6180–6189.

[42] C. Le Bris, *PhD thesis* (Ecole Polytechnique, 1993).

[43] C. L. Brooks and M. Karplus, Deformable stochastic boundaries in molecular-dynamics, *J. Chem. Phys.* **79**(12) (1983) 6312–6325.

[44] L. Bruneau and S. De Bièvre, A Hamiltonian model for linear friction in a homogeneous medium, *Comm. Math. Phys.* **109** (2002) 511–542.

[45] A. Brünger, C. L. Brooks, and M. Karplus, Stochastic boundary-conditions for molecular-dynamics simulations of ST2 water, *Chem. Phys. Lett.* **105**(5) (1984) 495–500.

[46] G. Bussi, A. Laio, and M. Parrinello, Equilibrium free energies from nonequilibrium metadynamics, *Phys. Rev. Lett.* **96**(9) (2006) 090601.

[47] M. Caffarel and P. Claverie, Development of a pure diffusion Quantum Monte-Carlo method using a full generalized Feynman-Kac formula. 2. Applications to simple systems, *J. Chem. Phys.* **88**(2) (1988) 1100–1109.

[48] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, and G. Turinici, High-order averaging schemes with error bounds for thermodynamical properties calculations by molecular dynamics simulations, *J. Chem. Phys.* **121**(21) (2004) 10346–10355.

[49] E. Cancès, F. Castella, P. Chartier, E. Faou, C. Le Bris, F. Legoll, and G. Turinici, Long-time averaging for integrable Hamiltonian dynamics, *Numer. Math.* **100**(2) (2005) 211–232.

[50] E. Cances, B. Jourdain, and T. Lelievre, Quantum Monte Carlo simulations of fermions. A mathematical analysis of the fixed-node approximation, *Math. Mod. Meth. Appl. Sci.* **16**(9) (2006) 1403–1440.

[51] E. Cancès, F. Legoll, and G. Stoltz, Theoretical and numerical comparison of sampling methods for molecular dynamics, *Math. Model. Numer. Anal.* **41**(2) (2007) 351–390.

[52] E. Cancès, C. Le Bris, and Y. Maday, *Méthodes mathématiques en chimie quantique*, volume 53 of *Mathématiques & Applications* (Springer-Verlag, Berlin, Heidelberg, 2006).

[53] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, and Y. Maday, Computational quantum chemistry: A primer, In *Handbook of Numerical Analysis (Special volume on computational chemistry)*, P. G. Ciarlet and C. L. Bris (Eds.), volume X (Elsevier, 2003), pp. 3–270.

[54] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events, *Chem. Phys. Lett.* **156**(5) (1989) 472–477.

[55] I. Catto, C. Le Bris, and P.-L. Lions, *The mathematical theory of thermodynamic limits: Thomas-Fermi type models* (Oxford University Press, New York, 1998).

[56] T. Çagin and B.M. Pettitt, Grand molecular dynamics: a method for open systems, *Mol. Simulat.* **6** (1991) 5–26.

[57] T. Çagin and B.M. Pettitt, Molecular dynamics with a variable number of molecules, *Mol. Phys.* **72** (1991) 169–175.

[58] D. Ceperley, G. V. Chester, and M. H. Kalos, Monte-carlo simulation of a many-fermion study, *Phys. Rev. B* **16**(7) (1977) 3081–3099.

[59] D. M. Ceperley, Path-integrals in the theory of condensed helium, *Rev. Mod. Phys.* **67**(2) (1995) 279–355.

[60] D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation, *J. Chem. Phys.* **68** (1978) 2959–2970.

[61] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, Oxford, 1987).

[62] S. Chandrasekhar, Stochastic problems in physics and astronomy, *Rev. Mod. Phys.* **15** (1943) 1–89.

[63] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, Spatially adaptive lattice coarse-grained Monte Carlo simulations for diffusion of interacting molecules, *J. Chem. Phys.* **121**(22) (2004) 11420–11431.

[64] Y. Chen, Another look at Rejection sampling through Importance sampling, *Institute of Statistics and Decision Science, Duke University - Discussion papers* **04-30** (2004).

[65] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics, *ChemPhysChem* **6**(9) (2005) 1809–1814.

[66] G. Ciccotti, T. Lelièvre, and E. Vanden-Eijnden, Sampling Boltzmann-Gibbs distributions restricted on a manifold with diffusions, *to appear in Comm. Pure Appl. Math.* (2007).

[67] G. Ciccotti and A. Tenenbaum, Canonical ensemble and non-equilibrium states by molecular-dynamics, *J. Stat. Phys.* **23**(6) (1980) 767–772.

[68] F. Cleri, S. R. Phillpot, D. Wolf, and S. Yip, Atomistic simulations of materials fracture and the link between atomic and continuum length scales, *J. Am. Ceramic Soc.* **81**(3) (1998) 501–516.

[69] A. J. Coleman, Structure of fermion density matrices, *Rev. Mod. Phys.* **35** (1963) 668–687.

[70] A. J. Coleman, Kummer variety, geometry of N-representability, and phase transitions, *Phys. Rev. A* **66**(2) (2002).

[71] A. J. Coleman and V. I. Yukalov, *Reduced Density Matrices*, volume 72 of *Lectures Notes in chemistry* (Springer, 2000).

[72] C.A. Coulson, Present state of molecular structure calculations, *Rev. Mod. Phys.* **132**(2) (1960) 170–177.

[73] R. Courant and K.O. Friedrichs, *Supersonic flow and shock waves* (Springer, 1991).

[74] G. E. Crooks and D. Chandler, Efficient transition path sampling for nonequilibrium stochastic dynamics, *Phys. Rev. E* **64**(2) (2001) 026109.

[75] E. Darve and A. Pohorille, Calculating free energies using average force, *J. Chem. Phys.* **115**(20) (2001) 9169–9183.

[76] E. Darve, M. A. Wilson, and A. Pohorille, Calculating free energies using a scaled-force molecular dynamics algorithm, *Mol. Simulat.* **28**(1-2) (2002) 113–144.

[77] T. Diaz de la Rubia, M. J. Caturla, E. Alonso, M. J. Fluss, and J. M. Perlado, Self-decay induced damage production and micro-structure evolution in fcc metals: An atomic-scale computer simulation approach, *Journal of Computer-Aided Materials Design* **5**(2-3) (1998) 243–264.

[78] C. Dellago and P. G. Bolhuis, Activation energies from transition path sampling simulations, *Mol. Simulat.* **30**(11-12) (2004) 795–799.

[79] C. Dellago, P. G. Bolhuis, and D. Chandler, On the calculation of reaction rate constants in the transition path ensemble, *J. Chem. Phys.* **110**(14) (1999) 6617–6625.

[80] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, Transition path sampling and the calculation of rate constants, *J. Chem. Phys.* **108**(5) (1998) 1964–1977.

[81] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Transition path sampling, *Advances In Chemical Physics* **123** (2002) 1–78.

[82] H. Deloof, S. C. Harvey, J. P. Segrest, and R. W. Pastor, Mean field stochastic boundary molecular-dynamics simulation of a phospholipid in a membrane, *Biochem.* **30**(8) (1991) 2099–2113.

