



HAL
open science

Adaptive biasing algorithms: mathematical analysis and applications in molecular dynamics

Lise Maurin

► **To cite this version:**

Lise Maurin. Adaptive biasing algorithms: mathematical analysis and applications in molecular dynamics. Probability [math.PR]. Sorbonne Université, 2021. English. NNT: 2021SORUS345 . tel-03589173

HAL Id: tel-03589173

<https://theses.hal.science/tel-03589173>

Submitted on 25 Feb 2022

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.

SORBONNE UNIVERSITÉ
LJLL

Doctoral School **École Doctorale Sciences Mathématiques de Paris Centre**
University Department **Laboratoire Jacques-Louis Lions**

Thesis defended by **Lise MAURIN**

Defended on **16th December, 2021**

In order to become Doctor from Sorbonne Université

Academic Field **Applied mathematics**

Adaptive biasing algorithms: mathematical analysis and applications in molecular dynamics

Thesis supervised by Jean-Philip PIQUEMAL Supervisor
Tony LELIÈVRE Supervisor
Pierre MONMARCHÉ Co-Supervisor

Committee members

<i>Referees</i>	Julian TUGAUT	Associate Professor at Université Jean Monnet
	Benedict LEIMKUHNER	Professor at University of Edinburgh
<i>Examiners</i>	Virginie EHRLACHER	Associate Professor at École des Ponts Paris-Tech
	Yvon MADAY	Professor at Sorbonne Université
<i>Supervisors</i>	Jean-Philip PIQUEMAL	Professor at Sorbonne Université
	Tony LELIÈVRE	Professor at École des Ponts Paris-Tech
	Pierre MONMARCHÉ	Associate Professor at Sorbonne Université

Keywords: molecular dynamics, diffusion process, adaptive biasing algorithm, functional inequality, free energy, alchemical transition

Mots clés : dynamique moléculaire, processus de diffusion, méthode de biais adaptatif, inégalité fonctionnelle, énergie libre, transitions alchimique

This thesis has been prepared at

Laboratoire Jacques-Louis Lions

Sorbonne Université
Campus Pierre et Marie Curie
4 place Jussieu
75005 Paris
France

☎ +33 1 44 27 42 98
Web Site <https://ljl1.math.upmc.fr/>



À Marine, Léana, Océane, Jennifer et Damien.

La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles ;
L'homme y passe à travers des forêts de symboles
Qui l'observent avec des regards familiers.

Comme de longs échos qui de loin se confondent
Dans une ténébreuse et profonde unité,
Vaste comme la nuit et comme la clarté,
Les parfums, les couleurs et les sons se répondent.

Il est des parfums frais comme des chairs d'enfants,
Doux comme les hautbois, verts comme les prairies,
- Et d'autres, corrompus, riches et triomphants,

Ayant l'expansion des choses infinies,
Comme l'ambre, le musc, le benjoin et l'encens,
Qui chantent les transports de l'esprit et des sens.

Charles Baudelaire, "*Correspondances*"

ADAPTIVE BIASING ALGORITHMS: MATHEMATICAL ANALYSIS AND APPLICATIONS IN MOLECULAR DYNAMICS**Abstract**

This thesis is dedicated to the study of adaptive biasing algorithms for molecular dynamics simulations, from both theoretical and numerical perspectives. The goal of molecular dynamics is to obtain macroscopic informations about a system of particles, given its microscopic description. Adaptive biasing algorithms are powerful tools in molecular dynamics, especially when one needs to compute a system's free energy. We will mainly focus on the *Adaptive Biasing Force* algorithm, whose key idea is to bias the interaction force between the particles in order to enhance the sampling of the system's configuration space. In particular, we will study its robustness in the case of non-conservative interaction forces. We will then proceed to design an enhanced sampling algorithm in the scope of alchemical transitions, where the system's evolution from an initial state to a final state is indexed by a parameter λ in $[0, 1]$. Such transitions are often used in pharmacology, as they allow the estimation of several free energies, such as the binding free energy of a ligand with a receptor, or even the solvation free energy of a compound in solvent. When coupled to the λ -dynamics method, which deals with the dynamic evolution of the parameter λ , the *Orthogonal Space Random Walk* (OSRW) sampling method may permit a better and quicker sampling of the configuration space. Drawing inspiration from this algorithm, we will implement an adaptive biasing method coupled to the λ -dynamics, and compare it with the original OSRW algorithm. This work led to the implementation of a new interface between the Tinker-HP program and the Collective Variables module software.

Keywords: molecular dynamics, diffusion process, adaptive biasing algorithm, functional inequality, free energy, alchemical transition

Résumé

L'objet de cette thèse porte sur l'étude tant théorique que numérique de méthodes de biais adaptatifs pour la dynamique moléculaire. Etant donnée une description microscopique d'un système de particules, le but de la dynamique moléculaire est d'en déduire des informations macroscopiques. Les méthodes de biais adaptatifs se révèlent être un outil efficace, notamment lorsqu'il est question de déterminer l'énergie libre d'un système. Nous accorderons une attention toute particulière à la méthode de l'*Adaptive Biasing Force*, dont le principe est de biaiser la force d'interaction entre les particules afin d'accélérer l'échantillonnage de l'espace des configurations. En particulier, nous étudierons sa robustesse dans le cas où les forces d'interaction s'avèrent être non-conservatives. Nous procéderons ensuite à la construction d'un algorithme d'accélération d'échantillonnage dans le cadre de transitions alchimiques, où l'évolution du système d'un état initial à un état final prédéfinis est indexée par un paramètre λ compris entre 0 et 1. De telles transitions s'avèrent utiles en pharmacologie, leur étude permettant de déterminer différentes énergies libres, telles que l'énergie libre de liaison d'un ligand avec un récepteur donné ou bien encore l'énergie libre de solvation d'une molécule ou d'un ion dans un solvant. Couplée à la méthode de la λ -dynamique, qui traite l'évolution dynamique du paramètre λ , la méthode d'échantillonnage de l'*Orthogonal Space Random Walk* (OSRW) pourrait permettre une accélération de l'échantillonnage des configurations. En s'inspirant de cette dernière, nous implémenterons une méthode de biais adaptatif couplée à une méthode de λ -dynamique, que nous comparerons à la méthode de l'OSRW originelle. Ce travail d'implémentation permettra la mise en place d'une nouvelle interface entre le logiciel Tinker-HP et le module Collective Variables.

Mots clés : dynamique moléculaire, processus de diffusion, méthode de biais adaptatif, inégalité fonctionnelle, énergie libre, transitions alchimique

Laboratoire Jacques-Louis Lions

Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France

Contents

Abstract	xi
Contents	1
Remerciements	5
Résumé succinct en français	9
Constants and conventions	13
Fundamental physical constants	13
Notations	13
I General introduction	17
Introductory foreword	19
1 A quick introduction to statistical physics	23
1.1 A matter of scales	23
1.2 The equations of motion at the microscopic scale	25
1.3 The ergodic hypothesis	26
1.4 Different thermodynamic ensembles	26
The microcanonical ensemble (NVE)	27
1.4.1 The canonical ensemble (NVT)	27
1.4.2 The grand canonical ensemble (μVT)	28
1.4.3 Other thermodynamic ensembles	29
1.5 State functions and thermodynamic potentials	29
1.5.1 Definition and the example of entropy	29
1.5.2 Thermodynamic potentials	30
1.5.3 Free energy	30
2 A quick introduction to molecular dynamics	33
2.1 Settings	33
2.1.1 The Born-Oppenheimer approximation	33
2.1.2 Interaction potentials	35
2.1.3 Boundary conditions	36
2.2 Sampling the canonical measure	37
2.2.1 Langevin dynamics	38

2.2.2	Overdamped Langevin dynamics	39
2.2.3	Numerical schemes	40
2.2.3.1	Overdamped Langevin dynamics: Euler-Maruyama scheme	40
2.2.3.2	Langevin dynamics: SPV and BBK schemes	40
3	Free energy calculations and sampling methods	43
3.1	What is metastability?	44
3.1.1	Definition and examples	44
3.1.2	Quantifying metastability	47
3.1.2.1	Convergence of averages and convergence of the law	47
3.1.2.2	Relative entropy and logarithmic Sobolev inequality	47
3.1.2.3	Properties and criteria for logarithmic Sobolev inequalities	50
3.2	Transition coordinates: configurational and alchemical cases	50
3.2.1	Conformational transitions	51
3.2.2	Alchemical transitions	53
3.2.3	Choosing the reaction coordinate	56
3.3	Enhanced sampling methods	56
3.3.1	Umbrella sampling	57
3.3.2	Adaptive biasing methods	59
3.3.2.1	Importance sampling and flat histogram property	59
3.3.2.2	Metadynamics	61
3.3.2.3	Adaptive Biasing Potential method	63
3.3.2.4	Adaptive Biasing Force method	64
3.3.2.5	Projected Adaptive Biasing Force method	65
3.3.2.6	The non-conservative case	66
3.4	Free energy differences in the alchemical setting	66
3.4.1	Motivation	67
3.4.1.1	Ligand binding affinity	67
3.4.1.2	Solvation free energy	68
3.4.2	Available methods	68
3.4.2.1	Free Energy Perturbation	69
3.4.2.2	Thermodynamic Integration	69
3.4.2.3	Wheighted Histogram Analysis method	70
3.4.2.4	Bennett Acceptance Ratio method	71
3.4.3	λ -dynamics and Orthogonal Space Random Walk sampling methods	72
3.4.3.1	λ -dynamics	72
3.4.3.2	The Orthogonal Space Random Walk method	74
4	Contributions of the thesis	77
	Study of the robustness of the Adaptive Biasing Force method with non-conservative forces	77
	Study of the Orthogonal Space Random Walk sampling method in the case of alchemical transitions	79
II Robustness of the Adaptive Biasing Force method with non-conservative forces and related topics		81

5	Robustness of the Adaptive Biasing Force algorithm	83
5.1	Introduction	83
5.1.1	Setting	84
5.1.2	Metastability, reaction coordinate and free-energy profiles	84
5.1.3	The Adaptive Biasing Force method	85
5.1.4	The non-conservative case	87
5.2	Main results	89
5.2.1	Precise statements of the results	89
5.3	Law of the transition coordinate	93
5.3.1	Proof of Proposition 3	93
5.3.2	Proof of Proposition 4	94
5.4	Existence of a stationary measure	99
5.4.1	Preliminary estimates for homogeneous diffusions	99
5.4.2	Proof of Theorem 1	103
5.4.3	Proof of Proposition 5	105
5.5	Long-time convergence	108
5.5.1	Intermediate results	109
5.5.2	Proof of Theorem 2	111
5.5.3	Proof of Theorem 3	119
5.5.4	Proof of Corollary 1	121
5.6	What remains to be done	121

III Alchemical free energies: the Orthogonal Space Random Walk sampling method 123

6	Study of the OSRW sampling method in the case of alchemical transitions	125
6.1	What is the OSRW method?	126
6.1.1	Limitations of free energy differences computation methods	126
6.1.2	Definition of the OSRW method	128
6.1.2.1	Adding a second reaction coordinate	128
6.1.2.2	Possible limitations	128
6.2	The Tinker-HP and Collective Variables module softwares	130
6.2.1	Tinker-HP	130
6.2.2	The Collective Variables module	130
6.2.3	The Tinker-HP–Colvars interface	131
6.3	What is λ -dynamics?	132
6.3.1	Some reminders on λ -dynamics	132
6.3.2	A classical application: ligand binding affinity	133
6.3.3	Limitations of the current λ -dynamics	135
6.4	Implementation of the λ -dynamics with softcore potentials	138
6.4.1	Foreword	138
6.4.1.1	Treatment of the collective variable λ	138
6.4.1.2	Extended potential and list of needed derivatives	139
6.4.2	Softcores for van der Waals interactions	139
6.4.2.1	The Halgren potential	139
6.4.2.2	Generic softcore	140
6.4.2.3	Useful derivatives	143
6.4.3	Softcores for electrostatic interactions	143

6.4.3.1	Particle Mesh Ewald	143
6.4.3.2	First attempt: "hardcore" potentials	146
6.4.3.3	Second attempt: proper softcore potential	147
6.4.4	Tinker-HP – Colvars interface: first update	152
6.4.5	First numerical results	153
6.4.5.1	On the mass of the collective variable λ	153
6.4.5.2	On the boundary conditions	155
6.4.5.3	Free energy profiles	156
6.5	Implementation of the OSRW method	167
6.5.1	Equations of motion	167
6.5.2	Tinker-HP – Colvars interface: second update	169
6.5.3	Numerical results: what is done and what remains to be done	169
6.5.3.1	OSRW with "softhard" λ -dynamics: inherent instabilities	170
6.5.3.2	OSRW with "softsoft" λ -dynamics	172
A	Notes on stochastic processes	175
A.1	Markov processes	175
A.2	Markov semigroups	175
A.3	Infinitesimal generator	177
A.4	Fokker-Planck-Kolmogorov equation	177
A.5	Hypoellipticity	178
A.6	Diffusion processes	179
B	Some notions in analytical mechanics	181
B.1	Some reminders of classical mechanics	181
B.1.1	Newton's second law	181
B.1.2	Work and kinetic energy	182
B.1.3	Conservative forces	182
B.2	Foundations of analytical mechanics	183
B.2.1	Relationship between potential and kinetic energies	183
B.2.2	Lagrangian	183
B.2.3	Constraints, degrees of freedom and generalised coordinates	184
B.2.4	Conjugate momentum	186
B.2.5	Generalised forces	186
B.2.6	Hamiltonian	187
C	Derivatives for the implementation of λ-dynamics and the OSRW method	189
C.1	Halgren potential	189
C.2	Softcore potential for van der Waals interactions	191
C.3	Softcore potential for electrostatic interactions	193
	References	197

Remerciements

Nul manuscrit ne verrait le jour sans les efforts de tout un monde, auquel je dois ma reconnaissance. Adeptes du violon, il est venue l'heure des remerciements.

J'aimerais tout d'abord remercier chaleureusement mes directeurs de thèse, Pierre Monmarché, Tony Lelièvre et Jean-Philip Piquemal pour m'avoir donné l'opportunité de réaliser cette thèse sous leur direction. En particulier, je souhaiterais remercier Pierre de m'avoir fait découvrir le très joli monde de la dynamique moléculaire, et de m'avoir fait l'honneur d'être sa première doctorante. Je souhaiterais remercier Tony pour son soutien durant le premier confinement, qui m'a beaucoup apporté tant sur le plan mathématique qu'humain, ainsi que pour ses conseils avisés. Enfin, j'aimerais remercier Jean-Philip pour sa totale confiance, sa bienveillance et son optimisme communicatif. Je remercie mes deux directeurs de thèse additionnels et officieux, à savoir Louis Lagardère et Jérôme Hénin¹, avec qui j'ai eu le grand plaisir de collaborer cette dernière année. Travailler à vos côtés aura été une grande source d'épanouissement et d'apprentissage. Cette thèse n'aurait pas vu le jour sans les conseils et l'investissement de Gabriel Stoltz, *deus ex machina* que je remercie chaleureusement pour son aide.

Je tenais à remercier chaleureusement Julian Tugaut, qui a accepté de rapporter cette thèse. Ses remarques pertinentes et son attention aux détails m'ont particulièrement touchée. *In the same manner, I would like to warmly thank Benedict Leimkuhler, who also took the time to review this manuscript, and to whom I am grateful for the incredibly motivating chats we had in the past months: molecular dynamics never sounded so fun!* J'adresse par ailleurs mes remerciements à Virginie Ehrlacher ainsi qu'à Yvon Maday qui me font l'honneur de faire partie de mon jury de thèse.

Les cheminements de la thèse sont toujours plus agréables lorsque l'on est bien accompagnée. J'aimerais de ce fait remercier Laurent Boudin, Michèle Thieullen et Matthieu Salanne d'avoir effectué mon suivi de thèse avec la plus grande bienveillance. Je remercie de même Catherine Drouet, Malika Larcher et Salima Lounici, pour toute leur aide (et patience...!). Un grand merci à Khashayar Dadras et Pierre-Henri Tournier pour leurs interventions à la vitesse de la lumière qui m'ont permis d'éviter beaucoup d'arrachage de cheveux². Un grand merci à Jean-Yves Chemin, Barbara Gris, Ayman Moussa et Diane Peurichard d'avoir amorcé un changement au sein du Pôle Écoute, agir à vos côtés s'est révélé très formateur.

Ces trois dernières années ont par ailleurs été l'occasion pour moi de rendre l'ascenseur à cette université qui m'est chère. J'aimerais remercier toute l'équipe du DU RESPE de m'avoir

¹ou *Jérôme et Louis*, comme il en est désormais l'usage.

²mais pas leur blanchiment...

permis d'effectuer mes premières heures d'enseignements dans un cadre riche et humain. Il en va de même d'Irina Kourkova, à qui j'adresse ma reconnaissance pour sa totale confiance. Un grand merci à toi Chantal, pour m'avoir donné l'opportunité d'enseigner en dehors de ma zone de confort. Enfin, il va sans dire Nathalie, que je te remercie pour ta gentillesse et ton professionnalisme à toute épreuve : les étudiants sont chanceux de t'avoir. À ces derniers, j'adresse ma gratitude entière, pour la joie et la fierté qu'ils m'ont apporté.

On ne s'engage pas dans une thèse sans un certain ~~masochisme~~ amour de la Science. Je tiens à ce titre à exprimer ma profonde gratitude envers plusieurs de mes professeurs, qui ont alimenté la curiosité de beaucoup d'élèves. Je remercie Laurent Boudin d'avoir été un mentor depuis tant d'année : Laurent, je te dois beaucoup. Un merci chaleureux à Alexandre Sine, pour avoir été un professeur et un adulte admirable, qui a su donner le goût des mathématiques à une adolescente un peu perdue. Cécile R., pour ton soutien, les dîners, les expos et les super cours de biologie : merci!

Pour amorcer une entrée graduelle dans le personnel, j'aimerais tout d'abord remercier les copains du boulotTM. Tout d'abord les anciens, Anne-Françoise, Pierre M. et JF, pour m'avoir prise sous leur aile en début de thèse. Les cuvées 2018, 2019, 2020, et 2021, trop nombreux pour tous les nommer, dont en vrac, Alex, Allen, Anatole, Fatima, Gontran, Houssam, Jeet, Nico, Lucas E., Katia, Po-Yi, Sylvain, Valentin, *la squadra italiana, el equipo hispánico* et les petits frères de thèse Lucas Journal et Pierre Le B(r)is : merci d'avoir été là! J'ai une pensée pour les "quelques" séances de cinéma qui ont animé cette rédaction, accompagnée de Rui, Ioanna-Maria, Jules C. et Maria. Merci à l'équipe de choc du GTT, j'ai nommé Matthieu D. et Yipeng, avec qui j'ai eu le grand plaisir d'organiser moults sessions zooms. À l'élite du deuxième, j'ai nommé Nelly, Hugo, Matthieu M., Justine, Jules G., et tous les nouveaux, merci d'avoir supporté le chauffage à fond, le caractère de cochon, et les bavardages à foison. À Igor, merci pour toutes ces journées dans le noir, les discussions interminables, et les meilleures spots du Morbihan. Un grand merci à Jules³, petit père des thésards, pour ton amitié, ta gouache et ton investissement sans faille dans la vie du labo. À Élise⁴ et Gabriela, merci pour votre sororité et votre amour du cidre et du café, respectivement! À la petite famille du 16-26-324, Ana, Cécile, David et Idriss qui m'ont acceptée comme pièce rapportée, merci pour tous les verres, engueulades et bon moments. À mon ami Ludo, merci pour les films, la Suze, et les super vacances. J'ai toute confiance en vos succès futurs.

En dehors des copains du boulotTM, il y a quelques docteurs et électrons libres⁵ que je tenais à remercier. Je remercie les copains du masterTM, à savoir Safaa, Lois, Michèle et Yekta, à qui je souhaite une belle continuation. Je remercie les *memelords* de Neurchi de TromatématikTM qui ont su soutenir mon talent pour la procrastination. Au copains de RennesTM, j'adresse ma gratitude pour tout votre soutien, et votre joie malgré la pluie : merci Lucile et Antoine M., Noémie, Amandine, Fabrice, Tatave, Camille et Juliette, Tom-tom, Léo, Clément, Grégoire, Antoine D. et Émilie. Enfin, j'adresse un merci tout spécial à mes copains de facTM Alice, Guobao, et Jérémy, pour ces moments (et ce tapis marseillais) inoubliables.

Deux figures féminines m'ont toujours poussée à aller plus loin. Alice, Chantal, merci d'avoir été des modèles de réussite académique et humaine, et de m'avoir épaulée tout ce temps.

³le seul, l'unique, l'inimitable.

⁴et les pièces rapportées, j'ai nommé Camille ainsi que *CyriletNaïla*

⁵entendre, vendus du Capital...

Il va sans dire que je remercie du fond du coeur Shihan Jean Château et Sensei Brigitte Baudry, qui me poussent quotidiennement à devenir une personne plus juste et plus forte. Merci aux membres du Cercle Kibukan, passés, présents et futurs, pour votre sens de la camaraderie et amour de la bagarre. Sempai Chris, Denis, Grégory et David, Léa, Fred, Charlotte, Djamal, et les autres: 押忍!

Je tiens à remercier mes compagnons de vie, sans qui cette dernière serait bien fade. Sans distinction aucune, merci Marine, Damien, Océane, Léana, Jennifer, Margot, Gaugau, Maxou, Warka, Cécile A. et Aurore : votre amour et soutien, je compte bien vous le rendre au centuple, jusqu'à ce que vie s'ensuive.

Rémi, merci pour ta présence. Quelle joie que de cheminer à tes côtés!

Naturellement, j'adresse toute ma gratitude à ma famille. À Hugo, Pauline et les deux petits démons, pour les week-ends animés. À Christine et Bernard, pour leur soutien incontestable. À mon grand-père André, pour son érudition, son amour de la terre, et ses histoires rigolotes. À ma grand-mère Michèle, pour sa joie de vivre, son amour de l'autre et ses talents culinaires inégalés. À ma tante Laure, pour son soutien digne de la meilleure des taties. À mes parents enfin, pour ce si beau cadeau qu'est cette vie.