[83] W. K. den Otter and W. J. Briels, The calculation of free-energy differences by constrained molecular-dynamics simulations, *J. Chem. Phys.* **109**(11) (1998) 4139–4146.

[84] A. Doucet, N. de Freitas, and N.J. Gordon, *Sequential Monte Carlo Methods in Practice*, Series Statistics for Engineering and Information Science (Springer, 2001).

[85] A. Doucet, P. Del Moral, and A. Jasra, Sequential monte carlo samplers, *J. Roy. Stat.. Soc. B* **68**(3) (2006) 411–436.

[86] D. Down, S.P. Meyn, and R.L. Tweedie, Exponential and uniform ergodicity of Markov processes, *Ann. Probab.* **23** (1995) 1671–1691.

[87] W. Dreyer and M. Kunik, Cold, thermal and oscilator closure of the atomic chain, *J. Phys. A* **33**(10) (2000) 2097–2129.

[88] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte-Carlo, *Phys. Lett. B* **195**(2) (1987) 216–222.

[89] M. Duflo, *Random iterative models* (Springer, Berlin, New York, 1997).

[90] G. E. Duvall, R. Manvi, and S. C. Lowell, Steady shock profile in a one-dimensional lattice, *J. Appl. Phys.* **40**(9) (1969).

[91] W. E and E. Vanden-Eijnden, *Metastability, conformation dynamics, and transition pathways in complex systems. Multiscale modelling and simulation*, volume 39 of *Lect. Notes Comput. Sci. Eng.* (Springer, Berlin, 2004).

[92] R. M. Erdahl, Representability, *Int. J. Quantum Chem.* **13**(6) (1978) 697–718.

[93] R. M. Erdahl, Two algorithms for the lower bound method of reduced density matrix theory, *Rep. Math. Phys.* **15** (1979) 147–162.

[94] P. Espanol, Dissipative particle dynamics for a harmonic chain: A first-principles derivation, *Phys. Rev. E* **53**(2) (1996) 1572–1578.

[95] P. Espanol, Dissipative particle dynamics with energy conservation, *Europhys. Lett.* **40**(6) (1997) 631–636.

[96] P. Espanol and M. Revenga, Smoothed dissipative particle dynamics, *Phys. Rev. E* **67**(2) (2003) 026705.

[97] P. Espanol, M. Serrano, and I. Zuniga, Coarse-graining of a fluid and its relation with dissipative particle dynamics and smoothed particle dynamics, *Int. J. Modern Phys. C* **8**(4) (1997) 899–908.

[98] P. Espanol and P. Warren, Statistical-mechanics of dissipative particle dynamics, *Europhys. Lett.* **30**(4) (1995) 191–196.

[99] R. Dautray et J.-L. Lions, *Analyse mathématique et calcul numérique pour les sciences et les techniques*, volume 1-3 (Masson, 1985).

[100] D. J. Evans, Homogeneous nemd algorithm for thermal-conductivity - application of non-canonical linear response theory, *Phys. Lett. A* **91**(9) (1982) 457–460.

[101] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in advanced mathematics (CRC Press, Chapman and Hall, 1991).

[102] H. Eyring, The activated complex in chemical reactions, *J. Chem. Phys.* **3**(2) (1935) 107–115.

[103] W. Fickett and W.C. Davis, *Detonation* (Dover Publication Inc., 2000).

[104] A.-M. Filip and S. Venakides, Existence and modulation of traveling waves in particle chains, *Comm. Pure Appl. Math.* **52** (1999) 693–735.

[105] C. Filippi and C. J. Umrigar, Multiconfiguration wave functions for quantum Monte-Carlo calculations of first-row diatomic molecules, *J. Chem. Phys.* **105**(1) (1996) 213–226.

[106] E. G. Flekkoy, P. V. Coveney, and G. De Fabritiis, Foundations of dissipative particle dynamics, *Phys. Rev. E* **62**(2) (2000) 2140–2157.

[107] H. A. Forbert and S. A. Chin, Fourth-order algorithms for solving the multivariable Langevin equation and the Kramers equation, *Phys. Rev. E* **63**(1) (2001) 016703.

[108] G. W. Ford, M. Kac, and P. Mazur, Statistical mechanics of assemblies of coupled oscillators, *J. Math. Phys.* **6** (1965) 504–515.

[109] B. M. Forrest and U. W. Suter, Accelerated equilibration of polymer melts by time-coarse-graining, *J. Chem. Phys.* **102**(8) (1995) 7256–7266.

[110] S. Fournais, M. Hoffmann-Ostenhof, T. Hoffmann-Ostenhof, and T. Ostergaard Sorensen, Sharp regularity results for Coulombic many-electron wave functions, *Commun. Math. Phys.* **255** (2005) 183–227.

[111] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, Molecular dynamics simulations of the complete satellite tobacco mosaic virus, *Structure* **14** (2006) 437–449.

[112] M. I. Freidlin and A. D. Wentzell, *Random perturbations of dynamical systems* (Springer, New-York, 1998).

[113] D. Frenkel and B. Smit, *Understanding Molecular Simulation, From Algorithms to Applications (2nd ed.)* (Academic Press, 2002).

[114] G. Friesecke, The multiconfiguration equations for atoms and molecules: charge quantization and existence of solutions, *Arch. Ration. Mech. Anal.* **169** (2003) 35–71.

[115] G. Friesecke and K. Matthies, Atomic scale localization of high energy solitary waves on lattices, *Physica D* **171** (2002) 211–220.

[116] G. Friesecke and R. L. Pego, Solitary waves on lattices: I. Qualitative properties, renormalization and continuum limit, *Nonlinearity* **12** (1999) 1601–1627.

[117] G. Friesecke and J. Wattis, Existence theorem for solitary waves on lattices, *Communun. Math. Phys.* **161** (1994) 391–418.

[118] M. Fukuda, B. J. Braams, M. Nakata, M. L. Overton, J. K. Percus, M. Yamashita, and Z. Zhao, Large-scale semidefinite programs in electronic structure calculation, *Math. Program. B* **109**(2-3) (2007) 553–580.

[119] B. Garcia-Archilla, J. M. Sanz-Serna, and R. D. Skeel, Long time step methods for oscillatory differential equations, *SIAM J. Sci. Comput.* **20**(3) (1998) 930–963.

[120] C. Garrod, M. V. Mihaillovic, and M. Rosina, The variational approach to the two-body density matrix, *J. Math. Phys.* **16**(4) (1975) 868–874.

[121] C. Garrod and J. K Percus, Reduction of the N-particle variational problem, *J. Math. Phys.* **5** (1964) 1756–1776.

[122] P. L. Geissler and C. Dellago, Equilibrium time correlation functions from irreversible transformations in trajectory space, *J. Phys. Chem. B* **108**(21) (2004) 6667–6672.

[123] T. Geyer, C. Gorba, and V. Helms, Interfacing Brownian dynamics simulations, *J. Chem. Phys.* **120**(10) (2004) 4573–4580.

[124] G. Gidofalvi and D. A. Mazziotti, Application of variational reduced-density-matrix theory to the potential energy surfaces of the nitrogen and carbon dimers, *J. Chem. Phys.* **122**(19) (2005) 194104.

[125] I. I. Gikhman and A. V. Skorokhod, *The theory of stochastic processes* (Springer, 2004).

[126] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Operators of Second Order* (Springer, 1998).

[127] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in practice* (Chapman and Hall, 1996).

[128] M. J. Gillan and M. Dixon, The calculation of thermal-conductivities by perturbed molecular-dynamics simulation, *J. Phys. C - Solid State Phys.* **16**(5) (1983) 869–878.

[129] D. T. Gillespie, General method for numerically simulating stochastic time evolution of coupled chemical-reactions, *J. Comput. Phys.* **22**(4) (1976) 403–434.

[130] D. T. Gillespie, Exact stochastic simulation of coupled chemical-reactions, *J. Phys. Chem.* **81**(25) (1977) 2340–2361.

[131] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, *J. Chem. Phys.* **115**(4) (2001) 1716–1733.

[132] D. T. Gillespie and L. R. Petzold, Improved leap-size selection for accelerated stochastic simulation, *J. Chem. Phys.* **119**(16) (2003) 8229–8234.

[133] J. A. Given and E. Clementi, Molecular-dynamics and Rayleigh-Benard convection, *J. Chem. Phys.* **90**(12) (1989) 7376–7383.

[134] D. Givon, R. Kupferman, and A. Stuart, Extracting macroscopic dynamics: Model problems and algorithms, *Nonlinearity* **17** (2004) R55–R127.

[135] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, Studies in Applied Mathematics (SIAM, 1989).

[136] H. Grabert, P. Hänggi, and P. Talkner, Microdynamics and nonlinear stochastic processes of gross variables, *J. Stat. Phys.* **22**(5) (1980) 537–552.

[137] G. Grimett and D. Stirzaker, *Probability and Random Processes* (Oxford University Press, 2001).

[138] O. V. Gritsenko and E. J. Baerends, Orbital structure of the Kohn-Sham exchange potential and exchange kernel and the field-counteracting potential for molecules in an electric field, *Phys. Rev. A* **64** (2001) 042506.

[139] L. Gross, Logarithmic Sobolev inequalities, *Amer. J. Math.* **97**(4) (1975) 1061–1083.

[140] H. Grübmuller, Predicting slow structural transitions in macromolecular systems - conformational flooding, *Phys. Rev. E* **52**(3) (1995) 2893–2906.

[141] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, Generalized Verlet algorithm for efficient molecular dynamics simulations with long range interaction, *Mol. Simulat.* **6** (1991) 121–142.

[142] X. Guerrault, B. Rousseau, and J. Farago, Dissipative particle dynamics simulations of polymer melts. I. Building potential of mean force for polyethylene and *cis*-polybutadiene, *J. Chem. Phys.* **121**(13) (2004) 6538–6546.

[143] A. Guionnet and B. Zegarlinski, Lectures on logarithmic Sobolev inequalities, In *Séminaire de Probabilités XXXVI*, volume 1801 of *Lecture Notes in Mathematics* (Springer Verlag, 2003), pp. 1–134.

[144] I. Gyöngy, Mimicking the one-dimensional marginal distributions of processes having an Ito differential, *Probab. Th. Rel. Fields* **71** (1986) 501–516.

[145] E. Hairer, C. Lubich, and G. Wanner, Geometric numerical integration illustrated by the Störmer-Verlet method, *Acta Numerica* **12** (2003) 399–450.

[146] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics* (Springer-Verlag, Berlin, Heidelberg, 2006).

[147] M. Hairer, A. M. Stuart, J. Voss, and P. Wiberg, Analysis of SPDEs arising in path sampling. Part 1: The gaussian case, *Comm. Math. Sci* **3** (2005) 587–603.

[148] O. H. Hald and R. Kupferman, Asymptotic and numerical analyses for mechanical models of heat baths, *J. Stat. Phys.* **106** (2002) 1121–1184.

[149] G. G. Hall, The molecular orbital theory of chemical valency. VIII. A method of calculating ionization potentials, *Proc. Roy. Soc. A* **205** (1951) 541–552.

[150] S. Hampton, P. Brenner, A. Wenger, S. Chatterjee, and J.A. Izaguirre, Biomolecular sampling: Algorithms, test molecules, and metrics, In *New Algorithms for Macromolecular Simulation*, B. Leimkuhler, C. Chipot, R. Elber, A. Laaksonen, A. Mark, T. Schlick, C. Schuette, and R. Skeel (Eds.), volume 49 of *Lecture Notes in Computational Science and Engineering* (Springer Verlag, Berlin and New York, 2006), pp. 103–123.

[151] C. Hartmann and Ch. Schütte, A constrained Hybrid Monte-Carlo algorithm and the problem of calculating the free energy in several variables, *Z. Angew. Math. Mech.* **85**(10) (2005) 700–710.

[152] R. Z. Has'minskii, *Stochastic Stability of Differential Equations* (Sijthoff and Noordhoff, 1980).

[153] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57** (1970) 97–109.

[154] A.J. Heim, N. Grønbech-Jensen, T. C. Germann, E. M., Kober, B. L. Holian, and P. S. Lomdahl, The influence of interatomic bonding potentials on detonation properties, *Arxiv preprint / cond-mat* **0601106** (2006).

[155] R. L. Henderson, A uniqueness theorem for fluid pair correlation functions, *Phys. Lett. A* **49**(3) (1974) 197–198.

[156] D. A. Hendrix and C. Jarzynski, A "fast growth" method of computing free energy differences, *J. Chem. Phys.* **114**(14) (2001) 5974–5981.

[157] J. Hénin and C. Chipot, Overcoming free energy barriers using unconstrained molecular dynamics simulations, *J. Chem. Phys.* **121**(7) (2004) 2904–2914.

[158] G. Henkelman and H. Jonsson, Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table, *J. Chem. Phys.* **115**(21) (2001) 9657–9666.

[159] F. Hérau and F. Nier, Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential, *Arch. Ration. Mech. Anal.* **171** (2004) 151–218.

[160] J. Hietarinta, T. Kuusela, and B. Malomed, Shock waves in the dissipative Toda lattice, *J. Phys. A* **28** (1995) 3015–3024.

[161] P. Hohenberg and W. Kohn, Inhomogeneous electron gas, *Phys. Rev. B* **136** (1964) 864–871.

[162] B. L. Holian, Atomistic computer simulations of shock waves, *Shock Waves* **5** (1995) 149–157.

[163] B. L. Holian, Formulating mesodynamics for polycrystalline materials, *Europhys. Lett.* **64**(3) (2003) 330–336.

[164] B. L. Holian, H. Flaschka, and D. W. McLaughlin, Shock waves in the Toda lattice: Analysis, *Phys. Rev. A* **24**(5) (1981) 2595–2623.

[165] B. L. Holian, W. G. Hoover, and H. A. Posch, Resolution of Loschmidt paradox - the origin of irreversible behavior in reversible atomistic dynamics, *Phys. Rev. Lett.* **59**(1) (1987) 10–13.