Résumé succinct en français

Contexte et motivations – Le présent mémoire s’articule autour de questions propres à la dynamique moléculaire. Le but de la dynamique moléculaire est d’étudier l’évolution de systèmes microscopiques ayant un grand nombre de particules (qui peuvent être des atomes, des molécules, ou encore même des protéines), afin d’en déduire diverses propriétés macroscopiques, comme par exemple l’énergie ou la température. Lier une description microscopique de la matière à sa description macroscopique repose sur des principes de mécanique statistique : un résultat classique nous dit qu’à l’équilibre thermodynamique, les positions des particules sont distribuées selon la mesure de Boltzmann-Gibbs $\mu \propto \exp(-\beta V)$, où V est l’énergie potentielle du système, et β est le bêta thermodynamique. À partir de cette distribution, de nombreuses propriétés macroscopiques peuvent être obtenues. Être capable d’échantillonner la mesure μ , à savoir être capable d’obtenir numériquement des positions distribuées selon μ , est un problème clef en dynamique moléculaire. Pour ce faire, on peut utiliser la dynamique de Langevin suramortie, où les positions du système sont représentées par un processus stochastique $(X_t)_{t \geq 0}$ satisfaisant l’équation différentielle stochastique suivante:

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

où $(W_t)_{t \geq 0}$ est un mouvement Brownien classique, et $\mathcal{F} = -\nabla V$ est la force d’interaction entre les particules. S’écrivant comme le gradient d’une énergie potentielle, \mathcal{F} est dite conservative. Un tel processus a de bonnes propriétés théoriques, mais est difficile à utiliser en pratique. En effet, le système obtenu est dit métastable : les particules peuvent se retrouver coincées dans des puits d’énergie potentielle et prendre beaucoup de temps à s’en échapper [56]. Le coût à fournir pour atteindre l’équilibre et échantillonner la mesure μ est alors trop important. Afin d’éviter la métastabilité, une idée est de considérer une coordonnée de réaction, à savoir une fonction ξ des positions représentant le système de manière simplifiée. Étant donnée cette coordonnée de réaction, on peut alors considérer la méthode de l’*Adaptive Biasing Force* (ABF) [39], qui consiste à biaiser la force \mathcal{F} dans la direction de ξ , à l’aide d’un biais s’adaptant à chaque pas de temps, et montrer la convergence en temps long de l’algorithme [57]. On peut aussi s’intéresser à une variante, la méthode de la *Projected Adaptive Biasing Force* (PABF) [2]. Une propriété intéressante de ces méthodes est celle de l’histogramme plat : le profil de l’énergie potentielle est lissé dans la direction de ξ , il n’y a plus de métastabilité, et l’équilibre thermodynamique est atteint bien plus rapidement.

La première partie de cette thèse est dédiée à l’étude des algorithmes ABF et PABF lorsque la force d’interaction \mathcal{F} n’est plus conservative. En effet, certains modèles d’approximation des forces d’interaction induisent une violation de la conservation de l’énergie mécanique du système Hamiltonien considéré, une conséquence directe étant que la force ne peut plus s’écrire comme le gradient d’une énergie. Il faut alors s’assurer qu’utiliser les algorithmes ABF et PABF a toujours

du sens, étant donné que la majeure partie des preuves de convergence (convergence de la loi du processus considéré vers une mesure d'équilibre, convergence du biais adaptatif vers un biais stationnaire) repose sur le fait que la force \mathcal{F} est conservative. Plus précisément, l'on doit s'assurer que (i) il existe une mesure et un biais stationnaires vers lesquels converger, et (ii) la propriété de l'histogramme plat est bien vérifiée. Le cas échéant, il reste à quantifier l'erreur commise lors des différents calculs de référence à l'aide de la nouvelle méthode. Le travail effectué au cours de cette thèse afin d'amener des réponses à ces questions est désormais disponible en tant que *preprint*, à la page arXiv:2102.09957.

Une question, encore ouverte aujourd'hui, se pose : quel est le choix optimal de la coordonnée de réaction ξ ? La seconde partie de cette thèse cherche à y répondre, et ce dans un cadre bien particulier, celui des transformations dites alchimiques. Une telle transformation consiste, à l'aide d'une coordonnée de réaction $\lambda \in [0, 1]$, à amener le système étudié d'un état initial $\lambda = 0$ à un état final $\lambda = 1$, en passant par des états intermédiaires $\lambda \in]0, 1[$ n'ayant *a priori* pas de sens physiquement. Étudier les transformations alchimiques s'avère utile dans de nombreux cas de figure, où l'on cherche à établir la différence d'énergie libre $\Delta_{0 \rightarrow 1} A = A(1) - A(0)$ du système considéré entre l'état initial et l'état final. C'est le cas par exemple de la transformation d'un ligand L_1 en un ligand L_2 , qui permet alors de déterminer lequel est le plus à même de se lier à un récepteur cible. La λ -dynamique introduite par C. L. Brooks, III [49], fait partie des nombreuses méthodes existantes pour estimer ces différences d'énergie libre : la coordonnée λ est alors considérée comme une variable dynamique additionnelle. La méthode de la λ -dynamique présente des avantages vis-à-vis des autres méthodes, plus classiques. Néanmoins, elle n'est pas sans limitation, la coordonnée λ n'étant *a priori* pas capable de capturer la métastabilité du système. Il est donc nécessaire, dans le cadre de transitions alchimiques, d'associer à la λ -dynamique une méthode d'accélération d'échantillonnage, dans la même ligne que la méthode ABF. Dans cette optique, W. Yang [87, 88, 66, 89] propose dans plusieurs de ses travaux un choix de coordonnée de réaction ainsi qu'un algorithme qui amèneraient à une exploration plus rapide et efficace de l'espace des configurations d'un système. Cependant, cette méthode reste à ce jour ni reproductible, ni mathématiquement validée. En collaboration avec Louis Lagardère (Laboratoire de Chimie Théorique, Sorbonne Université, France) et Jérôme Hénin (Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, CNRS, France), nous implémentons une méthode reproductible permettant de valider –ou invalider– ce choix de coordonnée de réaction, tout en s'assurant de sa cohérence mathématique. Ce travail est par ailleurs à l'origine d'une nouvelle interface entre deux logiciels de dynamique moléculaire, Tinker-HP [52] et Colvars, ce qui élargit conséquemment le champ des possibles en termes de simulations.

Structure – La première partie de ce manuscrit consiste en une introduction générale de la dynamique moléculaire, et est composée de quatre chapitres. Le Chapitre 1 présente brièvement les concepts essentiels de la mécanique statistique : on y rappellera la nécessité d'une description probabiliste des systèmes à l'échelle microscopique, les équations du mouvement dans un tel cas de figure, les différents ensembles thermodynamiques dans lesquels travailler, ainsi que les potentiels thermodynamiques usuels d'intérêt, dont l'*entropie* et l'*énergie libre de Helmholtz*. Le Chapitre 2 est dédié à la dynamique moléculaire en elle-même : on y présentera l'approximation de Born-Oppenheimer, sans laquelle la majeure partie des simulations ne serait possible, les potentiels interatomiques usuels ainsi que les différentes conditions au bord utilisées en dynamique moléculaire. Les dynamiques de Langevin et de Langevin suramortie seront présentées en détails, et l'on abordera rapidement les schémas numériques les plus connus utilisés pour leur discrétisation. Le Chapitre 3 quant à lui présentera plus en détail les motivations de cette thèse, à

savoir le calcul de différences d'énergie libre à l'aide d'algorithmes d'échantillonnage. Le concept de métastabilité et de coordonnée de réaction mentionnés plus haut y seront abordés de manière exhaustive. Les méthodes d'échantillonnage accéléré utilisant le concept de coordonnée de réaction afin d'éviter la métastabilité d'un système seront ensuite abordées. On portera une attention particulière aux méthodes de biais adaptatifs telles que la métadynamique, la méthode de l'*Adaptive Biasing Potential*, ainsi que la méthode de l'*Adaptive Biasing Force* (ABF) susmentionnée. On évoquera la première motivation de cette thèse, à savoir l'étude de la méthode ABF dans le cas de forces d'interaction non conservatives. On procédera alors à la présentation des méthodes classiques de calculs de différences d'énergie libre, afin de se focaliser sur l'une d'elles, la λ -dynamique, que nous introduirons succinctement, un chapitre entier y étant par la suite dédié. Enfin, nous terminerons avec la seconde motivation de cette thèse, qui revient à étudier la méthode d'échantillonnage de l'*Orthogonal Space Random Walk* (OSRW) proposée par Wei Yang, qui repose sur un choix particulier de coordonnée de réaction. Enfin, le Chapitre 4 détaillera les contributions de cette thèse ainsi que l'organisation précise du présent manuscrit.

La deuxième partie de ce manuscrit se focalise sur l'étude de la méthode de l'*Adaptive Biasing Force* dans le cas de forces non conservatives, et peut être lue de manière quasi-indépendante du reste de ce corpus. On y rappellera succinctement des concepts déjà introduits en partie I, pour ensuite présenter les résultats, et procéder à la démonstrations de ces derniers.

Enfin, la troisième et dernière partie regroupe les efforts produits à ce jour dans le cadre de l'étude et l'implémentation des méthodes de la λ -dynamique et de l'OSRW dans le logiciel *Tinker-HP* et le module *Collective Variables*. La méthode de l'OSRW sera introduite en détails, et nous en présenterons ses limitations théoriques, avant de motiver le besoin d'implémenter la λ -dynamique. On présentera au préalable le logiciel Tinker-HP ainsi que le module Collective variable, aussi dit Colvars, afin de définir la nouvelle interface mise en place permettant à ces deux programmes de communiquer dans le cadre de la λ -dynamique. Par la suite, la λ -dynamique, déjà abordée rapidement au Chapitre 3, sera présentée de manière exhaustive, et l'on détaillera une de ses applications clés en pharmacologie. L'implémentation de cette méthode, préliminaire à celle de la méthode de l'OSRW, appelle à utiliser de nouveaux potentiels d'interaction, dits *softcores*, dans le code du programme Tinker-HP, que l'on présentera en détails pour deux types d'interactions. La λ -dynamique permettant le calcul de différences d'énergie libre dans le cadre de transitions alchimiques, nous présenterons des résultats numériques que l'on comparera à des différences d'énergie libre de référence. Enfin, nous procéderons à la présentation de l'implémentation de la méthode OSRW, pour conclure sur les problèmes soulevés lors de cette dernière, qui à ce jour, restent ouverts.

Des annexes s'ajoutent à ce corpus afin d'en faciliter la lecture. L'annexe B consiste en de très brefs rappels de mécanique analytique: y est rappelé le passage de la mécanique Newtonienne à la mécanique Hamiltonienne, ainsi que le concept de coordonnées et de forces étendues, nécessaires à la bonne compréhension de l'étude de la λ -dynamique et de la méthode de l'OSRW présentée en partie III. L'annexe A quant à elle, est dédiée aux processus de diffusion, et peut être consultée en amont de la lecture de la partie II de ce manuscrit. On y rappelle, entre autres, les définitions de processus de Markov, de générateur infinitésimal, de mesure invariante et d'équation de Fokker-Planck, d'hypoellipticité ainsi que de diffusion. L'annexe C contient tous les calculs exhaustifs de gradients, nécessaires à l'implémentation des méthodes de λ -dynamics et d'OSRW, afin de faciliter la lecture de la partie III.

Constants and conventions

Fundamental physical constants

m_e	$9.10938356 \times 10^{-31}$ kg	electron mass
e	$1.60217662 \times 10^{-19}$ C	elementary charge
\mathcal{N}_A	$6.02214076 \times 10^{23}$ mol ⁻¹	Avogadro's number
k_B	$1.38064852 \times 10^{-23}$ m ² .kg.s ⁻² .K ⁻¹	Boltzmann constant
ε_0	$8.8541878128(13) \times 10^{-12}$ F.m ⁻¹	vacuum permittivity

Notations

• Physical systems

- ▷ N is the number of particles, d denotes the space dimension.
- ▷ $\mathcal{D} \subset \mathbb{R}^{dN}$ denotes the configuration space.
- ▷ $\mathcal{T}^* \times \mathcal{D}$ denotes the state space.
- ▷ q can denote either the position or the charge of a particle. In the latter case, the position is denoted by \mathbf{r} .
- ▷ p denotes the momentum of a particle.
- ▷ M is the mass matrix of a given system.

• Diffusion processes, invariant measures

- ▷ $\lambda(dq)$ denotes the Lebesgue measure on the configuration space \mathcal{D} .

- ▷ $\beta = (k_B T)^{-1}$ is the thermodynamic beta, with T being the absolute temperature of the system.
- ▷ $\mu_\beta \propto \exp(-\beta H)$ is the density of the canonical measure of a system with Hamiltonian H .
- ▷ $\mu \propto \exp(-\beta V)$ is the density of Boltzmann-Gibbs measure of a system with potential energy V .
- ▷ $(X_t)_{t \geq 0}$ with any capital letter other than W will denote a stochastic process.
- ▷ $(W_t)_{t \geq 0}$ denotes a standard Brownian motion of dimension to be specified.

• **Enhanced sampling methods**

- ▷ ξ is the standard notation for a generic transition coordinate.
- ▷ λ denotes the order parameter of an alchemical transition.
- ▷ A (resp. G) is the Helmholtz (resp. Gibbs) free energy of the system.
- ▷ F denotes the local mean force of a system.
- ▷ \mathcal{F} denotes the interaction force of a system. If $\mathcal{F} = -\nabla U$ where U is an energy, then \mathcal{F} is said to be *conservative*.

• **Set of functions** Let $\Omega \subset \mathbb{R}^{dN}$ be a Borel set.

- ▷ $\mathcal{C}_0^\infty(\Omega)$ denotes the set of infinitely differentiable functions with compact support on Ω .
- ▷ $\mathcal{C}^k(\Omega)$, with $k \in \mathbb{N}^*$ denotes the set of k -differentiable functions on Ω . $\mathcal{C}^0(\Omega)$ denotes the set of continuous functions.
- ▷ For any p in $[1, +\infty]$, $L^p(\Omega)$ denotes the Lebesgue space of order p , $L^\infty(\Omega)$ being the space of essentially bounded functions. The norm of a function f in $L^p(\Omega)$ is defined as:

$$\|f\|_{L^p(\Omega)} = \|f\|_p := \begin{cases} \left(\int_{\Omega} |f|^p \right)^{\frac{1}{p}} & \text{if } p \in [1, +\infty[, \\ \inf \{ M \geq 0 \mid |f(x)| \leq M \text{ for almost every } x \} & \text{if } p = +\infty \end{cases} .$$

- ▷ For any $p \in [1, +\infty]$ and $k \in \mathbb{N}^*$, we define the Sobolev space

$$W^{k,p}(\Omega) := \{f \in L^p(\Omega) \mid D^\alpha f \in L^p(\Omega), \forall |\alpha| \leq k\}$$

endowed with the norm:

$$\|f\|_{W^{k,p}(\Omega)} = \|f\|_{k,p} := \begin{cases} \left(\sum_{|\alpha| \leq k} \|D^\alpha f\|_p \right)^{\frac{1}{p}} & \text{if } p \in [1, +\infty[, \\ \max_{|\alpha| \leq k} \|D^\alpha f\|_\infty & \text{if } p = +\infty \end{cases} .$$

We will mainly work with the $H^1(\Omega) = W^{1,2}(\Omega)$ space.

▷ $H_0^1(\Omega)$ denotes the closure of the space $C_0^\infty(\Omega)$ with respect to the Sobolev norm $\|\cdot\|_{H^1}$.

Part I

General introduction

Nous devons donc envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome ; rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux.

PIERRE-SIMON LAPLACE
Essai philosophique sur les probabilités,
Oeuvres, Gauthier, Villars, 1886, vol. VII, 1, pp. 6-7.

Introductory foreword

Context and motivations – The present dissertation is centred around problems related to molecular dynamics. The goal of molecular dynamics is to study the evolution of microscopic systems composed of a great number of particles (be they atoms, molecules, or even proteins), in order to obtain macroscopic informations about it, such as the energy or the temperature. To relate a microscopic description of matter to its macroscopic description relies on the principles of statistical mechanics. One may use a classical result, which is that at thermodynamical equilibrium, the particles' positions are distributed according to the Boltzmann-Gibbs measure $\mu \propto \exp(-\beta V)$, where V is the potential energy of the system, and β is the thermodynamic beta. Given this distribution, one may obtain many macroscopic properties. Sampling the measure μ , in others words, being able to numerically obtain positions distributed according to μ , is a key problem in molecular dynamics. To do this, one may use overdamped Langevin dynamics, where the position vector of the system is represented by a stochastic process $(X_t)_{t \geq 0}$ which satisfies the following stochastic differential equation:

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

where $(W_t)_{t \geq 0}$ is a standard Brownian motion, and $\mathcal{F} = -\nabla V$ is the interaction force between the particles. As the force \mathcal{F} is written as the gradient of a potential energy, it is said to be conservative. Such a process has good theoretical properties, but it is difficult to use it in practice. Indeed, the obtained system is said to be metastable: the particles may remain trapped in energy wells and take a lot of time to free themselves [56]. Waiting to reach equilibrium and sampling the measure μ is thus too costly. In order to avoid metastability, one idea may be to consider a reaction coordinate, namely a mapping ξ of the positions, representing the system in a simplified manner. Given this reaction coordinate, one may then use the *Adaptive Biasing Force* (ABF) method [39], which consists of biasing the interaction force \mathcal{F} in the direction of ξ with an adaptive bias, and show the long-time convergence of the algorithm [57]. One may also consider a variation of ABF, namely the *Projected Adaptive Biasing Force* (PABF) method [2]. An interesting property of both methods is that of the flat histogram: the potential energy profile is flattened in the direction of ξ , avoiding metastability in this direction, and convergence to thermodynamic equilibrium is accelerated.

The first goal of this thesis is the study of the ABF and PABF algorithm in the case where the interaction force \mathcal{F} is no longer conservative. Indeed, some models used to approximate interaction forces may lead to a violation of the mechanical energy of the Hamiltonian system considered. A direct consequence being that the force can no longer be written as the gradient of an energy. One then needs to ensure that using the ABF and PABF methods is still relevant, as most of the long-time convergence proofs (convergence of the law of the process to an equilibrium measure, convergence of the adaptive bias to a stationary bias) rely on the fact that the force \mathcal{F} is

conservative. To be more precise, one has to ensure that (i) there exists a stationary measure and stationary bias towards which one can converge, and (ii) the flat histogram property is verified. If so, one then has to quantify the error committed while estimating the quantities of interest using the new algorithm. The work done during this thesis in order to answer such questions is now available as a preprint, at the page arXiv:2102.09957.

A question still open to this day arises: what is the optimal choice for the reaction coordinate ξ ? The second goal of this dissertation is to provide an answer to this issue, in a very specific case, that of alchemical transitions. Such a transition, given a reaction coordinate $\lambda \in [0, 1]$, consists of bringing the system at hand from an initial state $\lambda = 0$ to a final state $\lambda = 1$, while visiting temporary states $\lambda \in (0, 1)$ which are allowed to not make sense physically. Studying alchemical transitions proves to be useful in many cases, when one needs to estimate the free energy difference $\Delta_{0 \rightarrow 1} A = A(1) - A(0)$ of a system between the initial state and the final state. It is the case for example when a ligand L_1 is transformed in a ligand L_2 , so that one may determine which ligand between L_1 and L_2 is more prone to bind with a target receptor. Introduced by C. L. Brooks, III [49] and colleagues, λ -dynamics is among several already existing methods designed for the computation of free energy differences. The key idea behind λ -dynamics is to consider the coordinate λ as an additional dynamical variable. The method of λ -dynamics offers some advantages compared to the more classical methods. However, λ -dynamics is not without limitations, as the coordinate λ is *a priori* not designed to capture the system's metastability. This motivates the need to associate an accelerated sampling method to λ -dynamics, such as the ABF method. In this scope, W. Yang [87, 88, 66, 89] suggests using a new choice of reaction coordinate, coupled to a sampling method that is to this day not easily reproducible, nor mathematically validated. In collaboration with Louis Lagardère (Laboratoire de Chimie Théorique, Sorbonne Université, France) and Jérôme Hénin (Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, CNRS, France), we set on implementing a reproducible method which would validate –or invalidate– this choice of reaction coordinate, while ensuring its mathematical coherence. This work has moreover led to the creation of a new interface between two molecular dynamics softwares, Tinker-HP [52] and the Collective Variables module, which widen the field of possibilities in term of simulations.

Structure – The first part of this dissertation is a general introduction to the thesis' objectives, and is composed of four chapters. Chapter 1 briefly introduces the essential concepts of statistical mechanics: we will motivate the need to use a probabilistic description of physical systems at the microscopic scale, recall the equations of motion used in this scope, introduce the different thermodynamical ensemble one may work with, and eventually define the usual thermodynamical potentials, among which the *entropy* and the *Helmholtz free energy*. Chapter 2 is dedicated to molecular dynamics in itself: we will present the Born-Oppenheimer approximation without which most simulations would not be possible, the usual interatomic potentials along with the different kinds of boundary conditions used in molecular dynamics simulations. Langevin and overdamped Langevin dynamics will be introduced in details, and we will quickly review some numerical schemes used to discretise said dynamics. In Chapter 3 we will thoroughly introduce the main motivation for this thesis, namely the computation of free energy differences with sampling algorithms. The aforementioned concepts of metastability and of reaction coordinate will be carefully defined. The accelerated sampling algorithms relying on a reaction coordinate to by-pass metastability will then be introduced. We will pay special attention to adaptive biasing methods such as Metadynamics, the Adaptive Biasing Potential or the Adaptive Biasing Force mentioned above. We will present the first motivation of this thesis, namely the study of the

ABF method in the case where the interaction force is no longer conservative. We will then proceed to introduce the methods used to compute free energy differences, before focusing on one of them, called λ -dynamics. Said method will be briefly presented, as a whole chapter will be dedicated to its study later on. Eventually, we will sketch the second motivation of this thesis, which is to study the so-called *Orthogonal Space Random Walk* (OSRW) sampling method, introduced by Wei Yang and colleagues, whose design relies on the choice of a particular reaction coordinate. To finish, Chapter 4 will detail carefully the main contributions of this dissertation, along with its structure.

The second part of this corpus focuses on the study of the Adaptive Biasing Force algorithm in the non-conservative case. It consists of the single Chapter 5, which is written to be read independently. We will briefly recall some concepts already presented in part I, and then proceed to introduce the results and the associated proofs.

Eventually, the third and last part gathers all of the work related to the study and implementation of λ -dynamics and the OSRW method in the Tinker-HP and Collective Variables module softwares. It consists of the single Chapter 6. The OSRW method will be introduced in details and we will list its theoretical limitations, before motivating the need for an implementation of λ -dynamics. We will beforehand present the Tinker-HP software along with the Collective Variables module –also called Colvars–, in order to define the new interface put in place between both codes. We will later on exhaustively introduce λ -dynamics, which would have already been quickly presented in Chapter 3, and detail one of its key applications in drug design. The implementation of this method, which is prior to that of OSRW, requires the implementation of new interaction potentials, called *softcores*, into the Tinker-HP software. We will detail two kinds of softcores potentials, one for the van der Waals interactions, the other for electrostatic interactions. As λ -dynamics allows for the computation of free energy differences in the case of alchemical transitions, we will present some numerical results that will be compared to reference free energy differences. Finally, we will discuss the implementation of the OSRW method, to conclude on the issues that arose doing so, and which are to this day open problems.