[166] B. L. Holian and G .K. Straub, Molecular dynamics of shock waves in one-dimensional chains, *Phys. Rev. B* **18**(4) (1978) 1593–1608.

[167] B. L. Holian and G. K. Straub, Molecular dynamics of shock waves in three dimensional solids, *Phys. Rev. Lett.* **43** (1979) 1598.

[168] B. L. Holian, G. K. Straub, and R. G Petscheck, Molecular dynamics of shock waves in one-dimensional chains. II. Thermalization, *Phys. Rev. B* **19**(8) (1979) 4049–4055.

[169] R. Holley and D. Stroock, Logarithmic Sobolev inequalities and stochastic Ising models, *J. Stat. Phys.* **46**(5-6) (1987) 1159–1194.

[170] P. J. Hoogerbrugge and J. M. V. A. Koelman, Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics, *Europhys. Lett.* **19**(3) (1992) 155–160.

[171] W. G. Hoover, Canonical dynamics - Equilibrium phase-space distributions, *Phys. Rev. A* **31**(3) (1985) 1695–1697.

[172] F. C. Hoppensteadt, M. Rahman, and B. D. Welfert, $\sqrt{n}$-central limit theorems for Markov processes with applications to circular processes, *Preprint version available at the URL* `http://math.asu.edu/~bdw/PAPERS/CLT.pdf` (2003).

[173] A. M. Horowitz, A generalized guided Monte-Carlo algorithm, *Phys. Lett. B* **268**(2) (1991) 247–252.

[174] K. Hukushima and Y. Iba, Population annealing and its application to a spin glass, *AIP Conference Proceedings* **690**(1) (2003) 200–206.

[175] R. J. Hulse, R.L. Howley, and W.V. Wilding, Transient nonequilibrium molecular dynamic simulations of thermal conductivity. I. Simple fluids, *Int. J. Thermophys.* **26**(1) (2005) 1–12.

[176] G. Hummer, Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations, *New J. Phys.* **7** (2005) 34.

[177] G. Hummer and A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, *Proc. Nat. Acad. Sci. USA* **98**(7) (2001) 3658–3661.

[178] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: fifty years after kramers, *Rev. Mod. Phys.* **62**(2) (1990) 251–341.

[179] M. Iannuzzi, A. Laio, and M. Parrinello, Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics, *Phys. Rev. Lett.* **90**(23) (2003) 238302.

[180] Y. Iba, Extended ensemble Monte Carlo, *Int. J. Modern Phys. C* **12**(5) (2001) 623–656.

[181] W. Im, S. Berneche, and B. Roux, Generalized solvent boundary potential for computer simulations, *J. Chem. Phys.* **114**(7) (2001) 2924–2937.

[182] W. Im, S. Seefeld, and B. Roux, A Grand Canonical Monte Carlo-Brownian dynamics algorithm for simulating ion channels, *Biophys. J.* **79**(2) (2000) 788–801.

[183] J. A. Izaguirre, D. P. Catarello, J. M. Wozniak, and R. D. Skeel, Langevin stabilization of molecular dynamics, *J. Chem. Phys.* **114**(5) (2001) 2090–2098.

[184] J. A. Izaguirre and S. S. Hampton, Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules, *J. Comput. Phys.* **200**(2) (2004) 581–604.

[185] A. F. Izmaylov, V. N. Staroverov, G. Scuseria, E. R. Davidson, G. Stoltz, and E. Cancès, The effective local potential method: Implementation for molecules and relation to approximate optimized effective potential techniques, *J. Chem. Phys.* **126** (2007) 084107.

[186] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, *Phys. Rev. E* **56**(5) (1997) 5018–5035.

[187] C. Jarzynski, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.* **78**(14) (1997) 2690–2693.

[188] T. Juhasz and D. A. Mazziotti, Perturbation theory corrections to the two-particle reduced density matrix variational method, *J. Chem. Phys.* **121**(3) (2004) 1201–1205.

[189] J. Juraszek and P. G. Bolhuis, Sampling the multiple folding mechanisms of Trp-cage in explicit solvent, *Proc. Nat. Acad. Sci. USA* **103**(43) (2006) 15859–15864.

[190] T. Kato, *Perturbation theory for linear operators* (Springer, Berlin, 1980).

[191] A. D. Kennedy and B. Pendleton, Cost of the generalised hybrid Monte Carlo algorithm for free field theory, *Nuclear Phys. B* **607**(3) (2001) 456–510.

[192] G. King and A. Warshel, A surface constrained all-atom solvent model for effective simulations of polar solutions, *J. Chem. Phys.* **91**(6) (1989) 3647–3661.

[193] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science* **220**(4598) (1983) 671–680.

[194] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**(5) (1935) 300–313.

[195] W. Kohn and L. J. Sham, Self-consistent equations including exchange and correlation effects, *Phys. Rev.* **140** (1965) A1133–A1138.

[196] D. I. Kopelevich, A. Z. Panagiotopoulos, and I. G. Kevrekidis, Coarse-grained kinetic computations for rare events: application to micelle formation, *J. Chem. Phys.* **122** (2005) 044908.

[197] H. Kummer, N-representability problem for reduced density matrices, *J. Math. Phys.* **8**(10) (1967) 2063–2081.

[198] R. Kupferman and A. M. Stuart, Fitting SDE models to nonlinear Kac-Zwanzig heat bath models, *Physica D* **199** (2004) 279–316.

[199] R. Kupferman, A. M. Stuart, J. R. Terry, and P. F. Tupper, Long-term behaviour of large mechanical systems with random initial data, *Stochastics and Dynamics* **2**(4) (2002) 1–30.

[200] B. Lapeyre, E. Pardoux, and R. Sentis, *Méthodes de Monte Carlo pour les équations de transport et de diffusion*, volume 29 of *Mathématiques et applications* (Springer, 1998).

[201] J. L. Lebowitz and H. Spohn, Transport properties of the Lorentz gas - Fourier's law, *J. Stat. Phys.* **19**(6) (1978) 633–654.

[202] M. Ledoux, Logarithmic Sobolev inequalities for unbounded spin systems revisited, In *Séminaire de Probabilités XXXV*, volume 1755 of *Lecture Notes in Mathematics* (Springer Verlag, 2001), pp. 167–194.

[203] F. Legoll, *PhD thesis* (Université Paris VI, 2004).

[204] F. Legoll, M. Luskin, and R. Moeckel, Non-ergodicity of the Nosé-Hoover thermostatted harmonic oscillator, *Arch. Rational Mech. Anal.* **184** (2007) 449–463.

[205] B. J. Leimkuhler and S. Reich, *Simulating Hamiltonian dynamics*, volume 14 of *Cambridge monographs on applied and computational mathematics* (Cambridge University Press, 2005).

[206] B. J. Leimkuhler and C. R. Sweet, A Hamiltonian formulation for recursive multiple thermostats in a common timescale, *SIAM J. Appl. Dyn. Syst.* **4**(1) (2005) 187–216.

[207] T. Lelièvre, F. Otto, M. Rousset, and G. Stoltz, Long-time convergence of the Adaptive Biasing Force method, *in preparation* (2007).