Several annexes are added to this corpus in order to facilitate its reading. Annex B consists of small reminders on analytical mechanics: we will briefly detail the transition from Newtonian mechanics to Hamiltonian mechanics, and introduce the notion of extended coordinates and forces which prove to be essential for the study of λ -dynamics and the OSRW method presented in part II. Annex A is dedicated to a brief introduction to diffusion processes and can be read before part II. Among other things, will be recalled the concepts of a Markov process, the infinitesimal generator, invariant measures, Fokker-Planck equation, hypoellipticity and diffusion. Annex C contains all of the exhaustive computations of derivatives needed for the implementation of both λ -dynamics and the OSRW method, in order to ease the reading of part III.

Chapter 1

A quick introduction to statistical physics

Giving sense to the invisible drives us from the moment we are born. What is this world that surrounds us? Where does heat come from? Where do illnesses come from? How come certain phenomena governing Nature cannot be grasped by our own physical perception? Struggling to answer such questions gave birth to many conceptions of the world: some remained the norm for centuries, others stirred violent controversy. To name but a few, one can cite the pre-socratic notion of indivisible atoms developed by Leucippus and Democritus around 440 to 430 BC, which considered matter as being composed of indivisible entities called atoms. Neglected at first, our world now revolves around it. Another example, now an obsolete theory, is that of the triad of alchemy, or *Tria prima*, defined by Paracelsus in 1533 in his *Opus paragranum*. It attempted to explain natural phenomena, especially physiological ones, via the links between Sulfur, Salt and Mercury, and is now considered to be the starting point of modern occidental chemistry. One could also mention Black and Lavoisier's mid-eighteen century caloric theory, where heat was wrongly considered to be a fluid, and which eventually led to the foundation of classical thermodynamics. All of these theories interacted with the others, by either building bridges between ideas or raising walls against arguments. Eventually, the efforts of many led us into the 21st century viewing our world through the lens of yet another paradigm: that of modern statistical physics. In this chapter, we will present its key concepts.

1.1 A matter of scales

When it comes to describing physical systems, different scales come into play for both space and time measurements [43]. The macroscopic scale is the scale we are most at ease with: lengths are measured in *meters*, time in *seconds*, and our human eye can grasp the number of individuals in a given domain up to hundreds of people. In the microscopic scale, typical lengths are of the order of the Angström ($1\text{\AA} = 10^{-10}\text{m}$), and times are of the order of the femtosecond (10^{-15}s). This difference of scales applies for every physical quantities, such as energies, which are expressed in Joules at the macroscopic scale and in electron-volts at the microscopic scale, with $1\text{eV} = 1.6 \times 10^{-19}\text{J}$ [15]. A good way to visualise the gap between both scales is to look

at the number of units (electrons, atoms, molecules or ions, depending on the system at hand) in a given portion –more precisely, a mole– of matter. It is given by the Avogadro’s number $\mathcal{N}_A \simeq 6.02214076 \times 10^{23} \text{ mol}^{-1}$. As an example [15], in a 1mm^3 solid, if one considers the interatomic distances to be of the order of the Angström, there would be around $(10^{-3} \text{ m}/10^{-10} \text{ m})^3 = 10^{21}$ atoms. In comparison, on July, 2021, the estimated Earth’s population is around 7.795×10^9 human beings. One can also consider timescales: for an ideal gas at atmospheric pressure and room temperature, the mean time between molecular collisions is of the order 10^{-9} s [4], whereas heating up a pan of water from room temperature to boiling temperature can take several minutes.

As a consequence, being able to connect different scale regimes can prove to be difficult. Indeed, each scale is described using observable variables, and there is *a priori* no reason for said variables to be observable at other scales [43]. One can consider the example of the molecular scale, where the motion of the electrons are often considered to be negligible. At the macroscopic scale, as we shall see further on, the temperature of a given object cannot be expressed analytically as a function of the atoms’ positions. Even though several other regimes of scale exist, such as the subnuclear scale (where quarks, the components of protons and neutrons, become visible) or the mesoscopic scale (where cellular structures such as membranes of the order of the nanometer are visible), we will only concentrate on the microscopic and macroscopic scales.

We will consider physical systems composed of N particles, be they atoms, molecules, or ions. The size of the system will refer to the number N of particles. Examples of macroscopic information one can get on the system are the temperature T , the volume V , the pressure P , the work W or even the internal energy U . These quantities are defined using the tools of *thermodynamics*, a phenomenological theory used to describe the macroscopic evolution of large physical systems. It relies on the concept of equilibrium: a physical system will tend to remain in a stationary state, called *thermal equilibrium*. A perturbed system will consequently try to reach thermal equilibrium after a certain *relaxation time*. Thermodynamic variables, such as the ones listed above, can be divided into two categories:

- ▷ an **intensive** variable X is independent of the size of the system. It means that its value, in an homogeneous system, does not depend on where the measure has been made. The temperature is an example of an intensive variable (the boiling temperature of water does not change with the amount of water considered), along with the pressure, or even the colour of the system.
- ▷ an **extensive** variable X depends on the size of the system, and is proportional to it. For two given systems \mathcal{S}_1 and \mathcal{S}_2 , one has:

$$X(\mathcal{S}_1 \cup \mathcal{S}_2) = X(\mathcal{S}_1) + X(\mathcal{S}_2).$$

Direct examples of extensive variables are the mass, the volume or the amount of substance of a system.

The intensive property indicates that the atoms of the system are not impacted by the surface [43], whereas the extensive property is linked to the fact that atoms can interact with their surroundings only within a certain range. The fixed value of one of these macroscopic quantities, say the temperature, defines a *macrostate*. A macrostate is realised by several possible configurations of the system at the microscopic scale, called *microstates*. Characterising a microstate

depends on the nature of the system considered. In the scope of classical mechanics, a microstate is composed of the dynamical variables needed to describe the particles' motion: the positions and momenta. As for quantum systems, they will be overlooked in this text. Establishing the laws between the microscopic components of a given system in order to obtain macroscopic informations about it, such as its energy, its pressure or its temperature, is the goal of *statistical mechanics*.

1.2 The equations of motion at the microscopic scale

Let us consider a system of N particles with position vector $q = (\mathbf{q}_i)_{i \in \llbracket 1, N \rrbracket} \in \mathcal{D} \subset \mathbb{R}^{3N}$ and momentum vector $p = (\mathbf{p}_i)_{i \in \llbracket 1, N \rrbracket} \in \mathbb{R}^{3N}$. The space \mathcal{D} is called the *configuration space*. We define the *phase space* $\mathcal{T}^*\mathcal{D} = \mathcal{D} \times \mathbb{R}^{3N}$ which is the space where the couple (q, p) evolves. From the standpoint of classical mechanics, one can obtain the evolution of the i -th particle by applying Newton's second law:

$$\frac{d\mathbf{p}_i}{dt} = \mathcal{F}_i, \quad \text{and} \quad \frac{d\mathbf{q}_i}{dt} = m_i^{-1} \mathbf{p}_i,$$

where \mathcal{F}_i is the force acting on the i -th particle and m_i its mass. This equation yields the trajectory in time of the particle, and is entirely deterministic [21]. Another deterministic manner to describe the motion of a particle is to use the Hamiltonian equations of motion. In the Hamiltonian formalism, the energy of the system is given by the Hamiltonian function

$$H(q, p) = E^{kin}(p) + V(q), \quad (1.1)$$

where $E^{kin}(p) = \frac{1}{2} p^\top M^{-1} p$ is the kinetic energy of the particles, $M = \text{diag}(m_1 Id, \dots, m_N Id)$ the mass matrix of the system, and $V(q)$ is the potential energy. The potential energy V is assumed to be sufficiently smooth, namely, $V \in \mathcal{C}^2(\mathcal{D})$.

Given this energy function, one can obtain the equations of motion as follows:

$$\frac{\partial H}{\partial \mathbf{q}_i} = -\frac{d\mathbf{p}_i}{dt} \quad \text{and} \quad \frac{\partial H}{\partial \mathbf{p}_i} = \frac{d\mathbf{q}_i}{dt}, \quad \forall i \in \llbracket 1, N \rrbracket, \quad (1.2)$$

which yields the trajectory $(q_t, p_t)_{t \geq t_0}$ of the particles in the phase space. Starting from an initial point (q_0, p_0) , one can define the *Hamiltonian flow map* ϕ_t , namely the mapping $\phi_t : \mathcal{T}^*\mathcal{D} \rightarrow \mathcal{T}^*\mathcal{D}$ such that for each time $t \geq 0$, $\phi_t(q_0, p_0) = (q_t, p_t)$. We refer to the Annex B for further details on how to establish the Hamiltonian equations of motion (1.2) from the classical, Newtonian equations.

Remark 1. *Note that both kinetic and potential energies could depend on both the momenta and positions. We will exclude this case and focus only on Hamiltonians of the previous form (1.1), which are called separable.*

Solving (1.2) requires solving $6N$ equations, $3N$ on the positions q and $3N$ on the momenta p . Since N is usually of the order of \mathcal{N}_A , it is impossible to properly follow the evolution of each particle at each time t ! Fortunately, as discussed above, macroscopic quantities of interest are realised by several possible configurations (q, p) of the system. This justifies the use of *statistics* to describe the common, collective behaviour of our N particles.

1.3 The ergodic hypothesis

We are interested in physical *observables*, namely, smooth and bounded functions of the microstates (or more generally, degrees of freedom of the system) $\varphi \in \mathcal{C}_0^\infty(\mathcal{D} \times \mathbb{R}^{3N})$. As the observation time $T > 0$ at our macroscopic scale is larger than the usual microscopic timespans, a measurement of the observable φ corresponds to its trajectorial average $\bar{\varphi}_T$:

$$\bar{\varphi}_T := \frac{1}{T} \int_0^T \varphi(q_t, p_t) dt. \quad (1.3)$$

For a given observable $\varphi \in \mathcal{C}_0^\infty(\mathcal{D} \times \mathbb{R}^{3N})$, one can infer that for large observation times T , i.e when $T \rightarrow +\infty$, the average $\bar{\varphi}_T$ converges towards a limit $\bar{\varphi}$, which can be interpreted as the equilibrium value of the observable. In order to establish this limit $\bar{\varphi}$, we introduce the probability to find the particles at a given position of the phase space, *at equilibrium*. This requires to define a measure μ on the phase space, whose density with respect to the Lebesgue measure λ will also be denoted by μ , so that the probability can be written as

$$\mu(q, p) dqdp = \mu \left((\mathbf{q}_i)_{i \in [1, N]}, (\mathbf{p}_i)_{i \in [1, N]} \right) \prod_{i=1}^N dq_i dp_i.$$

One can then define the statistical average, or *thermodynamic average* of the observable φ with respect to the measure μ :

$$\mathbb{E}_\mu[\varphi] = \int_{\mathcal{T}^*\mathcal{D}} \varphi(q, p) \mu(q, p) dqdp. \quad (1.4)$$

We will now assume the *ergodic hypothesis*:

Assumption 1. *At thermodynamic equilibrium, the trajectorial average of a given observable is equal to its statistical average. In other words:*

$$\lim_{T \rightarrow +\infty} \bar{\varphi}_T = \bar{\varphi} = \mathbb{E}_\mu[\varphi]. \quad (1.5)$$

Remark 2. *The ergodic hypothesis is a fundamental assumption in statistical mechanics. It is based on the fact that the Hamiltonian flow ϕ_t is ergodic with respect to measure $\mu(dqdp) := \mu(q, p) dqdp$ on the Euclidean space $\mathcal{T}^*\mathcal{D}$. Let us consider a measurable space $(\mathcal{X}, \mathcal{B})$, a measure ν on \mathcal{X} and a family of measurable functions $(\phi_t)_{t \geq 0}$ from \mathcal{X} to itself such that $\phi_0 = Id$ and for all $t, s \geq 0$, $\phi_t \circ \phi_s = \phi_{t+s}$. The family $(\phi_t)_{t \geq 0}$ is said to be ergodic with respect to ν if, and only if,*

$$\forall B \in \mathcal{B}, \forall t \geq 0, \quad \phi_t^{-1}(B) = B, \quad \text{then either } \nu(B) = 0 \quad \text{or} \quad \nu(B) = 1.$$

For a nice introduction to the history of the ergodic hypothesis we refer to Calvin C. Moore's perspective paper [68], and we refer to [22] for a more mathematical standpoint.

Remark 3. *Along with the statistical average $\mathbb{E}_\mu[\varphi]$ of the observable $\varphi \in \mathcal{C}_0^\infty(\mathcal{T}^*\mathcal{D})$, one may also be interested in its fluctuations $\mathbb{E}_\mu[\varphi^2] - (\mathbb{E}_\mu[\varphi])^2$.*

1.4 Different thermodynamic ensembles

Now that the ergodic hypothesis is assumed, one needs to properly define the measure μ on the phase space. Putting constraints on the system, such as fixing the number of particles N , the

volume V or the pressure P , amounts to introducing a certain probability space, or *statistical ensemble*, on which to define the measure μ . For this reason, the chosen fixed constraints along with the associated probability distribution μ , define what we call a *thermodynamic ensemble*. This yields the existence of several, different probability measures, which depend on the system's constraints. We give an exhaustive list of the different possible thermodynamic ensembles, and refer to [56, Section 1.2.3] for their proper mathematical derivation.

The microcanonical ensemble (NVE)

Here we consider that our system is completely *isolated*: it cannot exchange particles with the environment, and its energy is constant. As a consequence, the number of particles N , the volume V and the energy E are fixed, so that one can define a thermodynamic ensemble called the *microcanonical ensemble*, often denoted by (NVE). The corresponding probability measure $\mu \equiv \mu_E$ is the normalized, uniform probability measure on the set of phase points of constant energy E :

$$\Sigma_E := \{(q, p) \in \mathcal{T}^*\mathcal{D} \mid H(q, p) = E\}.$$

In other words:

$$\mu_E(dqdp) = Z_{\mu_E}^{-1} \delta_{H(q,p)-E}(dqdp), \quad Z_{\mu_E} = \int_{\Sigma_E} H(q, p) dqdp = E \int_{\Sigma_E} dqdp, \quad (1.6)$$

where the measure $\delta_{H(q,p)-E}(dqdp)$ is obtained as follows. Let us consider the set of configurations whose energy is higher than E up to the small fluctuation ΔE :

$$I_{\Delta E} := \{(q, p) \in \mathcal{T}^*\mathcal{D} \mid E \leq H(q, p) \leq E + \Delta E\},$$

endowed with a uniform measure $\tilde{\mu}_{\Delta E}$ such that, for all test function φ ,

$$\tilde{\mu}_{\Delta E}(\varphi) = \int_{I_{\Delta E}} \varphi(q, p) \mu_{\Delta E}(dqdp) = \frac{1}{\Delta E} \int_{I_{\Delta E}} \varphi(q, p) dqdp.$$

Then, as $\Delta E \rightarrow 0$, one recovers a measure $\delta_{H(q,p)-E}(dqdp)$ supported on Σ_E , such that:

$$\int_{\Sigma_E} \varphi(q, p) \delta_{H(q,p)-E}(dqdp) = \lim_{\Delta E \searrow 0} \frac{1}{\Delta E} \int_{I_{\Delta E}} \varphi(q, p) dqdp.$$

1.4.1 The canonical ensemble (NVT)

One may also consider a system which is in contact with a *thermal bath*, or *thermostat*. In this case, the system still cannot exchange particles with the environment so that N remains fixed. On the other hand, the energy now fluctuates, and it is the *absolute temperature* T that is constant. The associated thermodynamic ensemble is the *canonical ensemble* (NVT). One defines the *canonical measure* μ_β as follows:

$$\mu_\beta(dqdp) = Z_{\mu_\beta}^{-1} e^{-\beta H(q,p)} dqdp, \quad Z_{\mu_\beta} = \int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p)} dqdp, \quad (1.7)$$

where $\beta = \frac{1}{k_B T}$ is the *thermodynamic beta* and k_B is the *Boltzmann constant*. The normalisation constant Z_{μ_β} is called the *partition function*, and plays a key role in defining thermodynamic

quantities such as the internal energy, the entropy, the free energy or even the heat capacity of the system.

Now, recalling the expression (1.1) of the separable Hamiltonian, one can rewrite the canonical measure so that the position and momentum vectors are now independent random variables:

$$\mu_\beta(dqdp) = \mu(dq) \times \nu(dp), \quad (1.8)$$

where

$$\nu(dp) := \left(\frac{\beta}{2\pi}\right)^{3N/2} \prod_{i=1}^N \frac{1}{m_i^{3/2}} e^{-\frac{\beta}{2} p^\top M^{-1} p} dp,$$

and

$$\mu(dq) := Z_\mu^{-1} e^{-\beta V(q)} dq, \quad Z_\mu = \int_{\mathcal{D}} e^{-\beta V(q)} dq, \quad (1.9)$$

is the *Boltzmann-Gibbs measure*. We will later on be interested in sampling the measure μ_β , i.e, in numerically obtaining positions and momenta (q, p) distributed according to μ_β . A pleasant asset of the decomposition (1.8) is that positions and momenta can be sampled independently, the momenta being easily sampled. One then only needs to focus on sampling the Boltzmann-Gibbs measure.

1.4.2 The grand canonical ensemble (μVT)

Systems can also be *open*, and exchange particles and energy with a reservoir. One way to treat this case is to consider an isolated system composed of the target system \mathcal{S}_1 , assumed to be small compared to the much bigger reservoir \mathcal{S}_2 . If we denote by N_1 (resp. N_2), T_1 (resp. T_2), E_1 (resp. E_2), and μ_1 (resp. μ_2) the number of particles, absolute temperature, energy and *chemical potential* of the system \mathcal{S}_1 (resp. \mathcal{S}_2) then the constraints can be read as:

- ▷ Thermal equilibrium is reached, so that temperatures and chemical potentials are fixed. The temperature T and chemical potential μ of the whole system \mathcal{S} satisfy $T = T_1 = T_2$ and $\mu = \mu_1 = \mu_2$.
- ▷ The total energy E and number of particles N of the whole system are conserved, i.e $N = N_1 + N_2$ and $E = E_1 + E_2$.

This thermodynamic ensemble is called the *grand canonical ensemble*, denoted by (μVT). The associated measure, along with the Hamiltonian H_1 of the small system \mathcal{S}_1 , change with the number of particles N_1 , so that one can write the so-called *grand canonical distribution*:

$$\mu_{GC}(dp, dq; N_1) := Z_{GC}^{-1} e^{-\beta(H_1(q,p;N_1) - \mu N_1)} dq dp,$$

where

$$Z_{GC} = \sum_{N_1=0}^N e^{-\beta(H_1(q,p;N_1) - \mu N_1)}$$

is the *grand canonical partition function*.

1.4.3 Other thermodynamic ensembles

Another thermodynamic ensemble is the *isobaric-isothermal ensemble*, denoted by (NPT), as it applies to systems under constant pressure P , temperature T and number of particles N . In which case, the associated measure is given as follows:

$$\mu_{NPT}(dqdp) := Z_{NPT}^{-1} e^{-\beta(H(q,p)+PV(q,p))} dqdp,$$

where H is the Hamiltonian of the system and V its volume.

One can also define the *isoenthalpic-isobaric ensemble*, denoted by (NPH), where H is the *enthalpy* of the system, to be defined in Table 1.1. In this case, the enthalpy –which can be seen as analogous to the energy– and the pressure, are constant. The (NPH) ensemble has not been covered extensively by the literature, and we refer to the two papers [79] and [78] of J. R. Ray and his co-authors on the subject.

1.5 State functions and thermodynamic potentials

1.5.1 Definition and the example of entropy

An important notion in thermodynamics is that of *state functions*, which relate several state quantities at equilibrium, independently of the path chosen by the system to reach its current state. For example, work and heat are not state functions, as they depend on the thermodynamic path. On the other hand, the pressure P , the volume V , the temperature T , the number of particles N or the chemical potential μ are classical state functions. Their expressions and the relationships between each others can be found in [21], and we will here focus on a less intuitive state function, which is the entropy of the system.

The entropy is an extensive function of state which quantifies the degree of disorder of a given system. If the system undergoes a reversible transformation, the variation of entropy dS can be expressed as the ratio between the produced heat δQ and the absolute temperature T :

$$dS = \frac{\delta Q}{T},$$

in other words, the entropy tends to capture the energy loss of the system. If one is working within the microcanonical ensemble and if Ω denotes the number of accessible microscopic configurations for a given macroscopic state, then the entropy can be rewritten according to the Boltzmann law:

$$S = k_B \ln \Omega.$$

If one works in the **canonical ensemble** (NVT), as we will further on, one can define as in [56] the statistical entropy, for a given measure μ of density π on the phase space:

$$\mathcal{G}(\mu) := - \int_{\mathcal{T}^* \mathcal{D}} \ln(\pi(q,p)) \pi(q,p) dqdp. \quad (1.10)$$

The statistical entropy, just like the entropy S , quantifies the degree of disorder of the system. It allows us to visualise how a thermodynamic ensemble (more precisely, its associated measure) should be the measure which yields the most disordered macrostate possible matching with given data [56]. We will see in Section 1.5.3 that the statistical entropy can be seen, in a way,

as the entropy of the system in the formalism of statistical mechanics. One can show that the canonical measure μ_β is the unique minimizer of \mathcal{G} under the constraint that the average energy

$$\bar{H} = \int_{\mathcal{T}^*\mathcal{D}} H(q, p) dq dp \text{ is constant.}$$

1.5.2 Thermodynamic potentials

Among state functions, one can identify *thermodynamic potentials*, which allow to predict the evolution of the system and deduce several of its physical properties. These thermodynamic potentials depend on the ensemble in which one is working: to each thermodynamic ensemble one can associate a thermodynamic potential. Let us list them here and refer to [21, 15] for their proper derivation.

Potential	Notation	Differential	Ensemble
Internal energy	$U(S, V, N)$	$dU = TdS - PdV + \mu dN$	Microcanonical
Helmholtz free energy	$A(N, V, T)$ or $F(N, V, T)$, $A = U - TS$	$dA = -SdT - PdV + \mu dN$	Canonical
Grand potential	$\phi_G(\mu, V, T)$, $\phi_G = U - TS - \mu N$	$d\phi_G = -PdV - SdT - Nd\mu$	Grand canonical
Enthalpy	$H(P, S, N)$, $H = U + PV$	$dH = TdS + VdP$	\emptyset
Gibbs free energy or free enthalpy	$G(N, P, T)$, $G = A + PV = H - TS$	$dG = -SdT + VdP + \mu dN$	Isothermal-isobaric

Table 1.1 – Different thermodynamic potentials.