[208] M. Lewin, Solutions of the multiconfiguration equations in quantum chemistry, *Arch. Ration. Mech. Anal.* **171**(1) (2004) 83–114.

[209] X. Li and W. E, Multiscale modeling of the dynamics of solids at finite temperature, *J. Mech. Phys. Sol.* **53** (2005) 1650–1685.

[210] E. H. Lieb, Density functional theory for Coulomb systems, *Int. J. Quant. Chem.* **24** (1983) 243–277.

[211] E. H. Lieb and B. Simon, The Hartree-Fock theory for Coulomb systems, *Commun. Math. Phys.* **53** (1977) 185–194.

[212] K. Lindenberg and V. Seshadri, Dissipative contributions of internal multiplicative noise. I. Mechanical oscilator, *Physica A* **109** (1981) 483–499.

[213] J.-L. Lions, Y. Maday, and G. Turinici, Résolution d'EDP par un schéma en temps "pararéel", *C. R. Acad. Sci., Paris, Sér. I, Math.* **332**(7) (2001) 661–668.

[214] P.-L. Lions, Solutions of Hartree-Fock equations for Coulomb systems, *Commun. Math. Phys.* **109** (1987) 33–97.

[215] J. S. Liu, *Monte Carlo strategies in Scientific Computing*, Springer Series in Statistics (Springer, 2001).

[216] C. Lo and B. Palmer, Alternative Hamiltonian for molecular dynamics simulations in the grand canonical ensemble, *J. Chem. Phys.* **102** (1995) 925–931.

[217] L. B. Lucy, Numerical approach to testing of fission hypothesis, *Astron. J.* **82**(12) (1977) 1013–1024.

[218] M. Lupkowski and F. van Swol, Ultrathin films under shear, *J. Chem. Phys.* **95** (1991) 1995–1998.

[219] A. P. Lyubartsev, M. Karttunen, P. Vattulainen, and A. Laaksonen, On coarse-graining by the inverse Monte Carlo method: Dissipative particle dynamics simulations made to a precise tool in soft matter modeling, *Soft Materials* **1**(1) (2003) 121–137.

[220] P.O. Löwdin, Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configuration interaction, *Phys. Rev.* **97**(6) (1955) 1474–1489.

[221] P. B. Mackenzie, An improved Hybrid Monte-Carlo method, *Phys. Lett. B* **226**(3-4) (1989) 369–371.

[222] J.-B. Maillet, L. Soulard, and G. Stoltz, A reduced model for shock and detonation waves. II. The reactive case, *Europhys. Lett.* **78**(6) (2007) 68001.

[223] M. J. Mandell, Properties of a periodic fluid, *J. Stat. Phys.* **15**(4) (1976) 299–305.

[224] X. Mao, *Stochastic differential equations and applications* (Horwood, Chichester, 1997).

[225] E. Marinari and G. Parisi, Simulated tempering - a new Monte-Carlo scheme, *Europhys. Lett.* **19**(6) (1992) 451–458.

[226] J. E. Marsden and M. West, Discrete mechanics and variational integrators, *Acta Numerica* **10** (2001) 357–514.

[227] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino, Self-healing umbrella sampling: A non-equilibrium approach for quantitative free energy calculations, *J. Phys. Chem. B* **110**(29) (2006) 14011–14013.

[228] M. G. Martin and J. I. Siepmann, Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes, *J. Phys. Chem. B* **102**(14) (1998) 2569–2577.

[229] G. J. Martyna, M. L. Klein, and M. Tuckerman, Nosé-Hoover chains - the canonical ensemble via continuous dynamics, *J. Chem. Phys.* **97**(4) (1992) 2635–2643.

[230] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Explicit reversible integrators for extended systems dynamics, *Mol. Phys.* **87**(5) (1996) 1117–1157.

[231] J. C. Mattingly, A. M. Stuart, and D. J. Higham, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, *Stoch. Proc. Appl.* **101**(2) (2002) 185–232.

[232] J. E. Mayer, Electron correlation, *Phys. Rev.* **100**(6) (1955) 1579–1586.

[233] D. A. Mazziotti, Variational minimization of atomic and molecular ground-state energies via the two-particle reduced density matrix, *Phys. Rev. A* **65**(6) (2002) 062511.

[234] D. A. Mazziotti, First-order semidefinite programming for the direct determination of two-electron reduced density matrices with application to many-electron atoms and molecules, *J. Chem. Phys.* **121**(22) (2004) 10957–10966.

[235] D. A. Mazziotti, Realization of quantum chemistry without wave functions through first-order semidefinite programming, *Phys. Rev. Lett.* **93**(21) (2004) 213001.

[236] D. A. Mazziotti, Variational two-electron reduced density matrix theory for many-electron atoms and molecules: Implementation of the spin- and symmetry-adapted T-2 condition through first-order semidefinite programming, *Phys. Rev. A* **72**(3) (2005) 032510.

[237] K. L. Mengersen and R. L. Tweedie, Rates of convergence of the Hastings and Metropolis algorithms, *Ann. Stat.* **24**(1) (1996) 101–121.

[238] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**(6) (1953) 1087–1091.

[239] S. P. Meyn and R. L. Tweedie, Stability of markovian processes. I. Criteria for discrete-time chains., *Adv. Appl. Probab.* **24** (1992) 542–574.

[240] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Communications and control engineering series (Springer-Verlag, London, New York, 1993).

[241] S. P. Meyn and R. L. Tweedie, Stability of markovian processes. II. Continuous-time processes and sampled chains, *Adv. Appl. Probab.* **25** (1993) 487–517.

[242] G. N. Milstein and M. V. Tretyakov, Quasi-symplectic methods for Langevin-type equations, *IMA J. Numer. Anal.* **23**(4) (2003) 593–626.

[243] R. A. Miron and K. A. Fichthorn, Accelerated molecular dynamics with the bond-boost method, *J. Chem. Phys.* **119**(12) (2003) 6210–6216.

[244] B. Mishra and T. Schlick, The notion of error in Langevin dynamics. I. Linear analysis, *J. Chem. Phys.* **105**(1) (1996) 299–318.

[245] V. Molinero and W.A. Goddard III, M3B: A coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures, *J. Phys. Chem. B* **108** (2004) 1414–1427.

[246] J. J. Monaghan, Smoothed particle hydrodynamics, *Ann. Rev. Astron. Astrophys.* **30** (1992) 543–574.

[247] P. Del Moral, *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications*, Springer Series Probability and its Applications (Springer, 2004).

[248] P. Del Moral and L. Miclo, Branching and interacting particle systems approximations of feynman-kac formulae with applications to nonlinear filtering, *Lecture notes in Mathematics* **1729** (2000) 1–145.

[249] J. J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. Fr.* **93** (1965) 273–299.

[250] H. Mori, Transport, collective motion, and Brownian motion, *Prog. Theor. Phys.* **33** (1965) 423–450.

[251] F. Müller-Plathe, A simple nonequilibrium molecular dynamics method for calculating the thermal conductivity, *J. Chem. Phys.* **106**(14) (1997) 6082–6085.