1.5.3 Free energy

The only thermodynamic potential we will focus on here is the free energy of a given system in the canonical (NVT) ensemble. As written above, the free energy is expressed in terms of thermodynamic variables as $A = U - TS$, but one may be interested into obtaining an alternative definition of the free energy within the formalism of statistical mechanics. To do so, we interpret the internal energy as the average energy with respect to the canonical measure:

$$\mathbb{E}_{\mu_\beta}[H] := \int_{\mathcal{T}^*\mathcal{D}} H(q, p) \mu_\beta(dq dp) = Z_{\mu_\beta}^{-1} \int_{\mathcal{T}^*\mathcal{D}} H(q, p) e^{-\beta H(q, p)} dq dp.$$

Recall that the statistical entropy given by (1.10) reads:

$$\begin{aligned}\mathcal{G}(\mu_\beta) &= - \int_{\mathcal{T}^*\mathcal{D}} \ln \left(\frac{e^{-\beta H(q,p)}}{Z_{\mu_\beta}} \right) \frac{e^{-\beta H(q,p)}}{Z_{\mu_\beta}} dqdp \\ &= -Z_{\mu_\beta}^{-1} \int_{\mathcal{T}^*\mathcal{D}} (-\beta H(q,p) - \ln(Z_{\mu_\beta})) e^{-\beta H(q,p)} dqdp \\ &= \beta \mathbb{E}_{\mu_\beta}[H] - \ln(Z_{\mu_\beta}).\end{aligned}$$

This yields, if one attempts to mimic the thermodynamic expression:

$$\mathbb{E}_{\mu_\beta}[H] - T\mathcal{G}(\mu_\beta) = \mathbb{E}_{\mu_\beta}[H] \left(1 - \frac{1}{k_B} \right) + T \ln(Z_{\mu_\beta}).$$

This does not satisfy us, as we wish to obtain a quantity which does not depend on the internal energy in the right-hand term. Now, if one defines the microcanonical entropy to be $k_B \mathcal{G}(\mu_\beta)$, then:

$$\mathbb{E}_{\mu_\beta}[H] - k_B T \mathcal{G}(\mu_\beta) = -\beta^{-1} \ln(Z_{\mu_\beta}),$$

and one defines the Helmholtz free energy as the logarithm of the partition function:

$$A = -\beta^{-1} \ln Z_{\mu_B}.$$

The (Helmholtz) free energy is a key quantity for chemists and physicists, and computing free energy differences will be one of the main motivations of this thesis. We will state the several uses of the free energy in Section 3.4.1 and the different manners of computing it in Sections 3.4.2 and 3.4.3.

A quick introduction to molecular dynamics

Now that we are endowed with the tools of statistical physics, we may look at ways of *numerically* predict the trajectories of a given microscopic system. Such is the motive of *molecular dynamics* (MD), a term that encompasses the numerical methods used to simulate the time evolution of interacting particles, under the assumption that atoms are classical particles.

Molecular dynamics traces back to the birth of modern computers in the early 1950s, where the first *in silico* experiments were designed. Seminal works include Metropolis, Rosenbluth, and Teller's work on the MANIAC I computer of the *Los Alamos* National laboratory [64]. The authors succeeded into obtaining proper macroscopic informations such as equations of state, by considering the molecules to be individual particles and using a modified Monte-Carlo scheme to compute trajectories. We refer to the perspective paper of E. J. Maginn and J. R. Elliott [61] for a complete review of the early days of Molecular Dynamics. The pace at which computer efficiency has evolved during the last decades has led to a broad range of applications for MD simulations. Starting in the 1970s, protein folding could now be simulated accurately, only twenty years after the first runs on the MANIAC I machine. Molecular dynamics is consequently a fast-paced field of study, and we will present its key notions in the current chapter. In Section 2.1 we introduce the Born-Oppenheimer approximation used in typical MD simulations, describe the different kinds of interaction potentials available along with the possible boundary conditions one can use. We then proceed to present the Langevin and overdamped Langevin dynamics in Section 2.2 used to sample the canonical measure (1.7), along with their different discretisation schemes.

2.1 Settings

2.1.1 The Born-Oppenheimer approximation

Most molecular dynamics simulations rely on the Born-Oppenheimer approximation, which proves to be useful for the quantum analysis of atoms and molecules. It assumes that one can treat the motion of nuclei and electrons separately. Let us quickly sketch the main idea behind the Born-Oppenheimer approximation, by considering the *molecular Hamiltonian* representing the

energy of the nuclei and electrons of a given molecule. Let us denote by $Q = (\mathbf{Q}_1, \dots, \mathbf{Q}_N) \in \mathbb{R}^{3N}$ (resp. $q = (\mathbf{q}_1, \dots, \mathbf{q}_n) \in \mathbb{R}^{3n}$) and $P = (\mathbf{P}_1, \dots, \mathbf{P}_N) \in \mathbb{R}^{3N}$ (resp. $p = (\mathbf{p}_1, \dots, \mathbf{p}_n) \in \mathbb{R}^{3n}$) the positions and momenta of the N nuclei (resp. n electrons) of the molecule. The molecular Hamiltonian is given by:

$$H(Q, P; q, p) = E_N^{kin}(P) + E_n^{kin}(p) + V_{NN}(Q) + V_{ee}(q) + V_{eN}(Q, q),$$

where

- $E_N^{kin}(P) = \sum_{i=1}^N \frac{\mathbf{P}_i^2}{2M_i}$ is the kinetic energy of the nuclei,
- $E_n^{kin}(p) = \sum_{i=1}^n \frac{\mathbf{p}_i^2}{2m_e}$ is the kinetic energy of the electrons,
- $V_{NN}(Q) = \sum_{\substack{i,j \\ i < j}} \frac{Z_i Z_j e^2}{4\pi\epsilon_0 |\mathbf{Q}_i - \mathbf{Q}_j|}$ is the potential energy resulting of Coulombic nucleus-nucleus repulsion,
- $V_{ee}(q) = \sum_{\substack{i,j \\ i < j}} \frac{e^2}{4\pi\epsilon_0 |\mathbf{q}_i - \mathbf{q}_j|}$ is the potential energy resulting of Coulombic electron-electron repulsion,
- $V_{eN}(Q, q) = - \sum_{i=1}^N \sum_{j=1}^n \frac{Z_i e^2}{4\pi\epsilon_0 |\mathbf{Q}_i - \mathbf{q}_j|}$ is the potential energy resulting Coulombic electron-nuclei attraction,

with M_i being the mass of the i -th nuclei, m_e the electron rest mass (namely, the mass of a stationary electron), Z_i the atomic mass of the i -th nucleus, ϵ_0 the vacuum permittivity and e the elementary charge.

Now, one wishes to solve the time-independent Schrödinger equation:

$$H(Q, P; q, p)\psi(Q, q) = E\psi(Q, q),$$

where $\psi(Q, q)$ is an eigenfunction of the molecular Hamiltonian associated to the (energy) eigenvalue E .

This is where the Born-Oppenheimer approximation comes into play: as the electrons are much lighter than the nuclei, the electrons can be considered to follow the nuclei motion adiabatically, meaning that they will be dragged by the nuclei without requiring a finite relaxation time. As a consequence, one can decompose the wavefunction $\psi(Q, q)$ using a separation of variables, so that:

$$\psi(Q, q) = \psi_e(Q, q)\psi_N(Q),$$

where $\psi_e(q, Q)$ is the electron wavefunction for a fixed nuclear configuration Q , and $\psi_N(Q)$ is the nuclear wavefunction.

The idea is then to consider the *clamped Hamiltonian* describing the energy of the electrons for a fixed nuclear configuration Q :

$$\begin{aligned} H_{elec}(Q; q, p) &= E_n^{kin}(p) + V_{NN}(Q) + V_{ee}(q) + V_{eN}(q, Q) \\ &\simeq E_n^{kin}(p) + V_{ee}(q) + V_{eN}(q, Q), \end{aligned}$$

where the nucleus-nucleus potential energy V_{NN} can be neglected, as here Q is only a parameter, making $V_{NN}(Q)$ a term that simply shifts the eigenvalues by a constant. One can then determine the electronic wavefunction by solving:

$$H_{elec}(Q; q, p)\psi_e(Q, q) = E_e(Q)\psi_e(Q, q).$$

Eventually, the nuclear wavefunction can be recovered by solving:

$$(E_N^{kin}(P) + E_e(Q) + V_{NN}(Q))\psi_N(Q) = E\psi_N(Q).$$

All simulations in this thesis will be done within the Born-Oppenheimer approximation.

2.1.2 Interaction potentials

Now that the Born-Oppenheimer approximation is made, we will from now on only consider the nuclei's motion. We are interested in interactions between neutrally charged particles such as atoms, molecules and even proteins. Particles are assumed to be rigid spheres interacting with each others via interaction potentials, or force fields. Depending on the choice of the force field, one can simulate from 10^4 to 10^{12} atoms, over period of times of the order of the *picosecond*, sometimes even of the *nanosecond*. When the potentials are empirical, i.e based on a man-made model, one can run simulations for up to 10^{12} atoms, to the cost of accuracy. Indeed, one would like to model the interactions as precisely as possible without relying on empirical models: this is the goal of *ab initio* molecular dynamics. With *ab initio* force fields, which take into consideration the quantum description of the interactions, the number of atoms accessible to simulation decreases greatly. Nevertheless, a recent breakthrough [44] reached the number of 10^8 atoms using machine learning. In this thesis, we will overlook *ab initio* potentials and focus only on empirical force fields. As an example, we recall here the expression of the Lennard-Jones potential, which is often used to describe short-ranged interactions in classical models. The Lennard-Jones potential describes the soft, attractive interactions of atoms or molecules, such as the van der Waals interactions. It is defined as follows:

$$V_{LJ}(r) := 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right],$$

where r is the distance between two particles, ε is the depth of the potential well, and σ is such that when $r = \sigma$, $V_{LJ}(r) = 0$. The Lennard-Jones potential admits one minima at $r = r^* = 2^{\frac{1}{6}}\sigma$ so that one can rewrite:

$$V_{LJ}(r) = 4\varepsilon \left[\left(\frac{r^*}{2^{\frac{1}{6}}r} \right)^{12} - \left(\frac{r^*}{2^{\frac{1}{6}}r} \right)^6 \right],$$

i.e

$$V_{LJ}(r) = 4\varepsilon \left[\frac{1}{4} \left(\frac{r^*}{r} \right)^{12} - \frac{1}{2} \left(\frac{r^*}{r} \right)^6 \right].$$

In other words,

$$V_{LJ}(r) = \varepsilon \left[\left(\frac{r^*}{r} \right)^{12} - 2 \left(\frac{r^*}{r} \right)^6 \right]. \quad (2.1)$$

Remark 4.

- ▷ Note that the Lennard-Jones potential (2.1) is singular. Nevertheless, as stated in Section 1.2, theoretical interaction potentials are taken to be at least in $\mathcal{C}^2(\mathcal{D})$.
- ▷ Several other empirical potentials exist, and we will discuss a generic model for the van der Waals interactions later on in Section 6.4.

2.1.3 Boundary conditions

We will work only with isolated systems. In this case, several choices of boundary conditions can be used for numerical simulations. We will restrict ourselves to the following two.

- ▷ The *periodic boundary conditions* are often used. In this case, particles interact with their physical neighbours, but also with their periodic images.
- ▷ As we shall see later on in Sections 6.4 and 6.5, one may apply reflecting boundary conditions when the system, or part of it, is constrained to a given region of the configuration space.

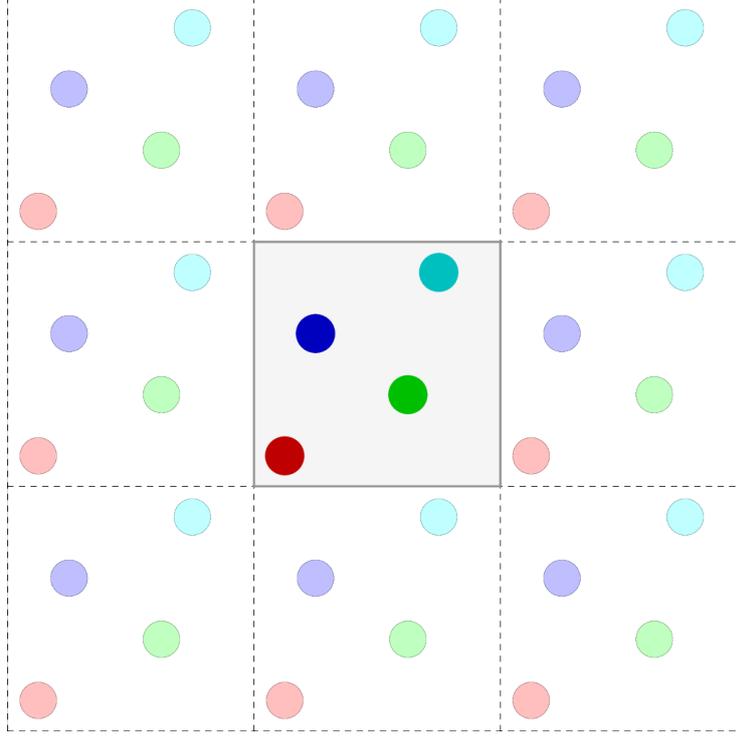


Figure 2.1 – Boundary conditions: the center cell is the reference cell where the real atoms live. The other cells are replicas obtained by translations of the reference cell, and the light-coloured atoms are the periodic images of the bright-coloured atoms.

2.2 Sampling the canonical measure

From now on, we will work within the canonical (NVT) ensemble. We will consider a system of N particles of positions $q = (\mathbf{q}_i)_{i \in [1, N]} \in \mathcal{D} \subset \mathbb{R}^{dN}$ and momenta $p = (\mathbf{p}_i)_{i \in [1, N]} \in \mathbb{R}^{dN}$, where d is the space dimension. Let us recall that the space \mathcal{D} is the configuration space and that we denote the phase space by $\mathcal{T}^*\mathcal{D} = \mathcal{D} \times \mathbb{R}^{3N}$.

We consider, as stated in Section 1.2, a separable Hamiltonian as given in (1.1):

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

where we assume that the potential energy V is sufficiently smooth, namely, $V \in \mathcal{C}^2(\mathcal{D})$. The associated measure, as said in section 1.4.1 is the canonical measure $\mu_\beta \propto e^{-\beta H}$ given by (1.7). We know that under the ergodic hypothesis, one can approach statistical averages of the form (1.4) by computing trajectorial averages of the form (1.3) for any observable $\varphi \in \mathcal{C}_b^\infty(\mathcal{T}^*\mathcal{D})$ using equality (1.5). As a consequence, one would like to be able to sample the canonical measure in order to compute thermodynamic averages $\mathbb{E}_{\mu_\beta}[\varphi]$ for any observable $\varphi \in \mathcal{C}_b^\infty(\mathcal{T}^*\mathcal{D})$. In order to do so, one can use *Langevin* or *overdamped Langevin dynamics*. These models are both given by continuous diffusion processes: we refer to the Annex A for the definition and basic properties

of Markov processes, along with a quick recap of diffusion processes.

2.2.1 Langevin dynamics

Langevin dynamics is a stochastic process $(q_t, p_t)_{t \geq 0}$ modelling the evolution in time of particles subject to friction and dissipation. It is a Hamiltonian system coupled with a reservoir of constant temperature T [56], termed a *thermostat* (not to be confused with a thermostat in the sense of classical physics). The process $(q_t, p_t)_{t \geq 0}$ satisfies the following stochastic differential equations:

$$\begin{cases} dq_t = \nabla_p H(q_t, p_t) dt, \\ dp_t = -\nabla_q H(q_t, p_t) dt - \gamma(q_t) \nabla_p H(q_t, p_t) dt + \sigma(q_t) dW_t, \end{cases} \quad (2.2)$$

where $(W_t)_{t \geq 0}$ is a standard $3N$ -dimensional Brownian motion, σ and γ are functions with values in $\mathcal{M}_{3N}(\mathbb{R})$, with γ being positive definite and σ being symmetric. In the scalar case, both parameters are assumed to be positive. The energy brought into the system by the fluctuation term $\sigma(q_t) dW_t$ is dissipated by the friction term $-\gamma(q_t) \nabla_p H(q_t, p_t) dt$. In most simulations, γ and σ are taken to be constants. In any case, these quantities satisfy the following *fluctuation-dissipation* condition:

$$\sigma \sigma^\top = \frac{2\gamma}{\beta}, \quad (2.3)$$

where $\beta := \frac{1}{k_B T}$.

When H is separable, one can rewrite the dynamics as:

$$\begin{cases} dq_t = M^{-1} p_t dt, \\ dp_t = -\nabla V(q_t) dt - \gamma(q_t) M^{-1} p_t dt + \sigma(q_t) dW_t. \end{cases}$$

The infinitesimal generator \mathcal{L} of the dynamics (2.2) reads:

$$\mathcal{L} = \{\cdot, H\} + \beta^{-1} e^{\beta H} \operatorname{div}_p (\gamma e^{-\beta H} \nabla_p \cdot),$$

where $\{\cdot, H\}$ denotes the Poisson bracket against the Hamiltonian H . For all test function φ , one has:

$$\{\varphi, H\} = \nabla_p H \cdot \nabla_q \varphi - \nabla_q H \cdot \nabla_p \varphi.$$

One can then show [56, Section 2.2.3] that:

- (i) Langevin dynamics (2.2) is reversible with respect to the canonical measure up to momenta reversal. Indeed, since the kinetic energy is symmetric, the momentum reversal $S(q, p) := (q, -p)$ leaves μ_β invariant. One consequently has for any pair of test functions $\varphi, \tilde{\varphi}$:

$$\int_{\mathcal{T}^* \mathcal{D}} \varphi \mathcal{L} \tilde{\varphi} d\mu_\beta = \int_{\mathcal{T}^* \mathcal{D}} \tilde{\varphi} \circ S \mathcal{L} (\varphi \circ S) d\mu_\beta.$$

- (ii) The canonical measure μ_β is an invariant measure for the Langevin dynamics, namely, for

all test function φ :

$$\int_{\mathcal{T}^*\mathcal{D}} \mathcal{L}\varphi \, d\mu_\beta = 0.$$

- (iii) The operator \mathcal{L} satisfies the parabolic Hörmander condition (A.5), and is by consequence *hypoelliptic*.

One may show that a consequence of (ii) and (iii) is that the process $(q_t, p_t)_{t \geq 0}$ is ergodic with respect to the canonical measure, see [56, Section 2.2.3.1]. In other words, for any initial condition (q_0, p_0) , for any observable $\varphi \in \mathcal{C}_b^\infty(\mathcal{T}^*\mathcal{D})$:

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \varphi(q_t, p_t) \, dt = \int_{\mathcal{T}^*\mathcal{D}} \varphi \, d\mu_\beta, \quad \text{almost surely.}$$

2.2.2 Overdamped Langevin dynamics

A much simpler dynamics also used to sample the canonical measure, or more precisely, the Boltzmann-Gibbs measure, is *overdamped Langevin dynamics*. It is the limit of Langevin dynamics as the friction γ goes to $+\infty$. In this *overdamped* limit, the friction forces become predominant, and one deals with a system with null average acceleration. Indeed, if one defines $\varepsilon := \gamma^{-1}$, Langevin dynamics with position and momentum vectors given by $(q_t^\varepsilon, p_t^\varepsilon) := (q_{t/\varepsilon}, p_{t/\varepsilon})$ reads:

$$\begin{cases} dq_t^\varepsilon = \frac{1}{\varepsilon} M^{-1} p_t^\varepsilon dt, \\ dp_t^\varepsilon = -\frac{1}{\varepsilon} \nabla V(q_t^\varepsilon) dt - \frac{1}{\varepsilon^2} M^{-1} p_t^\varepsilon dt + \sqrt{\frac{2\beta^{-1}}{\varepsilon^2}} dW_t, \end{cases}$$

and one can infer that when $\varepsilon \ll 1$, there is a separation of timescales for the positions and momenta, as the momenta $(p_t^\varepsilon)_{t \geq 0}$ will thermalise faster than the positions $(q_t^\varepsilon)_{t \geq 0}$. As a consequence, in the limit $\gamma \rightarrow +\infty$, momenta will be thermalised and the dynamics is then reduced to the positions. This leads us to define overdamped Langevin dynamics, the derivation of which can be found in details in [71, Chapter 6] and [56, Section 2.2.4]:

$$dq_t = -\nabla V(q_t) dt + \sqrt{2\beta^{-1}} dW_t, \quad (2.4)$$

where $(W_t)_{t \geq 0}$ is a $3N$ -dimensional Brownian motion.

Remark 5. *Langevin dynamics (2.2), where $\gamma < +\infty$, is often called underdamped Langevin dynamics, or kinetic Langevin dynamics.*

The infinitesimal generator \mathcal{L} of the dynamics (2.4) reads:

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1} \Delta,$$

and one can show that:

- (i) The overdamped Langevin dynamics is reversible with respect to the Boltzmann-Gibbs measure $\mu \propto e^{-\beta V}$ given by equation (1.9). One consequently has for any pair of test functions $\varphi, \tilde{\varphi}$:

$$\int_{\mathcal{D}} \varphi \mathcal{L} \tilde{\varphi} \, d\mu = \int_{\mathcal{D}} \tilde{\varphi} \mathcal{L} \varphi \, d\mu.$$

Note that here, reversibility is a notion from probability theory, which is not to be confused with the momentum reversibility property mentioned earlier in the previous Section.

- (ii) The Boltzmann-Gibbs measure μ is an invariant measure for the overdamped Langevin dynamics, namely, for all test functions φ :

$$\int_{\mathcal{D}} \mathcal{L}\varphi \, d\mu = 0.$$

- (iii) The operator \mathcal{L} satisfies the parabolic Hörmander condition (A.5), and is by consequence *hypoelliptic* (to be more precise, it is even *elliptic*).

As in the previous section, (ii) and (iii) yields the ergodicity of the process with respect to the measure μ .

Remark 6. Another way to study pathwise ergodicity is to look at the density π_t of law of the process. One can show that it satisfies the following Fokker-Planck equation:

$$\partial_t \pi_t = \mathcal{L}^* \pi_t, \tag{2.5}$$

where \mathcal{L}^* is the adjoint of the operator \mathcal{L} with respect to the $L^2(\mathrm{d}q)$ -scalar product. The Boltzmann-Gibbs measure μ is a stationary solution of (2.5). Provided $\partial_t - \mathcal{L}^* \mathcal{L}$ satisfies a parabolic Hörmander condition, i.e is *hypoelliptic*, then the ergodicity of the process (2.4) holds.

2.2.3 Numerical schemes

Knowing that both Langevin and overdamped Langevin dynamics are ergodic with respect to the canonical measure and Boltzmann-Gibbs measure respectively, one may wish to discretise them. Let us give a quick overview of how to do so. We denote by h the timestep, and we denote by q_n (resp. p_n) the position (resp. momentum) at time $t_n = nh$, with n a positive integer.

2.2.3.1 Overdamped Langevin dynamics: Euler-Maruyama scheme

There exists a simple way to discretise the overdamped Langevin dynamics, which is to use the classical Euler-Mayurama scheme [56]:

$$q_{n+1} = q_n - h\nabla V(q_n) + \sqrt{2\beta^{-1}h}G_n,$$

where G_n is a $3N$ -dimensional vector whose components are independent Gaussian random variables.

2.2.3.2 Langevin dynamics: SPV and BBK schemes

Langevin dynamics is more intricate to discretise, as it is the superposition of an Hamiltonian system with a stochastic process on the momenta. A key idea when discretising stochastic differential equations of the form

$$\mathrm{d}X_t = A(X_t)\mathrm{d}t + B(X_t)\mathrm{d}W_t,$$

where $(X_t)_{t \geq 0}$, A and B are of dimension m and $(W_t)_{t \geq 0}$ is a m -dimensional Brownian motion ($m \in \mathbb{N}^*$), is to establish a decomposition of the right-hand term's vector field into the sum of independant and exactly solvable parts. Each part can thus be discretised exactly, and one recovers the discretisation of the differential equation by combining the updates of each parts [54]. The resulting schemes are called *splitting schemes*, as decomposing the vector field amounts to decomposing the dynamics' infinitesimal generator as the sum of several differential operators. Splitting schemes are privileged schemes for the Langevin dynamics, as it can be shown that they are high-order methods [54, Section 2.4]. Let us list three well-known schemes:

▷ **Stochastic Position Verlet – [SPV]**

This method is obtained by re-writing the dynamics as:

$$d \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix} dt + \begin{pmatrix} 0 \\ -\nabla V(q)dt - \gamma p dt + \sigma M^{\frac{1}{2}} dW \end{pmatrix}$$

which leads to the following Stochastic Position Verlet scheme:

$$\begin{cases} q_{n+\frac{1}{2}} = q_n + \frac{h}{2} M^{-1} p_n \\ p_{n+1} = e^{-\gamma h} p_n - \eta \nabla V(q_n) + \sqrt{\beta^{-1}(1 - e^{-2\gamma h})} G_n \\ q_{n+1} = q_{n+\frac{1}{2}} + \frac{h}{2} M^{-1} p_{n+1}, \end{cases}$$

where G_n is a vector of $3N$ independent and identically distributed random variables, following a centered Gaussian law $\mathcal{N}(0, 1)$ and $\eta := (1 - e^{-\gamma h})/\gamma$. Note that $\lim_{\gamma \rightarrow +\infty} \eta = 0$ so that this scheme is not suitable for the discretization of the overdamped Langevin dynamics.

▷ **Brünger-Brooks-Karplus – [BBK]**

Rewriting (2.2) as:

$$d \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix} dt + \begin{pmatrix} 0 \\ -\nabla V(q) \end{pmatrix} dt + \begin{pmatrix} 0 \\ -\gamma p dt + \sigma M^{\frac{1}{2}} dW \end{pmatrix} \quad (2.6)$$

leads to the following Brünger-Brooks-Karplus scheme

$$\begin{cases} p_{n+\frac{1}{2}} = \frac{(1-\gamma h)}{2} p_n - \frac{h}{2} \nabla V(q_n) + \frac{1}{2} \sqrt{2\beta^{-1}\gamma h} M^{\frac{1}{2}} G_n \\ q_{n+1} = q_n + h M^{-1} p_{n+\frac{1}{2}} \\ p_{n+1} = \left(1 + \frac{\gamma h}{2}\right)^{-1} \left(p_{n+\frac{1}{2}} - \frac{h}{2} \nabla V(q_{n+1}) + \frac{1}{2} \sqrt{2\beta^{-1}\gamma h} M^{\frac{1}{2}} \tilde{G}_n\right), \end{cases}$$

where G_n and \tilde{G}_n are two $3N$ -dimensional and independent vectors following a centered Gaussian law $\mathcal{N}(0, 1)$. The next step's Gaussian random variable will be given by $G_{n+1} = \tilde{G}_n$.

▷ **Other schemes**

The decomposition (2.6) allows us to define several other schemes of the same family:

among them, BAOAB is the most renowned scheme, and we refer to [54, Section 7.3.1] for an exhaustive list of similar schemes and their definition.

Remark 7. *Note that the discretisation of the stochastic process $(q_t, p_t)_{t \geq 0}$ defines a Markov chain $(q_n, p_n)_{n \in \mathbb{N}}$, which we assume to still be ergodic with respect to an invariant measure μ_h . In which case, the distance between the measure μ_h and the target measure μ is of the order h . This yields a timestep-dependent error, called bias, in computing canonical averages against μ . Some algorithms relying the Metropolis-Hasting strategy can be used to get rid of such errors [54, Chapter 7], [56, Algorithms 2.9 and 2.11].*

Free energy calculations and sampling methods

The aims of molecular dynamics simulations are numerous, such as determining statistical averages of observables given by (1.4), or thermodynamic properties of the system at hand. Among them is the computation of the free energy, defined in Section 1.5.3. More precisely, the key quantity of interest is often the free energy difference of the system ΔA between two different equilibrium states. In order to estimate ΔA , one would need to be able to sample either the canonical measure μ_β or the Boltzmann-Gibbs measure μ both introduced in Section 1.4.1. By enabling us to sample one of those two measures, Langevin and overdamped Langevin dynamics allow for a proper sampling of the configuration space, and consequently would permit the computation of such free energy differences. However, one difficulty arises, that of *metastability*: due to the inherent nature of the system and of the target measure, the system may remain trapped in certain regions of the phase space, and one would have to wait for an unreasonable period of time in order to see the system explore the rest of the phase space. The sampling of the equilibrium measure by both dynamics is thus considerably slowed down. One consequently needs to build numerical methods bypassing metastability and enhancing the sampling of the phase space. A key idea used in this scope is to rely on a *reaction coordinate*, namely a *good* mapping of the state of the system, in a sense to be precised further on. Such mapping may then be used by an algorithm specifically designed to by-pass metastability, and one may eventually evaluate the target free energy difference ΔA . This chapter is dedicated to such algorithms.

In this chapter, we first properly define the notion of metastability in Section 3.1 and introduce in Section 3.2 the concept of *reaction coordinate*. We will distinguish two kinds of reaction coordinates, depending on the kind of transition considered for our system, which can be either conformational or alchemical. We will eventually review the question of how to properly choose a reaction coordinate. We then introduce in Section 3.3 the existing algorithms used to avoid metastability and enhance the sampling of the canonical measure. We put the emphasis on the *Adaptive Biasing Force method* and motivate the first problem to be treated in this thesis, namely its robustness in the case where the interaction force between the system's components is not conservative. We then focus on alchemical transitions and introduce in Section 3.4 the motivation behind computing free energy differences in such a setting, and proceed to list the

available methods used in this scope. Among these methods, we will focus on the λ -dynamics method and review its limitations. Eventually, we introduce the concept of the *Orthogonal Space Random Walk* (OSRW) sampling algorithm whose goal is to bypass the λ -dynamics method's inherent issues. We then describe the second problem of this thesis, which was to build a proper, reproducible sampling algorithm competing with the OSRW method.

From now on, we will work under the following assumption:

Assumption 2. *The dynamics of the system is given by overdamped Langevin dynamics (2.4) on $\mathcal{D} \subset \mathbb{R}^{dN}$ where d is the space dimension and N the number of particles.*

3.1 What is metastability?

3.1.1 Definition and examples

A system is said to be *metastable* when it may remain trapped in specific regions of the phase space, called *metastable regions*, for a very long period of time, before moving elsewhere. The overdamped Langevin dynamics we are working with is widely used, and known to be metastable. A first explanation is that the potential energy V has several local minima, and the system's components remain in potential wells for a great amount of time before jumping the energetic barrier and reaching another potential well, as shown in Figure 6.3.

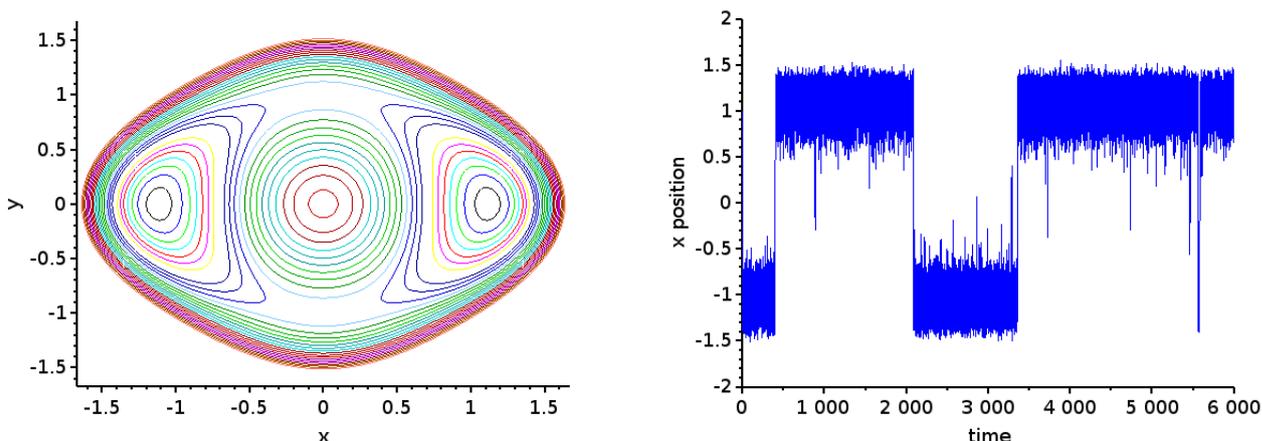


Figure 3.1 – Contour plot (left) of the potential given by the 2-dimensional potential $V(x, y) = \frac{1}{6}(4(1 - x^2 - y^2)^2 + 2(x^2 - 2)^2 + ((x + y)^2 - 1)^2 + ((x - y)^2 - 1)^2)$, and projected trajectory on x (right).

Metastability is commonly observed in biophysical and chemical phenomena. We list here two classical examples.

- **Example 1: configurational isomerism** – Two molecules are *stereoisomers* when they have the same formula and bonds, but their functional groups are in different orientations in the three-dimensional space. A classical example are the isomers of the 1,2-dichloroethene molecule $C_2H_2Cl_2$: the *cis*-isomer of the molecule is the conformation where the functional groups (i.e the

chlorine and hydrogen atoms, separately) are on the same side of a plane including the double bond, whereas the *trans*-isomer is the conformation where they are on opposing sides.

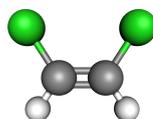


Figure 3.2 – *cis*-isomer of the 1,2-dichloroethene molecule.

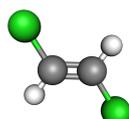


Figure 3.3 – *trans*-isomer of the 1,2-dichloroethene molecule.

Here, the system's components are the atoms, and to each configurational isomer is associated a local minimum of the system's potential energy. As a consequence, the system is metastable, and observing the change of configuration is considered a *rare event*. Let us illustrate it with the *cis-trans* isomers of the azobenzene $C_{12}H_{10}N_2$ molecule, which can undergo a change of configuration upon irradiation with light. In the dark, the molecule tends to be in the short *cis*-isomer configuration, and when irradiated, it undergoes a transition towards the longer *trans*-isomer configuration, which is thermally favoured [37].

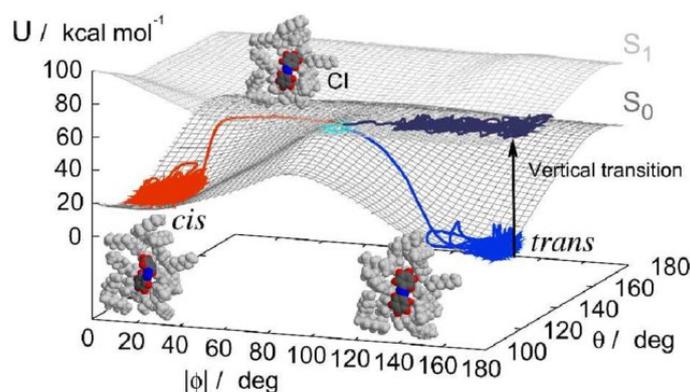
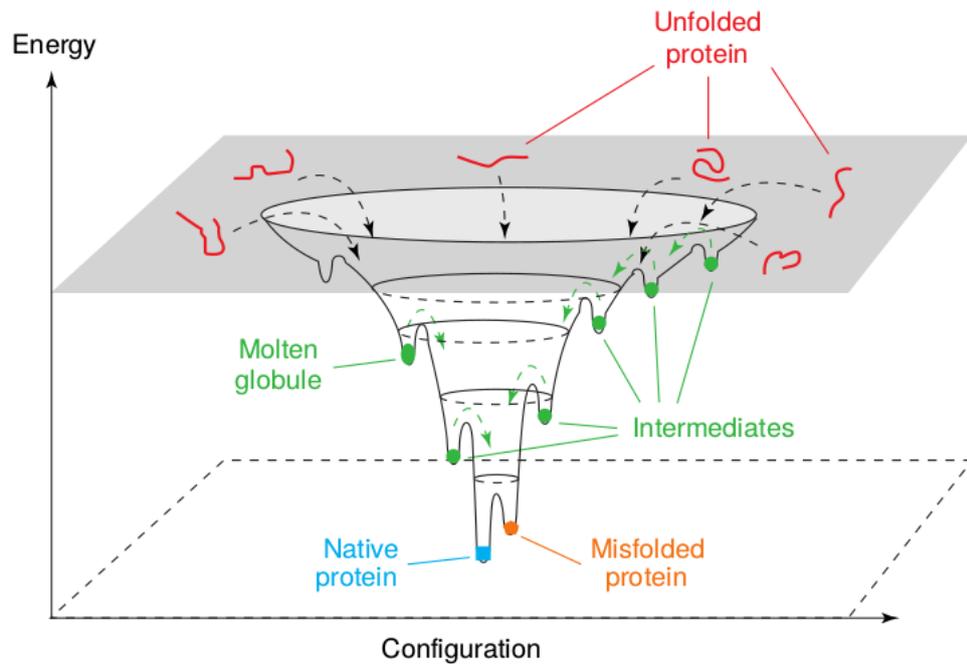


Figure 3.4 – Energy landscapes for a configurational transition of the azobenzene molecule from the *cis*-isomer configuration (in red) to the *trans*-isomer configuration (in blue). The two energy surfaces corresponds to the ground state S_0 and the first excited (singlet) state S_1 of the molecule. The molecule first needs to reach either the first excited state S_1 (or even the second S_2) to relax into either the *cis* or the *trans* configuration. The transition configuration is shown at the conical intersection CI, and the cyan region indicates the region of the excited energy landscape S_1 where the probability of configurational change is higher than 0.1. *Source: [85]*

• **Example 2: protein folding** – Other examples of metastability arise in the context of *protein folding*. A protein is a biomolecule composed of one or several amino acids chains, which can have many different configurations in the three-dimensional space. A protein can be unfolded, in which case its amino acids chains will be fully “uncoiled”, but will not be stable. An unfolded protein will eventually fold itself, depending on the interactions between the amino acids and the environment. According to the kind of amino acids composing its chains, the protein will either curl up around itself, avoiding contact between its core and the environment, or fold into an alternative configuration which allows higher interactions with the protein’s surroundings. A classical hypothesis is that one can view the protein’s journey, from an unfolded state to one of its curled-up, folded states, as an energy minimization problem, where the protein’s energy landscape is shaped as a funnel. Eventual potential wells can be located at the walls of the funnel, and the energy landscapes thus have several local minima. The global minimum of the energy landscape then corresponds to the protein’s *native* structure, i.e the properly assembled and functional form of the protein which is stabilized by several weak-interactions including the formation of hydrogen bonds and van der Waals interactions, whereas local minima correspond to denatured configurations.



T/BS

Figure 3.5 – Schematic representation of a funnel-shaped energy landscape. *Source: [77]*

Remark 8. Another explanation for the metastability of a system is the presence of entropic barriers. We refer to [56, Section 1.3.3.2] for a proper discussion on entropic and energetic barriers. For the sake of simplicity, and since our goal is here to give a hint on what metastability is, we will only talk of energetic wells, keeping in mind that doing so is reductive.

In a probabilistic setting metastability is linked to the *multimodality* of the Boltzmann-Gibbs measure μ given by equation (1.9): the probability density function of μ has several local maxima (or *modes*), which coincide with the local minima of the potential energy V . Consequently, high

probability regions corresponds to potential wells and low probability regions correspond to energy peaks.

3.1.2 Quantifying metastability

3.1.2.1 Convergence of averages and convergence of the law

Metastability has a direct impact on a system's rate of convergence towards the equilibrium. In particular, given an observable φ , one may be interested into quantifying the rate of convergence, called the *mixing time*, of the following limit:

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \int_0^T \varphi(q_t) dt = \mathbb{E}_\mu[\varphi], \quad (3.1)$$

where the trajectorial average of the observable φ converges to its statistical mean $\mathbb{E}_\mu[\varphi]$. One may also be interested in the rate of convergence, called *relaxation time*, of the law of the process $(X_t)_{t \geq 0}$ to the Boltzmann-Gibbs measure μ . Such convergence can be expressed as follows: for any continuous and bounded test function φ ,

$$\lim_{T \rightarrow +\infty} \mathbb{E}[\varphi(X_T)] = \mathbb{E}_\mu[\varphi]. \quad (3.2)$$

When the system is metastable, both convergences (3.1) and (3.2) are expected to be slowed down. In others words, metastability can be quantified by substantially long mixing time and relaxation times.

Note that both properties (3.1) and (3.2) are referred to as ergodicity properties. The long-time convergence of trajectorial averages (3.1) is often obtained after the long-time convergence of the law consider at the convergence of the law of the considered process. Here, we focus on overdamped Langevin dynamics (2.4). Its infinitesimal generator is given by:

$$\mathcal{L} = -\nabla V \cdot \nabla + \beta^{-1} \Delta,$$

and if we denote by π_t the probability density of the law of the process, as mentioned in Section A.4, it satisfies the following Fokker-Planck equation:

$$\partial_t \pi_t = \mathcal{L}^* \pi_t,$$

where $\mathcal{L}^* f = \operatorname{div}(\nabla V \cdot f + \beta^{-1} \nabla f)$ is the adjoint of the infinitesimal generator. Quantifying metastability then amounts to quantifying the rate of convergence of (3.2), i.e the rate of convergence of the density π_t towards the Boltzmann-Gibbs distribution $\mu \propto \exp(-\beta V)$. In order to quantify this convergence properly, let us introduce the key notions of *relative entropy*, *Fisher information* and *logarithmic Sobolev inequality*.

Remark 9. *There exist other approaches to the quantification of metastability. Among them, one relies on the concept of quasi-stationary distributions and another one on the Eyring-Kramers formula. We refer to [55, Section 3] and [16] for proper definitions and mathematical proofs.*

3.1.2.2 Relative entropy and logarithmic Sobolev inequality

For ν, μ two probability measures on the same space E , we will denote by $\nu \ll \mu$ the absolute continuity of ν with respect to μ . Now consider the relative entropy of ν with respect to μ :

$$\mathcal{H}(\nu|\mu) = \begin{cases} \int_E \ln \left(\frac{d\nu}{d\mu} \right) d\nu & \text{if } \nu \ll \mu, \\ +\infty & \text{otherwise.} \end{cases}$$

We recall the Csiszár-Kullback inequality:

$$\|\nu - \mu\|_{TV} \leq \sqrt{2\mathcal{H}(\nu|\mu)}, \quad (3.3)$$

where $\|\cdot\|_{TV}$ stands for the total variation norm. Let us recall that for two probability measures ν and μ on a probability space (Ω, \mathcal{F}) , the total variation norm is given by:

$$\|\nu - \mu\|_{TV} := 2 \sup_{A \in \mathcal{F}} |\nu(A) - \mu(A)|.$$

In particular, while the relative entropy is not a distance (it lacks the symmetry property), its convergence to zero implies the convergence in total variation norm of ν towards μ .

Similarly, let us define the Fisher information: for $\nu \ll \mu$,

$$I(\nu|\mu) = \int_E \left| \nabla \ln \left(\frac{d\nu}{d\mu} \right) \right|^2 d\nu.$$

The probability measure μ is said to satisfy a Logarithmic Sobolev Inequality LSI(ρ) of constant $\rho > 0$ if:

$$\forall \nu \ll \mu, \quad \mathcal{H}(\nu|\mu) \leq \frac{1}{2\rho} I(\nu|\mu).$$

From [70], if μ satisfies a log-Sobolev inequality with constant $\rho > 0$, then it also satisfies the so-called Talagrand inequality $\mathcal{T}(\rho)$ with constant $\rho > 0$:

$$\forall \nu \ll \mu, \quad W_2^2(\nu, \mu) \leq \frac{2}{\rho} \mathcal{H}(\nu|\mu), \quad (3.4)$$

where $W_2(\nu, \mu)$ is the Wasserstein distance with quadratic cost between the probability measures ν and μ . More precisely, if ν and μ are defined on a general Riemannian manifold Ω :

$$W_2^2(\nu, \mu) = \inf_{\pi \in \Pi(\nu, \mu)} \int_{\Omega \times \Omega} \omega(x, y)^2 d\pi(x, y),$$

where ω is the geodesic distance on Ω , and $\Pi(\nu, \mu)$ is the set of coupling probability measures, i.e probability measures on $\Omega \times \Omega$ whose marginals are ν and μ respectively. Note that since the content of the integral is positive in any case, the quantity $W_2^2(\nu, \mu)$ is well defined. In particular, it is finite provided the measures ν and μ have a second-order moment.