[252] M. Nakata, M. Ehara, and H. Nakatsuji, Density matrix variational theory: Application to the potential energy surfaces and strongly correlated systems, *J. Chem. Phys.* **116**(13) (2002) 5432–5439.

[253] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, Variational calculations of fermion second-order reduced density matrices by semidefinite programming algorithm, *J. Chem. Phys.* **114**(19) (2001) 8282–8292.

[254] H. Nakatsuji, Scaled Schrödinger equation and the exact wave function, *Phys. Rev. Lett.* **93**(3) (2004) 030403.

[255] H. Nakatsuji, General method of solving the Schrödinger equation of atoms and molecules, *Phys. Rev. A* **72**(6) (2005) 062110.

[256] R. M. Neal, An improved acceptance procedure for the Hybrid Monte-Carlo algorithm, *J. Comput. Phys.* **111**(1) (1994) 194–203.

[257] N. Niederreiter, *Random Number Generation and Quasi Monte-Carlo Methods* (Society for Industrial and Applied Mathematics, 1992).

[258] G. E. Norman and V. S. Filinov, Investigation of phase transitions by a monte-carlo method, *High Temp. (USSR)* **7** (1969) 216–222.

[259] S. Nosé, A molecular-dynamics method for simulations in the canonical ensemble, *Mol. Phys.* **52**(2) (1984) 255–268.

[260] S. Nosé, A unified formulation of the constant temperature molecular-dynamics methods, *J. Chem. Phys.* **81**(1) (1984) 511–519.

[261] H. Oberhofer, C. Dellago, and P. L. Geissler, Biased sampling of nonequilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods?, *J. Phys. Chem. B* **109**(14) (2005) 6902–6915.

[262] H. C. Ottinger, Brownian dynamics of rigid polymer-chains with hydrodynamic interactions, *Phys. Rev. E* **50**(4) (1994) 2696–2701.

[263] F. Otto and M. G. Reznikoff, A new criterion for the logarithmic Sobolev inequality, *J. Funct. Anal.* **243** (2007) 121–157.

[264] F. Otto and C. Villani, Generalization of an inequality by Talagrand, viewed as a consequence of the logarithmic Sobolev inequality, *J. Funct. Anal.* **173**(2) (2000) 361–400.

[265] G. Pagès, Sur quelques algorithmes récursifs pour les probabilités numériques, *ESAIM: Probability and Statistics* **5** (2001) 141–170.

[266] G. C. Papanicolaou, Some probabilistic problems and methods in singular perturbations, *Rocky Mountain J. Math.* **6**(4) (1976) 653–674.

[267] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality, *J. Chem. Phys.* **119**(6) (2003) 3559–3566.

[268] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization* (`http://www.maths.warwick.ac.uk/~stuart/book.pdf`, 2007).

[269] M. Peyrard, S. Odiot, E. Lavenir, and J. M. Schnur, Molecular-model for cooperative propagation of shock-induced detonations in energetic solids and its application to nitromethane, *J. Appl. Phys.* **57**(7) (1985) 2626–2636.

[270] G. D. J. Phillies, *Elementary Lectures in Statistical Mechanics* (Springer, 2000).

[271] Y. Pokern, A. M. Stuart, and P. Wiberg, Parameter estimation for partially observed hypo-elliptic diffusions, *submitted to J. Roy. Stat. Soc.* (2006).

[272] L. R. Pratt, A statistical-method for identifying transition-states in high dimensional problems, *J. Chem. Phys.* **85**(9) (1986) 5045–5048.

[273] L. R. Pratt and S. W. Haan, Effects of periodic boundary-conditions on equilibrium properties of computer-simulated fluids. I. Theory, *J. Chem. Phys.* **74**(3) (1981) 1864–1872.

[274] L. R. Pratt and S. W. Haan, Effects of periodic boundary-conditions on equilibrium properties of computer-simulated fluids. II. Application to simple liquids, *J. Chem. Phys.* **74**(3) (1981) 1873–1876.

[275] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B* **110**(8) (2006) 3533–3539.

[276] D. C. Rapaport, *The Art of Molecular Dynamics Simulations* (Cambridge University Press, 1995).

[277] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, volume I - IV (Academic Press, 1975-190).

[278] S. Reich, Backward error analysis for numerical integrators, *Siam J. Numer. Anal.* **36**(5) (1999) 1549–1570.

[279] D. Reith, M. Pütz, and F. Müller-Plathe, Deriving effective mesoscale potentials from atomistic simulations, *J. Comput. Chem.* **24** (2003) 1624–1636.

[280] A. Ricci and G. Ciccotti, Algorithms for Brownian dynamics, *Mol. Phys.* **101**(12) (2003) 1927–1931.

[281] J. M. Rickman and R. LeSar, Free-energy calculations in materials research, *Ann. Rev. Mater. Res.* **32** (2002) 195–217.

[282] M. Ripoll and P. Espanol, Dissipative particle dynamics with energy conservation: Heat conduction, *Int. J. Modern Phys. C* **9**(8) (1998) 1329–1338.

[283] G. Roberts and R. Tweedie, Exponential convergence of Langevin diffusions and their discrete approximations, *Bernoulli* **2** (1997) 314–363.

[284] G. O. Roberts and J. S. Rosenthal, Optimal scaling of discrete approximations to Langevin diffusions, *J. Roy. Stat. Soc. B* **60** (1998) 255–268.

[285] G. O. Roberts and R. L. Tweedie, Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms, *Biometrika* **83**(1) (1996) 95–110.

[286] D. Rodriguez-Gomez, E. Darve, and A. Pohorille, Assessing the efficiency of free energy calculation methods, *J. Chem. Phys.* **120**(8) (2004) 3563–3578.

[287] L. C. G. Rogers, Smooth transition densities for one-dimensional probabilities, *Bull. London Math. Soc* **17** (1985) 157–161.

[288] C. C. J. Roothan, New developments in molecular orbital theory, *Rev. Mod. Phys.* **23** (1951) 69–89.

[289] M. Rousset, On the control of an interacting particle estimation of Schrödinger ground-states, *SIAM J. Math. Anal.* **38** (2006) 824–844.

[290] M. Rousset, *PhD Thesis* (Université Paul Sabatier, Toulouse III, 2006).

[291] M. Rousset and G. Stoltz, An interacting particle system approach for molecular dynamics, *CERMICS Report* **283** (2005).

[292] M. Rousset and G. Stoltz, Equilibrium sampling from nonequilibrium dynamics, *J. Stat. Phys.* **123**(6) (2006) 1251–1272.

[293] D. Ruelle, *Statistical mechanics: rigorous results* (Benjamin, New York, 1969).

[294] J. P. Ryckaert and A. Bellemans, Molecular-dynamics of liquid alkanes, *Faraday Discussions* **66** (1978) 95–106.

[295] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes, *J. Comput. Phys.* **23**(3) (1977) 327–341.

[296] R. J. Sadus, *Molecular Simulation of Fluids* (Elsevier, 1999).

[297] F. Della Sala and A. Görling, Efficient localized Hartree-Fock methods as effective exact-exchange Kohn-Sham methods for molecules, *J. Chem. Phys.* **115**(13) (2001) 5718–5731.