From now on, we will slightly abuse notations and write $I(\nu|\mu)$, $\mathcal{H}(\nu|\mu)$ or $W_2(\nu|\mu)$ both in the case where ν and μ are probability measures, or probability density functions. The density of the Boltzmann-Gibbs measure is thus denoted by μ . Now, let us assume μ satisfies a logarithmic

Sobolev inequality of constant ρ . Denote by π_t the density of the law of the process (2.4), which satisfies the Fokker-Planck equation (2.5) rewritten as follows:

$$\begin{aligned}\partial_t \pi_t &= \mathcal{L}^* \pi_t \\ &= \operatorname{div} (\nabla V \pi_t + \beta^{-1} \nabla \pi_t) \\ &= \operatorname{div} (\beta \beta^{-1} \nabla V e^{-\beta V} e^{\beta V} \pi_t + \beta^{-1} \nabla \pi_t e^{-\beta V} e^{\beta V}) \\ &= \beta^{-1} \operatorname{div} (e^{\beta V} (\nabla \pi_t e^{-\beta V} - \pi_t \nabla e^{-\beta V})) \\ &= \beta^{-1} \operatorname{div} \left(e^{-\beta V} \frac{(\nabla \pi_t e^{-\beta V} - \pi_t \nabla e^{-\beta V})}{(e^{-\beta V})^2} \right),\end{aligned}$$

so that

$$\partial_t \pi_t = \beta^{-1} \operatorname{div} \left(\mu \nabla \left(\frac{\pi_t}{\mu} \right) \right).$$

Let us then consider the time-derivative of the relative entropy of the law π_t with respect to the Boltzmann-Gibbs measure μ :

$$\begin{aligned}\partial_t \mathcal{H}(\pi_t | \mu) &= \partial_t \int_{\mathcal{D}} \left[\ln \left(\frac{\pi_t}{\mu} \right) \pi_t \right] \\ &= \int_{\mathcal{D}} \left[\partial_t \ln \left(\frac{\pi_t}{\mu} \right) \pi_t + \ln \left(\frac{\pi_t}{\mu} \right) \partial_t \pi_t \right] \\ &= \int_{\mathcal{D}} \partial_t \pi_t + \int_{\mathcal{D}} \left[\ln \left(\frac{\pi_t}{\mu} \right) \beta^{-1} \operatorname{div} \left(\mu \nabla \left(\frac{\pi_t}{\mu} \right) \right) \right].\end{aligned}$$

Since $\int_{\mathcal{D}} \pi_t = 1$, and since the configuration space is compact, one has:

$$\begin{aligned}\partial_t \mathcal{H}(\pi_t | \mu) &= -\beta^{-1} \int_{\mathcal{D}} \left[\mu \nabla \left(\frac{\pi_t}{\mu} \right) \cdot \nabla \ln \left(\frac{\pi_t}{\mu} \right) \right] \\ &= -\beta^{-1} \int_{\mathcal{D}} \left[\frac{\mu}{\pi_t} \nabla \left(\frac{\pi_t}{\mu} \right) \cdot \nabla \ln \left(\frac{\pi_t}{\mu} \right) \pi_t \right] \\ &= -\beta^{-1} \int_{\mathcal{D}} \left[|\nabla \ln \left(\frac{\pi_t}{\mu} \right)|^2 \pi_t \right],\end{aligned}$$

hence

$$\partial_t \mathcal{H}(\pi_t | \mu) = -\beta^{-1} I(\pi_t | \mu). \quad (3.5)$$

A direct consequence of (3.5) is that since μ satisfies LSI(ρ), one has

$$\partial_t \mathcal{H}(\pi_t | \mu) \leq -2\beta^{-1} \rho \mathcal{H}(\pi_t | \mu),$$

and one can conclude using Gronwall's lemma, that for all $t \geq 0$, and for any initial condition π_0 :

$$\mathcal{H}(\pi_t | \mu) \leq \mathcal{H}(\pi_0 | \mu) e^{-2\beta^{-1} \rho t}.$$

The Csiszár-Kullback inequality (3.3) then yields for all $t \geq 0$:

$$\|\pi_t - \mu\|_{TV} \leq \sqrt{2\mathcal{H}(\pi_t | \mu)} \leq \sqrt{2\mathcal{H}(\pi_0 | \mu)} e^{-\beta^{-1} \rho t}.$$

One can then conclude that the density π_t converges exponentially fast in the long-time limit towards the density of the Boltzmann-Gibbs measure μ in L^1 -norm with rate $\beta^{-1}\rho$. The exponential rate allows us to quantify metastability: for a fixed temperature T (and hence a fixed β), the smaller the logarithmic Sobolev constant ρ is, the greater the metastability [56, Section 2.3.2.1].

3.1.2.3 Properties and criteria for logarithmic Sobolev inequalities

Since logarithmic Sobolev inequalities are a key tool for quantifying the metastability of a given process, one may be interested in establishing if a given stationary measure of the Boltzmann-Gibbs form, i.e. $\mu \propto e^{-\beta V}$, satisfies a logarithmic Sobolev inequality. Let us first list an interesting property:

Property 1. If $\mu = \prod_{n=1}^N \mu_n$ and each measure $\mu_n(q) dq$ satisfies $LSI(\rho_n)$ with $\rho_n > 0$ then μ satisfies $LSI(\rho)$ with $\rho = \min\{\rho_n \mid n \in \llbracket 1, N \rrbracket\}$.

Now, let us state two criteria for a measure π_∞ to satisfy a logarithmic Sobolev inequality [56].

Proposition 1 (Bakry-Emery criterion). Let $V : \mathbb{R}^{dN} \rightarrow \mathbb{R}$ be an α -convex function, namely a function such that for some $\alpha > 0$,

$$\forall x, y \in \mathbb{R}^{dN}, \quad y^\top \nabla^2 V(x) y \leq \alpha |y|^2,$$

where $\nabla^2 V$ denotes the Hessian matrix of V . The measure $\pi_\infty \propto e^{-V}$ then satisfies a $LSI(\rho)$ with $\rho \geq \alpha$.

Proposition 2 (Holley-Strook criterion). Let V be a function such that the measure $\pi_\infty \propto e^{-V}$ satisfies a $LSI(\rho)$ with $\rho > 0$. Consider a bounded function \tilde{V} . Then, the measure $\tilde{\pi}_\infty \propto e^{-(V-\tilde{V})}$ satisfies $LSI(\tilde{\rho})$ with

$$\tilde{\rho} \geq \rho e^{\inf(\tilde{V}) - \sup(\tilde{V})}.$$

3.2 Transition coordinates: configurational and alchemical cases

Metastability is a major obstacle in MD simulations. Now that we know it can be quantified, we wish to answer the question: how does one avoid metastability? Is there a way to **reduce** metastability? Many algorithms have been designed in order to do so. Among them, some rely on the concept of *transition coordinates*, or *reaction coordinates*, or *collective variables*. Most of the time the three denominations are used without distinction, even though the terms do not have the same meaning depending on the system at hand. Transition coordinates, often denoted by ξ , are mappings designed to provide a *coarse-grained information* on the system's state. Let

us take the example 1 in Section 3.1.1 of configurational isomerism. The two possible configurations of the 1,2-dichloroethane molecule are the two *cis-trans* isomers of Figures 3.2 and 3.5. To each isomer corresponds a potential well, and one can grasp that the change of configuration solely relies on the fact that the dihedral angle between the double-bond and the chlorine atoms changes, as shown in Figure 3.6. In a simulation, the change of dihedral angle would be the only slow motion of the molecule's dynamics: a good idea is thus to consider the reaction coordinate ξ which maps the position q onto the corresponding dihedral angle.

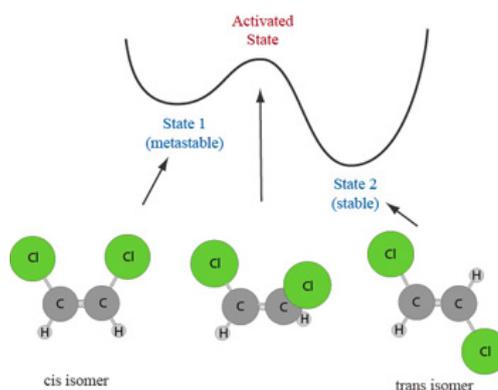


Figure 3.6 – Representation of the energy landscape of the 1,2-dichloroethane molecule *Source: <http://chemcollective.org/chem/entropy/reactcoord.php>*

A transition coordinate can be very different depending on the system considered: it can be the dihedral angle of a molecule, the signed distance to a hypersurface of \mathcal{D} , or, in the case of proteins or small biomolecules, it can also be the *root-mean-square deviation* of atomic positions, namely the average distance between the protein's backbone atoms.

Choosing a transition coordinate is inherent to the kind of transition one wishes to study. Indeed, as said above, one aim of MD simulations is to compute free energy differences between two equilibrium states. One has the intuition that a transition coordinate might be able to properly characterise the transition of the system from a given initial state to a target final state. Let us list the two different kinds of possible transitions, and how one defines the free energy difference in each case.

3.2.1 Conformational transitions

In *conformational transitions*, the system is characterised by its Hamiltonian H given by (1.1). In this case, the term *reaction coordinate* is preferred, and one considers ξ to be a geometric function of the positions: it can be a dihedral angle, a distance, and so on. Namely, a *reaction coordinate* is a mapping of the microstates built to capture the slow components of the dynamics. It is a mapping of the form

$$\xi : \mathcal{D} \subset \mathbb{R}^{dN} \rightarrow \mathcal{M}, \quad (3.6)$$

where \mathcal{M} is a manifold of lesser dimension $m \ll dN$. If for example ξ is the dihedral angle of a molecule then $\mathcal{M} = \mathbb{T}$, whereas if it is the signed distance to a hypersurface of \mathcal{D} , then $\mathcal{M} = \mathbb{R}$.

In any case, $\xi(q) \in \mathcal{M}$ is the macroscopic state of a microscopic state $(q, p) \in \mathcal{T}^*\mathcal{D}$. The reaction coordinate ξ then gives a foliation of the configuration space:

$$\mathcal{D} = \bigsqcup_{z \in \mathcal{M}} \Sigma_z = \bigsqcup_{z \in \mathcal{M}} \{q \in \mathcal{D} \mid \xi(q) = z\}.$$

We assume that the submanifolds $(\Sigma_z)_{z \in \mathcal{M}}$ are simply connected. Now, denoting by σ_{Σ_z} the measure on Σ_z induced by the Lebesgue measure on \mathcal{D} , one can define the measure $\delta_{\xi(q)-z}(dq)$ by

$$\delta_{\xi(q)-z}(dq) = \frac{1}{\sqrt{\det G(q)}} \sigma_{\Sigma_z}(dq),$$

where $G = (\nabla \xi)^\top \nabla \xi$, in other words,

$$G_{i,j} = \nabla \xi_i \cdot \nabla \xi_j, \quad \text{for all } (i, j) \in \llbracket 1, m \rrbracket^2. \quad (3.7)$$

We make the assumption that ξ is such that $\det G > 0$, which is equivalent to $\text{rank}(\nabla \xi) = m$, with $\nabla \xi$ denoting the Jacobian matrix of the mapping ξ , i.e $\nabla \xi = (\partial_i \xi_j)_{(i,j) \in \llbracket 1, dN \rrbracket \times \llbracket 1, m \rrbracket}$. Note that in any case, ξ will be assumed to be differentiable throughout this thesis. Furthermore, we assume that

$$\sup_{i \in \llbracket 1, m \rrbracket} \left| \sum_{j=1}^m G_{i,j}^{-1}(q) \nabla \xi_j(q) \right| < +\infty, \quad \forall q \in \mathcal{D}.$$

This allows us to define the *marginal distribution* μ^ξ of the Boltzmann-Gibbs measure with respect to ξ :

$$\mu^\xi(dz) = \xi_* \mu(dq) = \left(Z_\mu^{-1} \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq) \right) dz, \quad (3.8)$$

where Z_μ is given in (1.9). The notation $\xi_* \mu(dq)$ indicates that μ^ξ is the image of μ by ξ . Assuming V and ξ are such that $Z_{\Sigma_z} := \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq) < +\infty$, the free energy A of the system is then the log-density of the marginal distribution:

$$e^{-\beta A(z)} dz = \mu^\xi(dz),$$

in other words:

$$A(z) = -\beta^{-1} \ln \left(Z_\mu^{-1} \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq) \right) \quad \forall z \in \mathcal{M}. \quad (3.9)$$

If one considers two end states z_0 and z_1 in \mathcal{M} , the free energy difference is given by

$$\Delta_{0 \rightarrow 1} A = A(z_1) - A(z_0) = -\beta^{-1} \ln \left(\frac{\int_{\Sigma_{z_1}} e^{-\beta V(q)} \delta_{\xi(q)-z_1}(dq)}{\int_{\Sigma_{z_0}} e^{-\beta V(q)} \delta_{\xi(q)-z_0}(dq)} \right)$$

and one can interpret the above quantity as the relative likelihood of states in the final set Σ_1 compared to states in the initial set Σ_0 .

The mapping $z \mapsto A(z)$ is called the *potential of mean force*: indeed, its gradient ∇A is often called the *mean force*, as it can be seen as an average force applied to the system for constant values of ξ . One has the following result [56, Lemma 3.9]:

$$\nabla A(z) = \int_{\Sigma_z} F d\mu^\xi(\cdot|z), \quad (3.10)$$

where $\mu^\xi(dq|z)$ is the Boltzmann-Gibbs measure conditioned to a fixed value z of the reaction coordinate ξ , and F is the *local mean force*, which is the vector with components $(F_i)_{i \in \llbracket 1, m \rrbracket}$ given by:

$$F_i = \sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \cdot \nabla V - \beta^{-1} \operatorname{div} \left(\sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \right),$$

where $G_{i,j}^{-1}$ denotes the (i, j) -component of the inverse of the matrix G defined above.

Remark 10. *All of the above reasoning also applies to Langevin dynamics, with the free energy of the system defined as*

$$e^{-\beta A(z)} dz = \mu_\beta^\xi(dz),$$

where $\mu_\beta^\xi(dz)$ is the marginal distribution of the canonical measure μ_β given by (1.7),

$$\mu_\beta^\xi(dz) = \xi \star \mu_\beta(dq) = \left(Z_{\mu_\beta}^{-1} \int_{\Sigma_z \times \mathbb{R}^{dN}} e^{-\beta H(q,p)} \delta_{\xi(q)-z}(dq) dp \right) dz.$$

3.2.2 Alchemical transitions

An alchemical transition is a chemical reaction where the system evolves from an initial state (*reactant* state) A towards a final state (*product* state) B , and during which the nature of the system's components may be changed. For example, one may be interested in changing the nature of a molecule's atom or add/remove components to/from a chemical compound. Such transitions are indexed by an external parameter λ , independent of the microstate (q, p) . As a consequence, the transition coordinate is not a geometric mapping of the microstate $q \in \mathcal{D}$: the coupling parameter λ is a scalar transition coordinate whose values are between 0 and 1, $\lambda = 0$ characterising the initial state and $\lambda = 1$ the final state. Note that intermediate states, characterised by intermediate values $\lambda \in (0, 1)$ are **allowed to not make sense physically**: alchemical transitions are **purely numerical transitions**. Classical examples of alchemical transitions (or *reactions*) include the solvation of a chemical moiety (an ion, a molecule or a protein) in a box of solvent, the docking of a protein unto a given receptor, or the gradual modification of physical constants such as the intensity of a magnetic field applied to a spins system or the constants in empirical forces (such as ε and σ in the Lennard-Jones potential (2.1)).

The scalar transition coordinate λ can be seen as a classical reaction coordinate provided one considers the extended system of extended microstate $\mathbf{x} = (q, p; \lambda) \in \mathcal{T}^* \mathcal{D} \times [0, 1]$. The reaction coordinate is then $\xi(\mathbf{x}) = \lambda$. To each value of $\lambda \in (0, 1)$ corresponds an extended Hamiltonian $H(\cdot, \cdot; \lambda)$ which characterises the system. Working within the canonical ensemble, the canonical measure according to which the microstates will be distributed is given by:

$$\mu_{\beta,\lambda}(dqdp) := Z_\lambda^{-1} e^{-\beta H(q,p;\lambda)} dqdp, \quad Z_\lambda = \int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;\lambda)} dqdp.$$

Implicitly, the measures $(\mu_{\beta,\lambda})_{\lambda \in [0,1]}$ are defined at fixed $\beta = (k_B T)^{-1}$ i.e, at fixed absolute temperature. The logical thermodynamic ensemble to work with is thus the isobaric-isothermal ensemble (NPT). As a consequence, the natural energy to consider is the *Gibbs free energy* or *free enthalpy* G defined in Section 1.5.2. We wish to work with the Helmholtz free energy A of the canonical ensemble (NVT). Let us show that under certain assumptions, one can assume the two free energies to be somewhat equal. In classical mechanics, the Helmholtz free energy is given by

$$A(\lambda) = -\beta^{-1} \ln(Z_\lambda), \quad Z_\lambda = c \int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;\lambda)} dqdp, \quad (3.11)$$

where $c = (N!h^{dN})^{-1}$, with h denoting the Planck's constant [40, Equation 6c]. The Gibbs free energy is given by

$$G(\lambda) = -\beta^{-1} \ln(\mathcal{Z}_\lambda), \quad \mathcal{Z}_\lambda = c \int \int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;\lambda) - \beta PV} dqdpdV,$$

where P is the pressure, and V is the volume of the system, which should not be confused with the system's potential energy [23]. Let us recall the relationship between the two free energies:

$$G(\lambda) = A(\lambda) + PV.$$

One then has :

$$\begin{aligned} \frac{dA(\lambda)}{d\lambda} &= \frac{\int_{\mathcal{T}^*\mathcal{D}} \partial_\lambda H e^{-\beta H(q,p;\lambda)} dqdp}{\int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;\lambda)} dqdp} = \mathbb{E}_{\mu_{\beta,\lambda}} [\partial_\lambda H] \\ \frac{dG(\lambda)}{d\lambda} &= \frac{\int \int_{\mathcal{T}^*\mathcal{D}} \partial_\lambda H e^{-\beta H(q,p;\lambda) - \beta PV} dqdpdV}{\int \int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;\lambda) - \beta PV} dqdpdV} = \mathbb{E}_{\mu_{(NPT),\lambda}} [\partial_\lambda H] \end{aligned}$$

where $\mu_{(NPT),\lambda} \propto \exp(-\beta H(q,p;\lambda) - \beta PV)$ is the equilibrium measure associated to the (NPT) ensemble. The free energy (resp. free enthalpy) difference between the initial state A and the final state B is then:

$$\Delta_{0 \rightarrow 1} A(V, T) = A(1) - A(0) = -\beta^{-1} \ln \left(\frac{Z_1}{Z_0} \right) = -\beta^{-1} \ln \frac{\int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;1)} dqdp}{\int_{\mathcal{T}^*\mathcal{D}} e^{-\beta H(q,p;0)} dqdp}$$

$$\text{i.e } \Delta_{0 \rightarrow 1} A(V, T) = \int_0^1 \mathbb{E}_{\mu_{\beta,\lambda}} [\partial_\lambda H] d\lambda$$

and

$$\Delta_{0 \rightarrow 1} G(P, T) = G(1) - G(0) = -\beta^{-1} \ln \left(\frac{Z_1}{Z_0} \right) = -\beta^{-1} \ln \frac{\int \int_{\mathcal{T}^* \mathcal{D}} e^{-\beta H(q, p; 1) - \beta P V} dq dp dV}{\int \int_{\mathcal{T}^* \mathcal{D}} e^{-\beta H(q, p; 0) - \beta P V} dq dp dV}$$

i.e. $\Delta_{0 \rightarrow 1} G(P, T) = \int_0^1 \mathbb{E}_{\mu_{(NPT), \lambda}} [\partial_\lambda H] d\lambda.$

Remark 11. *We shall keep in mind the following notation, that is often found in the MD literature: for a given thermodynamic ensemble, i.e for a given probability measure on the phase space ν ,*

$$\mathbb{E}_\nu [\cdot] = \langle \cdot \rangle_\nu.$$

As said above, the free enthalpy difference $\Delta_{0 \rightarrow 1} G$ can be naturally computed in the scope of a simulation run with constant pressure and temperature. Nonetheless, it is possible to calculate it with a simulation done in the canonical ensemble (NVT) [33]. By denoting P and V (resp. P_B and V_B) the pressure and volume of the initial state A (resp. final state B), one has [33, Equation (3.139)]:

$$\Delta_{0 \rightarrow 1} G(P) = \Delta_{0 \rightarrow 1} A(V) - \int_P^{P_B} (V_B(\mathbf{p}) - V) d\mathbf{p}.$$

The correction is approximated by

$$- \int_P^{P_B} (V_B(\mathbf{p}) - V) d\mathbf{p} = -\frac{1}{2} (P_B - P) \Delta V = \frac{(\Delta V)^2}{2\kappa V}$$

where ΔV is the volume variation at constant pressure P , and κ is the isothermal compressibility constant. This term is usually small, even negligible. For example, adding a water molecule in a water box of constant volume V containing 1000 water molecules will produce a pressure variation of around 22 bar, and a Helmholtz free energy correction only of the order of -1 kJ.mol^{-1} [33].

We will from now make the following assumption:

Assumption 3.

1. The extended Hamiltonian is separable, so that the equilibrium measure is of the Boltzmann-Gibbs form:

$$\mu_\lambda(\mathrm{d}q) = \frac{e^{-\beta V(q;\lambda)}}{\int_{\mathcal{D}} e^{-\beta V(q;\lambda)} \mathrm{d}q} \mathrm{d}q.$$

i.e $\mu_\lambda(\mathrm{d}q)$ is the marginal distribution of the canonical distribution $\mu_{\beta,\lambda}$ with respect to the momenta.

2. We work within the approximation

$$\Delta_{0 \rightarrow 1} A(V, T) \approx \Delta_{0 \rightarrow 1} G(P, T)$$

and $\frac{\mathrm{d}A(\lambda)}{\mathrm{d}\lambda} = \mathbb{E}_{\mu_\lambda} [\partial_\lambda H] = \mathbb{E}_{\mu_\lambda} [\partial_\lambda V]$.

3. The free energy difference between the initial and final states is given by

$$\Delta_{0 \rightarrow 1} A = -\beta^{-1} \ln \frac{\int_{\mathcal{D}} e^{-\beta V(q;1)} \mathrm{d}q}{\int_{\mathcal{D}} e^{-\beta V(q;0)} \mathrm{d}q}$$

3.2.3 Choosing the reaction coordinate

Except for the alchemical case where λ is a simple, scalar transition coordinate, designing a good reaction coordinate is a difficult problem, and one often relies on chemical intuition. Nonetheless, recent works have attempted to tackle the question of automatic learning of the reaction coordinate. We refer to [32],[29] and [81] for more insight on collective variables learning.

3.3 Enhanced sampling methods

Let us recall our main problem: the overdamped Langevin dynamics that we wish to use in order to sample the Boltzmann-Gibbs measure (1.9) is metastable. We thus need to build numerical methods to avoid metastability and enhance the sampling of the phase space. Among the available methods designed in this scope, we can distinguish between two classes of algorithms: those which use transition coordinates, and those which do not. Let us focus on the former. The key idea of algorithms using a transition coordinate ξ to bypass metastability is to manipulate either the interaction force between the system's components, or the potential energy, by adding a bias depending on ξ . Let us give a quick list of these methods. For the following subsections, we will consider the more general case of *configurational transitions* and denote by ξ the reaction coordinate.