[298] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès, and M. Caffarel, An efficient sampling algorithm for Variational Monte-Carlo, *J. Chem. Phys.* **125**(11) (2006).

[299] T. Schlick, *Molecular Modeling and Simulation* (Springer, 2002).

[300] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, General atomic and molecular electronic-structure system, *J. Comput. Chem.* **14**(11) (1993) 1347–1363.

[301] C. Schütte, *Habilitation Thesis* (Freie Universität Berlin, 1999).

[302] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, A direct approach to conformational dynamics based on Hybrid Monte-Carlo, *J. Comput. Phys.* **151**(1) (1999) 146–168.

[303] C. Schütte and W. Huisinga, Biomolecular conformations can be identified as metastable sets of molecular dynamics, In *Handbook of Numerical Analysis (Special volume on computational chemistry)*, P. G. Ciarlet and C. L. Bris (Eds.), volume X (Elsevier, 2003), pp. 699–744.

[304] G. E. Scuseria and V. N. Staroverov, Progress in the development of exchange-correlation functionals, In *Theory and Applications of Computational Chemistry: The First Forty Years*, C. E. Dykstra, G. Frenking, K. S. Kim, and G. E. Scuseria (Eds.) (Elsevier, Amsterdam, 2005), pp. 669–724.

[305] M. Serrano, G. De Fabritiis, P. Espanol, and P. V. Coveney, A stochastic Trotter integration scheme for dissipative particle dynamics, *Mathematics Computers In Simulation* **72**(2-6) (2006) 190–194.

[306] T. Shardlow, Splitting for dissipative particle dynamics, *SIAM J. Sci. Comp.* **24**(4) (2003) 1267–1282.

[307] T. Shardlow and Y. B. Yan, Geometric ergodicity for dissipative particle dynamics, *Stoch. Dynam.* **6**(1) (2006) 123–154.

[308] R. T. Sharp and G. K. Horton, A variational approach to the unipotential many-electron problem, *Phys. Rev.* **90** (1953) 317.

[309] Y. Shim and J. G. Amar, Rigorous synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth, *Phys. Rev. B* **71**(11) (2005).

[310] R. D. Skeel, In *The graduate student's guide to numerical analysis*, M. Ainsworth, J. Levesley, and M. Marletta (Eds.), Springer Series in Computational Mathematics (Springer-Verlag, 1999), pp. 119–176.

[311] R. D. Skeel and J. A. Izaguirre, An impulse integrator for Langevin dynamics, *Mol. Phys.* **100**(24) (2002) 3885–3891.

[312] J. C. Slater, A simplification of the Hartree-Fock method, *Phys. Rev.* **81** (1951) 385–390.

[313] L. I. Slepyan, Dynamics of a crack in a lattice, *Sov. Phys. Dokl.* **26**(5) (1981) 538–540.

[314] L. I. Slepyan, Dynamic factor in impact, phase transition and fracture, *J. Mech. Phys. Solids* **48** (2000) 927–960.

[315] D. Smets and M. Willem, Solitary waves with prescribed speed on infinite lattices, *J. Funct. Anal.* **149** (1997) 266–275.

[316] P. Soderlind, J. A. Moriarty, and J. M. Wills, First-principles theory of iron up to earth-core pressures: Structural, vibrational, and elastic properties, *Phys. Rev. B* **53**(21) (1996) 14063–14072.

[317] M. R. Sorensen and A. F. Voter, Temperature-accelerated dynamics for simulation of infrequent events, *J. Chem. Phys.* **112**(21) (2000) 9599–9606.

[318] H. Spohn, Kinetic equations from Hamiltonian dynamics: Markovian limits, *Rev. Mod. Phys.* **53**(3) (1980) 569–615.

[319] H. Spohn, *Large scale dynamics of interacting particles* (Springer, New York, 1991).

[320] M. Sprik and G. Ciccotti, Free energy from constrained molecular dynamics, *J. Chem. Phys.* **109**(18) (1998) 7737–7744.

[321] V. N. Staroverov, G. Scuseria, and E.R. Davidson, Optimized effective potentiels yielding Hartree-Fock energies and densities, *J. Chem. Phys.* (2006).

[322] M. L. Stedman, W. M. C. Foulkes, and M. Nekovee, An accelerated Metropolis method, *J. Chem. Phys.* **109**(7) (1998) 2630–2634.

[323] G. Stoltz, Shock waves in an augmented one-dimensional atom chain, *Nonlinearity* **18**(5) (2005) 1967–1985.

[324] G. Stoltz, A reduced model for shock and detonation waves. I. The inert case, *Europhys. Lett.* **76**(5) (2006) 849–855.

[325] G. Stoltz, Path sampling with stochastic dynamics: Some new algorithms, *J. Comput. Phys.* **225** (2007) 491–508.

[326] A. Strachan and B. L. Holian, Energy exchange between mesoparticles and their internal degrees of freedom, *Phys. Rev. Lett.* **94**(1) (2005) 014301.

[327] A. Strachan, A. C. T. van Duin, D. Chakraborty, S. Dasgupta, and W. A. Goddard, Shock waves in high-energy materials: The initial chemical events in nitramine RDX, *Phys. Rev. Lett.* **91**(9) (2003) 098301.

[328] O. Stramer and R. L. Tweedie, Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients, *Statistica Sinica* **7** (1997) 577–593.

[329] J. E. Straub, M. Borkovec, and B. J. Berne, Molecular-dynamics study of an isomerizing diatomic in a Lennard-Jones fluid, *J. Chem. Phys.* **89**(8) (1988) 4833–4847.

[330] A.M. Stuart, J. Voss, and P. Wiberg, Conditional path sampling of SDEs and the Langevin MCMC method, *Commun. Math. Sci.* **2**(4) (2004) 685–697.

[331] S. X. Sun, Equilibrium free energies from path sampling of nonequilibrium trajectories, *J. Chem. Phys.* **118**(13) (2003) 5769–5775.

[332] Z. Sun, M. M. Soto, and W. A. Lester Jr., Characteristics of electron movement in variational Monte Carlo simulations, *J. Chem. Phys.* **100**(2) (1994) 1278–1289.

[333] C. R. Sweet, *PhD Thesis* (University of Leicester, 2004).

[334] E. B. Tadmor, M. Ortiz, and R. Phillips, Quasicontinuum analysis of defects in solids, *Phil. Mag. A* **73** (1996) 1529–1563.

[335] D. Talay, Second-order discretization schemes of stochastic differential systems for the computation of the invariant law, *Stochastics and Stochastic Reports* **29** (1990) 13–36.

[336] D. Talay, Approximation of invariant measures of nonlinear hamiltonian and dissipative stochastic differential equations, *Publication du L.M.A.-C.N.R.S.* **152** (1999) 139–169.

[337] D. Talay, Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme, *Markov Proc. Rel. Fields* **8** (2002) 163–198.

[338] J. D. Talman and W. F. Shadwick, Optimized effective atomic central potential, *Phys. Rev. A* **14**(1) (1976) 36–40.

[339] R. Tehver, F. Toigo, J. Koplik, and J. R. Banavar, Thermal walls in computer simulations, *Phys. Rev. E* **57**(1) (1998) R17–R20.

[340] A. Tenenbaum, G. Ciccotti, and R. Gallico, Stationary non-equilibrium states by molecular-dynamics - Fourier law, *Phys. Rev. A* **25**(5) (1982) 2778–2787.

[341] J. Tersoff, Modeling solid-state chemistry: Interatomic potentials for multicomponent systems, *Phys. Rev. B* **39** (1989) 5566–5568.

[342] S. Teufel, *Adiabatic perturbation theory in quantum dynamics*, volume 1821 of *Lecture Notes in Mathematics* (Springer-Verlag, Berlin, Heidelberg, New York, 2003).

[343] J. Thouvenin, *Détonique*, Collection du Commissariat à l'Energie Atomique (Eyrolles, 1997).

[344] M. Toda, *Theory of Nonlinear Lattices* (Springer, 1981).

[345] G. M. Torrie and J. P. Valleau, Non-physical sampling distributions in Monte-Carlo free-energy estimation - Umbrella sampling, *J. Comput. Phys.* **23**(2) (1977) 187–199.

[346] M. E. Tuckerman and G. J. Martyna, Understanding modern molecular dynamics: Techniques and applications, *J. Phys. Chem. B* **104**(2) (2000) 159–178.

[347] M. E. Tuckermann, B. J. Berne, and G. J. Martyna, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* **97** (1992) 1990–2001.

[348] R. L. Tweedie, Topological conditions enabling the use of Harris methods in discrete and continuous time, *Acta Appl. Math.* **34** (1994) 175–188.

[349] R. L. Tweedie, Markov chains: Structure and applications, In *Handbook of Statistics*, D. N. Shanbhag and C. R. Rao (Eds.), volume 19 (North-Holland/Elsevier, 2001), pp. 817–851.

[350] C. J. Umrigar, Accelerated Metropolis method, *Phys. Rev. Lett.* **71**(3) (1993) 408–411.

[351] C. J. Umrigar and C. Filippi, Energy and variance optimization of many-body wave functions, *Phys. Rev. Lett.* **94**(15) (2005) 150201.

[352] C. J. Umrigar, M. P. Nightingale, and K. J. Runge, A diffusion monte-carlo algorithm with very small time-step errors, *J. Chem. Phys.* **99**(4) (1993) 2865–2890.

[353] A. C. T. van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard III, ReaxFF: A reactive force field for hydrocarbons, *J. Phys. Chem. A* **105** (2001) 9396–9409.

[354] T. S. van Erp, Efficiency analysis of reaction rate calculation methods using analytical models. I. The two-dimensional sharp barrier, *J. Chem. Phys.* **125**(17) (2006) 174106.

[355] T. S. van Erp and P. G. Bolhuis, Elaborating transition interface sampling methods, *J. Comput. Phys.* **205**(1) (2005) 157–181.

[356] T. S. van Erp, D. Moroni, and P. G. Bolhuis, A novel path sampling method for the calculation of rate constants, *J. Chem. Phys.* **118**(17) (2003) 7762–7774.

[357] E. Vanden-Eijnden and F. Tal, Transition state theory: Variational formulation, dynamical corrections, and error estimates, *J. Chem. Phys.* **123** (2005) 184103.

[358] W. F. VanGunsteren and H. J. C. Berendsen, Algorithms for brownian dynamics, *Mol. Phys.* **45**(3) (1982) 637–647.

[359] S. Venakides, P. Deift, and R. Oba, The Toda shock problem, *Comm. Pure Appl. Math.* **14** (1991) 1171–1242.

[360] L. Verlet, Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, *Phys. Rev.* **159** (1967) 98–103.

[361] H. J. Viljoen, L. L. Lauderback, and D. Sornette, Solitary waves and supersonic reaction front in metastable solids, *Phys. Rev. E* **65**(2) (2002) 026609.

[362] G. H. Vineyard, Frequency factors and isotope effects in solid state rate processes, *J. Phys. Chem. Solids* **3** (1957) 121–127.

[363] T. J. H. Vlugt and B. Smit, On the efficient sampling of pathways in the transition path ensemble, *PhysChemComm* **2** (2001) 2.

[364] A. F. Voter, Hyperdynamics: Accelerated molecular dynamics of infrequent events, *Phys. Rev. Lett.* **78**(20) (1997) 3908–3911.

[365] A. F. Voter, A method for accelerating the molecular dynamics simulation of infrequent events, *J. Chem. Phys.* **106**(11) (1997) 4665–4677.

[366] A. F. Voter, Parallel replica method for dynamics of infrequent events, *Phys. Rev. B* **57**(22) (1998) R13985–R13988.

[367] A. F. Voter, F. Montalenti, and T. C. Germann, Extending the time scale in atomistic simulation of materials, *Ann. Rev. Mater. Res.* **32** (2002) 321–346.

[368] F. G. Wang and D. P. Landau, Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram, *Phys. Rev. E* **64**(5) (2001) 056101.

[369] W. Wang and R. D. Skeel, Analysis of a few numerical integration methods for the Langevin equation, *Mol. Phys.* **101**(14) (2003) 2149–2156.

[370] E. Weinan, W. Q. Ren, and E. Vanden-Eijnden, Finite temperature string method for the study of rare events, *J. Phys. Chem. B* **109**(14) (2005) 6688–6693.

[371] E. P. Wigner, Effects of electron interaction on the energy levels of electrons in metals, *Trans. Faraday Soc.* **34** (1938) 678–685.

[372] H. J. Woo, A. R. Dinner, and B. Roux, Grand canonical Monte Carlo simulations of water in protein environments, *J. Chem. Phys.* **121**(13) (2004) 6392–6400.

[373] S. Yang, J. N. Onuchic, and H. Levine, Effective stochastic dynamics on a protein folding energy landscape, *J. Chem. Phys.* **125** (2006) 054910.

[374] F. M. Ytreberg and D. M. Zuckerman, Single-ensemble nonequilibrium path-sampling estimates of free energy differences, *J. Chem. Phys.* **120**(23) (2004) 10876–10879.

[375] E. Zeidler, *Nonlinear Functional Analysis and its Applications. I. Fixed-Point Theorems* (Springer, 1986).

[376] Z. J. Zhao, B. J. Braams, M. Fukuda, M. L. Overton, and J. K. Percus, The reduced density matrix method for electronic structure calculations and the role of three-index representability conditions, *J. Chem. Phys.* **120**(5) (2004) 2095–2104.

[377] G. Zhislin, Discussion of the spectrum of the Schrödinger operator for systems of many particles, *Tr. Mosk. Mat. Obs.* **9** (1960) 81–128.

[378] D. M. Zuckerman and T. B. Woolf, Systematic finite-sampling inaccuracy in free energy differences and other nonlinear quantities, *J. Stat. Phys.* **114**(5-6) (2004) 1303–1323.

[379] R. Zwanzig, Nonlinear generalized Langevin equations, *J. Stat. Phys.* **9** (1973) 215–220.

[380] R. W. Zwanzig, High-temperature equation of state by a perturbation method I. Nonpolar gases, *J. Chem. Phys.* **22**(8) (1954) 1420–1426.