3.3.1 Umbrella sampling

Umbrella sampling (US) is a widely used method first introduced by G. M. Torrie and J. P. Valleau in 1977 [86]. The main idea of US is to fix some values $(z_i)_{i \in \llbracket 1, K \rrbracket}$ of the transition coordinate ξ . It can either be applied to the configurational transitions (in which case $(z_i)_{i \in \llbracket 1, K \rrbracket} \in \mathcal{M}^K$) or to alchemical transitions (in which case $(z_i)_{i \in \llbracket 1, K \rrbracket} = (\lambda_i)_{i \in \llbracket 1, K \rrbracket} \in [0, 1]^K$). Doing so, one sets K independent windows in which to run simulations: for each value z_i one obtains an independent trajectory of the system. Now, the idea is to bias the potential energy V of the original Hamiltonian (1.1) with a different bias for each window. In the i -th window, one considers the *biased* potential energy:

$$U_i(q) := V(q) + \tilde{V}_i(q),$$

where \tilde{V}_i is an harmonic potential of constant k :

$$\tilde{V}_i(q) = \frac{k}{2} (\xi(q) - z_i)^2.$$

Our goal is to obtain the unbiased free energy $A \circ \xi$ of the system, defined by equation (3.9): we thus need the unbiased stationary distribution $\pi_i^\xi(z) dz$ of the reaction coordinate for the i -th window. The corresponding biased distribution of the reaction coordinate is given by the following density [46]: for all $z \in \mathcal{M}$,

$$\tilde{\pi}_i^\xi(z) = \frac{\int_{\Sigma_z} e^{-\beta(V(q) + \tilde{V}_i \circ \xi(q))} \delta_{\xi(q) - z}(dq)}{\int_{\mathcal{D}} e^{-\beta(V(q) + \tilde{V}_i \circ \xi(q))} dq} = e^{-\beta \tilde{V}_i(z)} \frac{\int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q) - z}(dq)}{\int_{\mathcal{D}} e^{-\beta(V(q) + \tilde{V}_i \circ \xi(q))} dq}$$

so that

$$\pi_i^\xi(z) = \frac{\int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q) - z}(dq)}{\int_{\mathcal{D}} e^{-\beta V(q)} dq} = \tilde{\pi}_i^\xi(z) e^{\beta \tilde{V}_i(z)} \frac{\int_{\mathcal{D}} e^{-\beta(V(q) + \tilde{V}_i \circ \xi(q))} dq}{\int_{\mathcal{D}} e^{-\beta V(q)} dq}.$$

Hence

$$\pi_i^\xi(z) = \tilde{\pi}_i^\xi(z) e^{\beta \tilde{V}_i(z)} \mathbb{E}_\mu \left[e^{-\beta \tilde{V}_i \circ \xi} \right], \quad (3.12)$$

and the unbiased free energy for the i -th window may be written as follows:

$$A_i(z) = -\beta^{-1} \ln \left(\tilde{\pi}_i^\xi(z) \right) - \tilde{V}_i(z) - \beta^{-1} \ln \left(\mathbb{E}_\mu \left[e^{-\beta \tilde{V}_i \circ \xi} \right] \right).$$

In order to obtain the free energy on the whole reaction coordinate space, one needs to compute the energy $F_i := \beta^{-1} \ln \left(\mathbb{E}_\mu \left[e^{-\beta \tilde{V}_i \circ \xi} \right] \right)$ for each $i \in \llbracket 1, K \rrbracket$. One has:

$$e^{-\beta F_i} = \mathbb{E}_\mu \left[e^{-\beta \tilde{V}_i \circ \xi} \right] = \int_{\mathcal{D}} e^{-\beta \tilde{V}_i \circ \xi(q)} \mu(dq),$$

where μ^ξ is the global, unbiased distribution given by (3.8). The quantity F_i is the free energy associated to the introduction of the i -th window potential \tilde{V}_i , and cannot be determined by direct sampling. There exists several methods such as the *Weighted Histogram Analysis method*, presented in Section 3.4.2 or *Umbrella Integration* are used to tackle this problem. This high-

lights the fact that the windows need to be built such that the distributions $(\pi_i^\xi)_{i \in [1, K]}$ should sufficiently overlap with each other [56, Section 2.4.1.4].

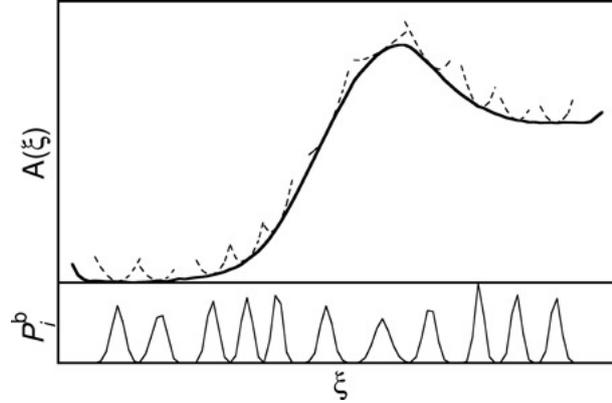


Figure 3.7 – Scheme of the umbrella sampling method: the free energy profile $A \circ \xi$ (thick solid curve) is represented. The contributions of the free energies $(A_i)_{i \in [1, K]}$ are represented by the dashed curves. The notation P_i^b stands for the biased distribution for the i -th window, in other words, $P_i^b \equiv \pi_i^\xi$. *Source:*[46]

Note that there is an inherent issue with the Umbrella Sampling method: it cannot *a priori* avoid metastability. Indeed, the choice of the values $(z_i)_{i \in [1, K]}$ of the transition coordinate ξ is done ahead of the simulation, without prior knowledge of the energy landscape, as shown in Figure 3.8 below.

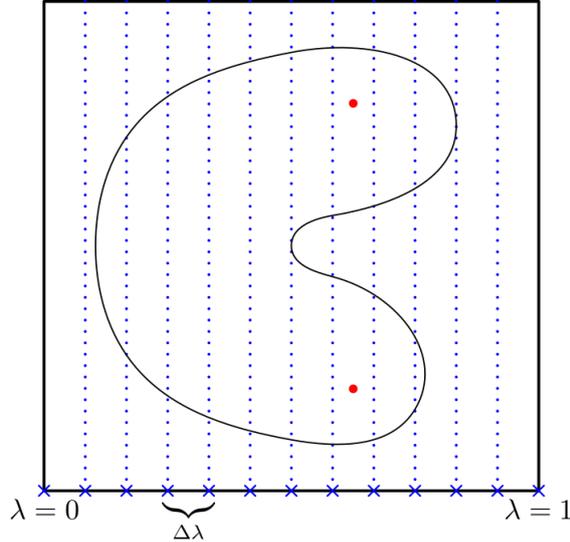


Figure 3.8 – An example of metastability for the Umbrella Sampling method. A level set of an energy is shown here, along with a discretisation of the interval $[0, 1]$ in which the parameter λ evolves. The bin containing the two red dots has two separate configuration regions to sample. As the dynamics is restrained to the bin, it will not be able to escape the region to visit the other, leading to metastability.

3.3.2 Adaptive biasing methods

The key idea behind the umbrella sampling method is that of biasing the potential energy of the system. In the same manner, one may consider another class of enhanced sampling algorithms, which rely on *adaptive biases* to modify the dynamics in a single, non-equilibrium trajectory of the system.

3.3.2.1 Importance sampling and flat histogram property

The adaptive biasing methods to be defined below are *importance sampling techniques*. The goal of importance sampling is to modify the dynamics so that the new stationary measure is easier to sample than the reference target measure (in our case, it is the Boltzmann-Gibbs measure μ given by (1.9)). Let us sketch the main idea of importance sampling, by considering overdamped Langevin dynamics (2.4) with modified potential $V(q) + \tilde{V}(q)$, where V is the potential energy of the original system and \tilde{V} a biasing potential:

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

This modified dynamics is ergodic with respect to the new stationary measure:

$$\tilde{\mu}(dq) = Z_{\tilde{\mu}}^{-1} e^{-\beta(V+\tilde{V})(q)} dq, \quad Z_{\tilde{\mu}} = \int_{\mathcal{D}} e^{-\beta(V+\tilde{V})(q)} dq.$$

Now, consider an observable $\varphi \in \mathcal{C}_0^\infty(\mathcal{D})$. Provided one has simulated a trajectory of the system using the modified potential $V + \tilde{V}$, one can calculate:

$$\int_{\mathcal{D}} \varphi \mu(dq) = \frac{\int_{\mathcal{D}} \varphi e^{-\beta V(q)} dq}{\int_{\mathcal{D}} e^{-\beta V(q)} dq} = \frac{\int_{\mathcal{D}} \varphi e^{\beta \tilde{V}(q)} e^{-\beta(V(q)+\tilde{V}(q))} dq}{\int_{\mathcal{D}} e^{\beta \tilde{V}(q)} e^{-\beta(V(q)+\tilde{V}(q))} dq}$$

i.e

$$\int_{\mathcal{D}} \varphi \mu(dq) = \frac{\int_{\mathcal{D}} \varphi e^{\beta \tilde{V}(q)} \tilde{\mu}(dq)}{\int_{\mathcal{D}} e^{\beta \tilde{V}(q)} \tilde{\mu}(dq)}.$$

Importance sampling highlights the need to build methods which sample *nice* measures. As a first intuition, one might think of using a reaction coordinate ξ as in (3.6) to capture the slow movements of the dynamics, and to bias the dynamics in the direction of ξ . Given a reaction coordinate ξ , the free energy of the system can be rewritten as:

$$A(z) = -\beta^{-1} \ln \left(Z_{\mu}^{-1} \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq) \right) \quad \forall z \in \mathcal{M}. \quad (3.13)$$

Let us now consider the following modified overdamped Langevin dynamics, where a constant bias, equal to the free energy A of the system, is added to the potential energy:

$$dq_t = -\nabla (V - A \circ \xi)(q_t)dt + \sqrt{2\beta^{-1}}dW_t, \quad (3.14)$$

where $(W_t)_{t \geq 0}$ is a standard dN -dimensional Brownian motion. The stationary distribution of the modified dynamics (3.14) is given by the measure

$$\mu_A(dq) = Z_{\mu_A}^{-1} e^{-\beta(V - A \circ \xi)(q)} dq, \quad Z_{\mu_A} = \int_{\mathcal{D}} e^{-\beta(V - A \circ \xi)(q)} dq.$$

One can check that the image of μ_A by ξ is the uniform measure, up to a normalisation constant. Indeed,

$$\begin{aligned} \mu_A^\xi(dz) &:= \xi_* \mu_A(dq) = \left(Z_{\mu_A}^{-1} \int_{\Sigma_z} e^{-\beta(V(q) - A \circ \xi(q))} \delta_{\xi(q) - z}(dq) \right) dz \\ &= \left(Z_{\mu_A}^{-1} e^{\beta A(z)} \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q) - z}(dq) \right) dz \\ &= \left(Z_{\mu_A}^{-1} Z_\mu e^{\beta A(z)} e^{-\beta A(z)} \right) dz, \quad \text{since } A \text{ is given by (3.13)} \\ &= (Z_{\mu_A}^{-1} Z_\mu) dz, \end{aligned}$$

and provided \mathcal{M} is compact, or A is restricted to a compact domain of \mathbb{R}^{dN} , one has that $Z_{\mu_A}^{-1} Z_\mu$ is finite, and that μ_A^ξ is the uniform distribution, namely $\mu_A^\xi := \xi_* \mu_A = \lambda(\mathcal{M})^{-1} \mathbf{1}_{\mathcal{M}}$, with $\lambda(\mathcal{M})$ being the Lebesgue measure on \mathcal{M} . Since, contrary to the reference measure $\xi_* \mu$, the uniform measure is no longer multimodal, we expect a faster sampling of the phase space, provided ξ is well chosen so that μ_A is less multimodal than μ . In other words, the dynamics (3.14) will asymptotically reach a *flat histogram property* in the direction of the reaction coordinate ξ : the exploration of \mathcal{M} by $(\xi(q_t))_{t \geq 0}$ is faster for dynamics (3.14) than for overdamped Langevin dynamics (2.4).

Although this change of potential can accelerate the phase space sampling, the free-energy A is *a priori* unknown. The main idea to avoid this issue will be to approximate on-the-fly either A or its derivative with respect to the reaction coordinate, ∇A . Let us here introduce the *Metadynamics*, *Adaptive Biasing Potential*, *Adaptive Biasing Force* and *Projected Adaptive Biasing Force* methods, which are all designed with the goal of satisfying the flat histogram property in mind. These methods all use an adaptive bias B_t , which depends on the reaction coordinate, to modify either the potential energy or the interaction force of the system. When the bias B_t modifies the potential energy, one considers overdamped Langevin dynamics (2.4) with a drift term equal to $-\nabla(V(q_t) - B_t \circ \xi(q_t))$, whereas if the bias modifies the interaction force, the drift term is equal to $-\nabla V(q_t) + B_t \circ \xi(q_t)$. The bias B_t can be defined using either the law of the process considered, in which case the process is an approximation of an interacting particles system, or the (un-)weighted occupation measure, in which case one works with a *self-interacting* diffusion process. We choose to present the methods as they have been introduced historically; as a consequence, the bias may be defined differently in the three following sections. Nonetheless, let us keep in mind that both definitions of the bias are acceptable for each algorithm.

3.3.2.2 Metadynamics

Metadynamics, or MetaD, is a sampling method introduced by A. Laio and M. Parrinello in 2002 [53]. It relies on the idea of biasing the potential energy V of the system with an *history-dependent* bias B_t , which can be written as the sum of Gaussian potentials added along the reaction coordinate space. The considered dynamics is thus given by:

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

Here, the reaction coordinate ξ has K components, in other words $\xi(q) = (\xi_1(q), \dots, \xi_K(q))$ for all $q \in \mathcal{D}$. Let us insist on notations: while Umbrella Sampling considered K independent copies of the system in order to obtain K trajectories, metadynamics considers only one trajectory, with a decomposition of ξ into K components. The repulsive Gaussian potentials added throughout the simulation to the potential energy are centered on the already explored points of the reaction coordinate space.

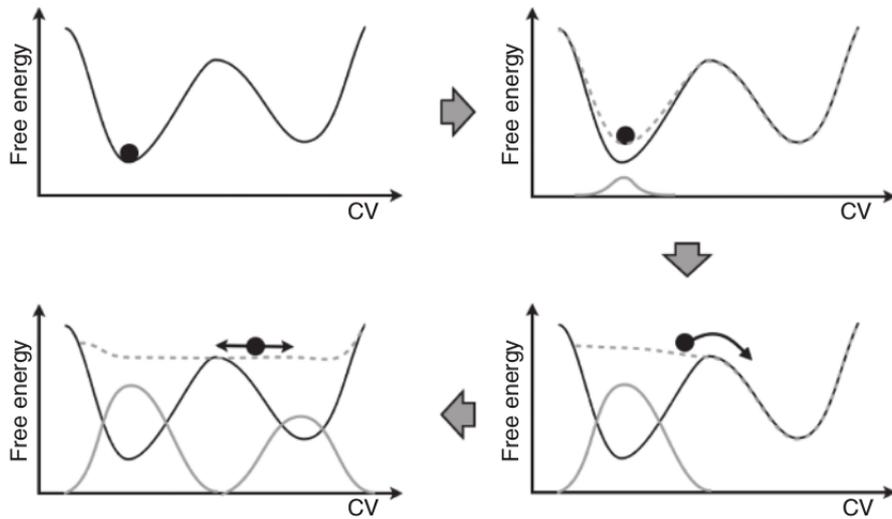


Figure 3.9 – Scheme of the metadynamics method. The left upper corner represents the unbiased system: the energy landscape has two minima, and the system is metastable. The right upper corner represents the first application of a Gaussian potential bias: the energy landscape is transformed unto the gray dashed landscape. The right lower corner represents the moment where, after several deposition times, the system can escape the first metastable region. The left lower corner represents the long-time limit where all metastable regions have been *filled* and the free energy landscape has been flattened in the direction of ξ . The solid gray line yields a rough approximation of the negative of the unbiased free energy $A \circ \xi$. (*Source of diagram:*[18])

Let us denote by τ the deposition time of the repulsive Gaussian potentials. At time t , the biasing potential may be written:

$$B_t \circ \xi(q_t) = \sum_{n=1}^{\lfloor t/\tau \rfloor} \omega_G e^{-\sum_{k=1}^K \frac{(\xi_k(q_t) - z_k(n\tau))^2}{2\sigma_k^2}}, \quad (3.15)$$

where σ_k is the preassigned width of the Gaussian potential associated to the k -th reaction coordinate ξ_k , and $z_k(n\tau) \in \mathcal{M}$ is the selected value of $\xi_k(q_t)$, at time $t = n\tau$.

The goal of metadynamics is to discourage the exploration of states already visited by the dynamics. It also provides an immediate estimation of the free energy surface. Indeed, Gaussian potentials are most likely to be added at local minima of the total free energy surface. In the long-time limit, one would like for the bias potential to converge towards the negative free energy associated to the reaction coordinate ξ [59], namely:

$$\lim_{t \rightarrow +\infty} B_t \circ \xi = -A \circ \xi + C,$$

where C is a constant. There are nonetheless two inherent problems to metadynamics as presented above:

- (i) A common issue with sampling methods using collective variables is that the reaction coordinate has to be chosen beforehand. Metadynamics is no exception.
- (ii) The estimation of the free energy is known not to properly converge in the long-time limit, as the height ω_G in (3.15) is constant. As a consequence, the free energy potential obtained with the biasing potential will tend to oscillate around the true value of the unbiased free energy. Furthermore, the risk of pushing the dynamics towards irrelevant sampling regions is not null.

Problem (i) is inherent to the structure of metadynamics, but problem (ii) is not: in order to tackle the latter, A. Barducci, G. Bussi and M. Parrinello have built in 2008 a variation of metadynamics, the *Well-Tempered Metadynamics* (WTMetaD) method [6]. The idea is to re-write (3.15) as:

$$B_t \circ \xi(q_t) = \sum_{n=1}^{\lfloor t/\tau \rfloor} \omega_G(n\tau) e^{-\sum_{k=1}^K \frac{(\xi_k(q_t) - z_k(n\tau))^2}{2\sigma_k^2}}, \quad (3.16)$$

where

$$\omega_G(n\tau) = \omega_0 e^{-\frac{B_{n\tau}(z(n\tau))}{k_B \Delta T}}$$

is now an history-dependent height. Here, $z(n\tau) = (z_1(n\tau), \dots, z_K(n\tau))$ and ΔT is a tunable parameter. Using (3.16) allows the dynamics, in the long-time limit, to spend more time in regions where smaller Gaussians are used, in other words, around the deepest wells of the energy surface. One indeed has on \mathcal{D} [6] :

$$\lim_{t \rightarrow +\infty} B_t \circ \xi = -\frac{\Delta T}{T + \Delta T} A \circ \xi + C,$$

where C is a constant, and T is the absolute temperature of the system. When $\Delta T = 0$ one recovers classical molecular dynamics, whereas the limit $\Delta T \rightarrow +\infty$ amounts to using the classical MetaD method. We refer to [6] for a proper construction of the WTMetaD algorithm.

3.3.2.3 Adaptive Biasing Potential method

Let us consider the overdamped Langevin dynamics (2.4) with initial condition $q_0 = q \in \mathcal{D}$. Since the process is ergodic, the empirical distribution

$$\bar{\mu}_t := \frac{1}{t} \int_0^t \delta_{q_s} ds$$

converges in distribution almost surely towards μ in the long-time limit [7]. The main idea behind the *Adaptive Biasing Potential method* (ABP) was originally sketched in 2006 by S. Marsili and colleagues [62]. It consists of biasing the potential energy V with an adaptive bias in the direction of the reaction coordinate (3.6). The biased potential energy at time $t > 0$ is:

$$U_t(q) = V(q) - B_t \circ \xi(q), \quad \forall q \in \mathcal{D}.$$

The dynamics of the ABP method is given by:

$$\begin{cases} dq_t &= -\nabla(V - B_t \circ \xi)(q_t)dt + \sqrt{2\beta^{-1}}dW_t \\ \bar{\pi}_t &= \frac{\bar{\pi}_0 + \int_0^t e^{-\beta B_s \circ \xi(q_s)} ds}{1 + \int_0^t e^{-\beta B_s \circ \xi(q_s)} ds} \\ B_t(z) &= -\beta^{-1} \ln \left(\int_{\mathcal{D}} K(z, \xi(q)) \bar{\pi}_t(dq) \right), \quad \forall z \in \mathcal{M} \end{cases} \quad (3.17)$$

where $(W_t)_{t \geq 0}$ is standard dN -dimensional Brownian motion, and $K : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+^*$ is a smooth kernel such that $\int_{\mathcal{M}} K(z, \tilde{z}) dz = 1$ for all $\tilde{z} \in \mathcal{M}$, which allows for B_t to be in $\mathcal{C}^\infty(\mathcal{M})$ for each time $t > 0$, and for $\bar{\pi}_t$ to have a density. We endow the dynamics with arbitrary and deterministic initial conditions $q_0 = q \in \mathcal{D}$, $\bar{\pi}_0 \in \mathcal{P}(\mathcal{D})$ (the set of Borel probability distributions on \mathcal{D} , endowed with the usual topology of weak convergence of probability distributions) and $B_0 \in \mathcal{C}^\infty(\mathcal{M})$. The unknowns of the dynamics (3.17) are the positions $(q_t)_{t \geq 0}$, the weighted empirical distribution $\bar{\pi}_t$ at each time t , and the adaptive bias $B_t \circ \xi$ at each time t .

The question of the convergence of the method has first been treated by M. Benaïm and C.-E. Bréhier in 2016 [7] with the following toy model: the configuration space \mathcal{D} is given by $\mathcal{D} = \mathbb{T}^m \times \mathbb{T}^{n-m}$, i.e each configuration q can be decomposed as $q = (x, y)$, and the reaction coordinate is such that $\xi(x, y) = x \in \mathbb{T}^m$. In which case, one can show that:

(i) Almost surely, the empirical distribution converges in the long-time limit towards the Boltzmann-Gibbs measure in $\mathcal{P}(\mathbb{T}^n)$ [7, Theorem 1.1].

(ii) Consider the stationary bias $B_\infty \circ \xi$ defined as follows:

$$B_\infty(z) = -\beta^{-1} \ln \left(\int_{\mathbb{T}^n} K(z, \xi(q)) \mu(dq) \right), \quad \forall z \in \mathbb{T}^m.$$

Almost surely, the adaptive bias $B_t \circ \xi$ converges in the long-time limit towards the stationary bias $B_\infty \circ \xi$ in $\mathcal{C}^k(\mathbb{T}^m)$ for all $k \in \mathbb{N}$ [7, Corollary 1.2].

(iii) The stationary bias B_∞ is an approximation of the free energy A of the unbiased system

as defined in (3.9), and one may notice that

$$e^{-\beta B_\infty(z)} = \int_{\mathbb{T}^m} K(z, \tilde{z}) e^{-\beta A(\tilde{z})} d\tilde{z}, \quad \forall z \in \mathbb{T}^m,$$

so that, provided the kernel K is built suitably, the bias $B_t \circ \xi$ will almost surely converge in the long-time limit to the unbiased free energy A .

- (iv) The flat histogram property is satisfied, namely the image by ξ of the distribution $\pi_t := \frac{1}{t} \int_0^t \delta_{(X_s, Y_s)} ds$ converges almost surely to the uniform distribution in $\mathcal{P}(\mathbb{T}^m)$.

Of course one may consider a more generic setting, with a reaction coordinate ξ given as in (3.6), and a more generic configuration space \mathcal{D} . We refer to the extensive work of M. Benaïm and C.-E. Bréhier [8] where the well-posedness of the dynamics and long-time convergence of the method is studied thoroughly and for different configuration spaces, diffusion processes and reaction coordinates.

Remark 12. *As stated beforehand, the bias B_t can be defined either using the reweighted occupation measure $\bar{\pi}_t$ defined in (3.17) (or the non-weighted measure $\bar{\pi}_0 + \int_0^t e^{-\beta B_s \circ \xi(q_s)} ds$ for that matter), or the law π_t of the process. In the latter case, the ABP method is given as follows [56, Section 5.1.1]:*

$$\begin{cases} dq_t &= -\nabla(V - B_t \circ \xi)(q_t) dt + \sqrt{2\beta^{-1}} dW_t \\ \frac{dB_t(z)}{dt} &= -\beta^{-1} \ln(\pi_t^\xi(z)), \quad \forall z \in \mathcal{M}, \end{cases}$$

where π_t^ξ is the image of the law π_t by the reaction coordinate ξ (with a slight abuse of notations where probability density functions and measures are not distinguished). Namely,

$$\pi_t^\xi(z) = \int_{\Sigma_z} \pi_t(q) \delta_{\xi(q)-z} dq \quad \forall z \in \mathcal{M},$$

the measure $\delta_{\xi(q)-z}(dq)$ being given by equation (3.2.1).

3.3.2.4 Adaptive Biasing Force method

Contrary to the ABP method, *Adaptive Biasing Force method* [24, 39] uses an adaptive bias that is built so that it approaches in the long-time limit the *gradient* of the free energy ∇A instead of looking directly for the free energy A . In order to do so, one does not bias the potential energy V , but the *interaction force* $\mathcal{F} = -\nabla V$. This is motivated by the fact that the free energy A can be rewritten as (3.10):

$$\nabla A(z) = \int_{\Sigma_z} F(q) d\mu^\xi(dq|z),$$

where F is the local mean force defined in Section 3.2.1. The ABF algorithm is given by:

$$\begin{cases} dq_t &= (-\nabla V(q_t) + B_t \circ \xi(q_t)) \nabla \xi(q_t) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t(z) &= \mathbb{E}[F(q_t) | \xi(q_t) = z] \quad \forall z \in \mathcal{M}, \end{cases} \quad (3.18)$$

Namely, the bias B_t is given by

$$B_t(z) = \mathbb{E}[F(q_t) | \xi(q_t) = z] = \int_{\Sigma_z} F(q) \pi_t^\xi(dq|z) \quad \forall z \in \mathcal{M},$$

where $\pi_t^\xi(\cdot|z)$ is the conditional measure at a fixed value $\xi(q) = z$ of the reaction coordinate, obtained from the instantaneous law of the process π_t .

Now, one can check that μ_A is a fixed point of the Fokker-Planck equation associated to the process. In other words, if $q_0 \sim \mu_A$, then $q_t \sim \mu_A$ for all $t \geq 0$ and $(q_t)_{t \geq 0}$ is exactly the diffusion (3.14).

Starting from another initial distribution, using entropy estimates and functional inequalities as defined in Section 3.1.2.2, it has been proven in [57], under mild assumptions, that this fixed point is in fact an attractor of the dynamics, in the sense that B_t converges to ∇A in the long-time limit, and the law of q_t converges to μ_A . Furthermore, the flat histogram property is satisfied.

Of course, the ABF method can be defined with a self-interacting process, as done in C.-E. Bréhier, M. Benaïm and P. Monmarché's work [9]. Choosing to introduce the ABF method with a process which approximates a system of N interacting particles (i.e to define the bias with the instantaneous law of the process as in (3.18)) is a purely *arbitrary choice*, as the theoretical study of the algorithm proved to be easier in this case. We refer to [57] for a detailed construction of the former ABF method and for the long-time convergence proofs, along with the very complete review [58, Part 4, page 777] on adaptive importance sampling methods. We also refer to [58, Remark 4.4] and [56, Section 5.1.1.5] for discussions on which method to use between the ABP or ABF methods.

Remark 13.

▷ In some cases \mathcal{M} is not bounded, for example when ξ is a distance. If so, an additional confining potential $W \circ \xi$ is needed in the drift [57].

▷ As discussed in [57], the algorithm (3.18) can be modified in order to obtain a diffusive behaviour for the law of $\xi(X_t)$. Additional terms depending on ξ are added to obtain the following variant:

$$\begin{cases} dX_t &= (-\nabla V + B_t \circ \xi - \nabla W \circ \xi + \beta^{-1} \nabla \ln(|\nabla \xi|^{-2})) |\nabla \xi|^{-2}(X_t) dt + \sqrt{2\beta^{-1}} |\nabla \xi|^{-1}(X_t) dW_t \\ B_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z], \quad \forall z \in \mathcal{M}. \end{cases}$$

In this case the long-time convergence of B_t towards ∇A is stronger than in the case of (3.18), in that it requires fewer hypotheses.

3.3.2.5 Projected Adaptive Biasing Force method

We might also consider a variant of the ABF method, namely the *Projected Adaptive Biasing Force* (PABF) algorithm, introduced by H. Alrachid and T. Lelièvre in 2015 [2]:

$$\begin{cases} dX_t &= (-\nabla V(X_t) + B_t(\xi(X_t)) \nabla \xi(X_t)) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t &= P_{L^2(\lambda)}(G_t) \\ G_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z] \quad \forall z \in \mathcal{M}, \end{cases} \quad (3.19)$$

where $P_{L^2(\lambda)}(f)$ stands for the Helmholtz projection with respect to the Lebesgue measure λ of a vector field f on an open bounded set $\mathcal{M} \subset \mathbb{R}^{dN}$ with Lipschitz boundary $\partial\mathcal{M}$ [3]. In other words, it is the gradient of the minimizer on $\{g \in H^1(\mathcal{M}), \int_{\mathcal{M}} g dx = 0\}$ of

$$g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 dx.$$

More generally, if ν is a continuous positive measure on \mathcal{M} , the Helmholtz projection with respect to ν is the minimizer on $\{g \in H^1(\mathcal{M}), \int_{\mathcal{M}} g dx = 0\}$ of $g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 \nu(dx)$.

The long-time convergence of bias B_t and of the dynamics' (3.19) law π_t has been shown in [2, Theorem 1], but for a version of the algorithm where the classical Helmholtz projection in $L^2(\lambda)$ is replaced by the Helmholtz projection in the weighted space $L^2(\pi_t^\xi)$. This choice of projection is motivated in [2] by some cancellations in the computations of the proofs. Nevertheless, as already noted in [2], the classical Helmholtz projection is used in practice. Showing the long-time convergence of the PABF algorithm with classical Helmholtz projection (3.19) is one of the problems solved in the scope of this thesis.

3.3.2.6 The non-conservative case

For both the ABF and PABF method, the interaction force $\mathcal{F} = -\nabla V$ is the gradient of a potential energy. It is said to be *conservative*: if the system were completely Hamiltonian, the system's mechanical energy (i.e the sum of the potential and kinetic energies) would be conserved throughout time. Most empirical force fields are built to be conservative. Nonetheless, as stated in Section 2.1.2, one would wish to use force fields closer to reality. For this reasons, many MD simulations resort to *ab initio* force fields. It has been shown [74, 69, 20] that some *ab initio* approximations could lead to *hysteresis*, namely, a violation of the conservation of energy, making the interaction force *a priori* non conservative. In this case, one is interested in knowing if, by controlling the error made on the force $-\nabla V$, one can deduce an estimation of the error made on the system's free energy. The robustness of a diffusion's invariant measure with respect to the perturbation of its drift is a classical problem, but note that in the ABF case, the adaptive procedure makes the question more subtle. Moreover, the convergence of the ABF method in such a context cannot be deduced from the aforementioned convergence analysis. Establishing the well-posedness of both the ABF and PABF algorithms along with their long-time convergence in the scope of non-conservative forces is also one of the problems that will be tackled in this thesis.

3.4 Free energy differences in the alchemical setting

We have shown that metastability could be avoided using various enhanced sampling methods such as the Adaptive Biasing Force method. A common goal of the methods introduced above is to compute the free energy of the system associated to a given transition coordinate, and more specifically free energy differences, as presented in Section 3.2. The methods presented in Section 3.3 have all been introduced using a reaction coordinate ξ , in other words, in the scope of configurational sampling. Nevertheless, the motivations for calculating free energy differences appear clearer when considering alchemical transitions. Let us give two examples of why free energy differences are key quantities to target.

3.4.1 Motivation

3.4.1.1 Ligand binding affinity

An important problem in biochemistry is to determine the binding affinity of a complex called *ligand* with a target complex called *receptor*. The receptor can either be a protein or a nucleic acid. The nature of this binding is usually functional: it can for example allow the transmission of a signal, the triggering of catalysis, or even modulate enzymatic activity. A ligand which provokes a physiological reaction when binding with a receptor is an *agonist* ligand: morphine, which binds with the μ -opioid receptor OP3 is an example of agonist ligand. On the contrary, a ligand which blocks the activity of a receptor after binding with it is an *antagonist* ligand: beta-blockers are a common example of antagonist ligands. The binding is usually reversible, and is a result of non-covalent interactions between the ligand and receptor, like van der Waals interactions, ionic interactions or hydrogen bonding. The force resulting from the interaction between the two entities is defined by the *dissociation constant*, often called *affinity (constant)*. The greater the interaction forces are between the ligand and the receptor, the greater the *affinity* is. A strong affinity can be explained by a strong occupation of the receptor by the ligand. A ligand with strong affinity with a given receptor can impact the latter in a non-negligible manner, especially when the energy resulting from the binding is sufficiently large to cause a conformational transition of the receptor. Such configurational change can impact the habitual operation of an ionic channel or of an enzyme linked to the receptor. Evaluating the affinity of a ligand with a receptor is consequently of great importance, especially in *drug design*. Note that in the scope of pharmacology, the *potency* of a given drug, namely the required concentration of the drug to induce a physiological response of given intensity, is not directly linked to the drug's affinity.

There exist different manners to determine affinity. The first is to consider the *dissociation constant* K_d (or $pK_d = -\log(K_d)$), which corresponds to the reaction constant of the dissociation between the ligand and receptor. The second is to establish the necessary concentration in order for the ligand to bind with 50% of the receptor: it is given by the *inhibition constant* K_i . Indirectly, affinity can be evaluated by *putting ligands in competition*, where one evaluates the *half maximal inhibitory concentration* IC_{50} , namely, the required ligand concentration to replace 50% of the reference ligand's concentration. The concentrations K_d and IC_{50} are linked via the Cheng-Prusov equation, whose expression depends on the kind of receptor and ligands considered. The binding of an agonist ligand with its receptor can be evaluated with two quantities: its *efficacy*, i.e the intensity of the induced physiological response to the binding, or its *half maximal effective concentration* EC_{50} , i.e the required concentration to trigger a physiological response whose intensity is half of the maximal response. All of these quantities can be determined by computing the (Gibbs) free energy difference ΔG between the initial state where the ligand-receptor complex is dissociated and the final state where the ligand is bound to the receptor. Most of the time ΔG is used to determine the dissociation constant K_d , from which one deduces the value of the K_d constant and indirectly, the value of the IC_{50} constant.

Ligand-competition is a key experiment to lead numerically in the case of alchemical transitions, as it determines which ligand, among a prechosen set of ligands, is the most likely to bind to a given receptor. Figure 3.10 is an example of ligand-competition between two ligands (L8 and L9) for a given receptor (FKBP-12). In Section 6.3.2 we will introduce how the binding affinity of a ligand-receptor complex can be evaluated.

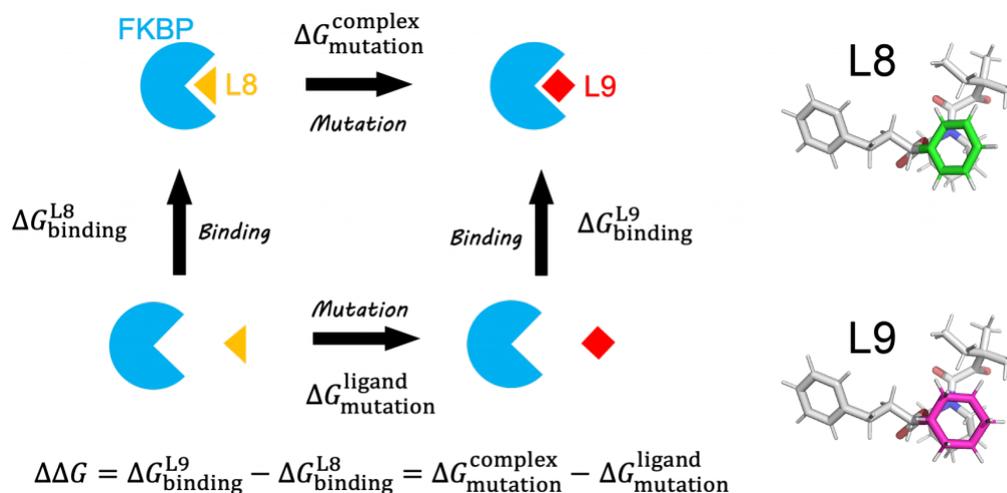


Figure 3.10 – FK506 binding-proteins, or FKBP, are immunophilins interacting with the immunosuppressive drugs FK506 and rapamycin. These receptor proteins are present in all eukaryotes and can induce diverse cellular functions [48]. On the left is the thermodynamic cycle of ligand binding between the FKBP-12 and ligands L8 and L9. On the right are the L8 and L9 ligands representations. *Source: GENESIS Tutorial 15.1, <https://www.r-ccs.riken.jp/labs/cbrt/>*

3.4.1.2 Solvation free energy

Solvation is the process of dissolution of a given compound, called the *solute*, into a liquid, called the *solvent*, the resulting mixture being called a *solution*. In the case where the solvent is water, solvation is often referred to as *hydration*. Solutes can either be ions, neutral molecules, or even proteins. Understanding the behaviour of molecules or proteins in solvents is fundamental in industrial chemistry and biochemistry. As an example, protein folding is greatly influenced by the protein's hydration in water: the hydrophobic effect, which is the tendency of nonpolar moieties to aggregate and exclude water molecules, is a driving force in protein folding. Indeed, some of the protein's sequences may combine hydrophobic (i.e., non-polar) and hydrophilic (i.e., polar) regions. A direct consequence is the hydrophobic effect, as hydrophobic regions will tend to aggregate and repel water molecules, strongly influencing the protein's folding [19].

Just as in the ligand binding processes, a solvation process will lead solvent and solute molecules to be reorganised into solvation complexes. Interactions between solute and solvent include bond formation, hydrogen bonding, van der Waals forces and electrostatic forces. Depending on the strength of the overall interaction force, one can deduce many properties of the solute, such as solubility, reactivity or even colour. As a consequence, the free energy difference between the initial state where the solute is not dissolved in the solvent, and the final state where the solute has been maximally dissolved, is a key quantity to study when it comes to solvation processes. In Section 6.4.5, we will conduct a computation of an ion's solvation free energy.

3.4.2 Available methods

As shown above, the study of a chemical system and of its thermodynamical properties leads to computing free energy differences. Free energy differences allow us to consider many problems, such as the formation of cavities in water, the solvation properties of molecules and ions, the

relative reaction rate between two reactants, or even the structure and stability of a ligand-receptor complex. Let us restrict ourselves to the alchemical case, where the system evolves from an initial state 0 to a final state 1, and list the different tools available to compute the free energy difference $\Delta_{(0 \rightarrow 1)}A$.

3.4.2.1 Free Energy Perturbation

The *Free Energy Perturbation* (FEP) is a method first introduced by R. W. Zwanzig in 1954 [90]. It is one of the most classical methods used to compute free energy differences in the alchemical setting. The FEP method yields:

$$\Delta_{(0 \rightarrow 1)}A = A_1 - A_0 = -\beta^{-1} \ln \left\langle e^{-\beta(H_1 - H_0)} \right\rangle_0$$

where A_0 and H_0 (resp. A_1 and H_1) is the free energy and Hamiltonian of the system at the initial state 0 (resp. final state 1). The notation $\langle \cdot \rangle_0$ denotes the ensemble average with respect to the measure $\mu_0(dqdp) \propto \exp(-\beta H_0(q, p))dqdp$.

Indeed, if one considers the partition functions of the initial state,

$$Z_0 = \int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H_0(q, p)} dqdp,$$

and of the final state,

$$Z_1 = \int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H_1(q, p)} dqdp,$$

since

$$\Delta_{(0 \rightarrow 1)}A = A_1 - A_0 = -\beta^{-1} \ln(Z_1) + \beta^{-1} \ln(Z_0) = -\beta^{-1} \ln \frac{Z_1}{Z_0}, \quad (3.20)$$

one has:

$$\begin{aligned} \Delta_{(0 \rightarrow 1)}A &= \beta^{-1} \ln \left(\frac{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H_1(q, p)} dqdp}{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H_0(q, p)} dqdp} \right) \\ &= \beta^{-1} \ln \left(\frac{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta(H_1 - H_0)(q, p)} e^{-\beta H_0(q, p)} dqdp}{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H_0} dqdp} \right) \\ &= -\beta^{-1} \ln \left\langle e^{-\beta(H_1 - H_0)} \right\rangle_0. \end{aligned}$$

3.4.2.2 Thermodynamic Integration

Another way to approach free energy differences is via *thermodynamic integration* (TI). Thermodynamic integration uses the fact that the system considered is extended: along with the microstate $(q, p) \in \mathcal{T}^*\mathcal{D}$, we consider a coupling parameter $\lambda \in [0, 1]$ so that the system's total energy is given by an extended Hamiltonian $H(q, p; \lambda)$. The partition function then becomes:

$$Z_\lambda = \int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H(q, p; \lambda)} dqdp.$$

And the Helmholtz free energy derivative with respect to λ reads:

$$\begin{aligned} \frac{\partial A}{\partial \lambda} &= -\beta^{-1} \frac{\partial}{\partial \lambda} \ln Z_\lambda = -\beta^{-1} \frac{1}{Z_\lambda} \frac{\partial Z_\lambda}{\partial \lambda} \\ &= \int_{\mathcal{D} \times \mathbb{R}^{dN}} \frac{\partial H(q, p; \lambda)}{\partial \lambda} \frac{e^{-\beta H(q, p; \lambda)}}{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H(q, p; \lambda)} dq dp} dq dp \\ &= \left\langle \frac{\partial H(q, p; \lambda)}{\partial \lambda} \right\rangle_\lambda \end{aligned}$$

where $\langle \cdot \rangle_\lambda$ denotes the ensemble average with respect to the measure $\mu_{\beta, \lambda}(dq dp) \propto \exp(-\beta H(q, p; \lambda)) dq dp$. Then one has:

$$\Delta_{(0 \rightarrow 1)} A = A_1 - A_0 = \int_0^1 \left\langle \frac{\partial H(q, p; \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda.$$

3.4.2.3 Weighted Histogram Analysis method

The *Weighted Histogram Analysis method* (WHAM) is a method commonly used in order to deduce free energy differences in post-processing of an Umbrella Sampling simulation. It was first introduced by S. Kumar and his colleagues in 1992 [51]. Let us keep the notation of Section 3.3.1: as said in the aforementioned section, the constant F_i is the free energy associated to the introduction of a bias in the i -th window. Provided one can estimate the quantities $(F_i)_{i \in \llbracket 1, K \rrbracket}$, one will be able to determine the total free energy difference $\Delta_{(0 \rightarrow 1)} A$ between the initial and final state of the system, by summing the free energy differences associated to the windows:

$$\Delta_{(0 \rightarrow 1)} A = \sum_{i=1}^{K-1} (F_{i+1} - F_i).$$

The WHAM method provides a way to this. Its first step is to build a density estimator $\hat{\mu}^\xi(z)$ of the *marginal distribution* μ^ξ of the Boltzmann-Gibbs measure with respect to ξ , namely

$$\mu^\xi(dz) := \xi \star \mu(dq) = \left(Z_\mu^{-1} \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq) \right) dz.$$

The estimator is given by:

$$\hat{\mu}^\xi(z) = \sum_{i=1}^K \pi_i^\xi(z) \times \frac{N_i e^{-\beta(\tilde{V}_i(z) - F_i)}}{\sum_{j=1}^K N_j e^{-\beta(\tilde{V}_j(z) - F_j)}},$$

where N_k represents the number of data collected in the k -th window to sample the biased distribution density π_k^ξ .

From equation (3.12), the unbiased distribution π_i^ξ of the i -th windows is given by:

$$\pi_i^\xi(z) = \tilde{\pi}_i^\xi(z) e^{\beta \tilde{V}_i(z)} e^{-\beta F_i},$$

so that one has:

$$\hat{\mu}^\xi(\mathrm{d}z) = \sum_{i=1}^K \frac{N_i \tilde{\pi}_i^\xi(z)}{\sum_{j=1}^K N_j e^{-\beta(\tilde{V}_j(z) - F_j)}}. \quad (3.21)$$

And the free energies $(F_i)_{i \in \llbracket 1, K \rrbracket}$ are then computed using the optimal estimation [84]:

$$F_j = \mathbb{E}_\mu \left[e^{-\beta \tilde{V}_j \circ \xi} \right] = \int_{\mathcal{M}} e^{-\beta \tilde{V}_j(z)} \mu^\xi(z) \mathrm{d}z. \quad (3.22)$$

Equations (3.21) and (3.22) are solved self-consistently, which is usually done via an iteration procedure: we refer to [84] for further details on the WHAM method.

3.4.2.4 Bennett Acceptance Ratio method

The *Bennett Acceptance Ratio method* is a method introduced by C. H. Bennett in 1976 [10]. It is used to compute free energy differences relying on the ratio of partition functions as in equation (3.20):

$$\Delta_{(0 \rightarrow 1)} A = -\beta^{-1} \ln \frac{Z_1}{Z_0}.$$

Indeed, computing the direct partition function defining the free energy A as in (3.9) is usually impossible, whereas the ratio $Q_{(0 \rightarrow 1)} := Z_1/Z_0$ happens to be much simpler to determine. First, let us notice that for a given weighting function W of the positions, one has:

$$\frac{Z_1}{Z_0} = \frac{Z_1 \int_{\mathcal{D}} W(q) e^{-\beta(V_1 + V_0)(q)} \mathrm{d}q}{Z_0 \int_{\mathcal{D}} W(q) e^{-\beta(V_1 + V_0)(q)} \mathrm{d}q} = \frac{\int_{\mathcal{D}} e^{-\beta V_1(q)} \mathrm{d}q}{\int_{\mathcal{D}} W(q) e^{-\beta V_0(q)} e^{-\beta V_1(q)} \mathrm{d}q} \cdot \frac{\int_{\mathcal{D}} W(q) e^{-\beta V_1(q)} e^{-\beta V_0(q)} \mathrm{d}q}{\int_{\mathcal{D}} e^{-\beta V_0(q)} \mathrm{d}q}$$

i.e

$$Q_{(0 \rightarrow 1)} = \frac{\langle W e^{-\beta V_1} \rangle_0}{\langle W e^{-\beta V_0} \rangle_1},$$

where the notation $\langle \cdot \rangle_0$ (resp. $\langle \cdot \rangle_1$) denotes the ensemble average with respect to the measure $\mu_0(\mathrm{d}q) \propto \exp(-\beta V_0(q)) \mathrm{d}q$ (resp. $\mu_1(\mathrm{d}q) \propto \exp(-\beta V_1(q)) \mathrm{d}q$). Note that this would require to sample both ensembles μ_0 and μ_1 , i.e to run two distinct simulations. It is nonetheless possible to compute $Q_{(0 \rightarrow 1)}$ with a single simulation. Indeed, by choosing W as

$$W = \exp(\beta \min(V_0, V_1)),$$

one gets that:

$$Q_{(0 \rightarrow 1)} = \frac{\langle M(\beta(V_1 - V_0)) \rangle_0}{\langle M(\beta(V_0 - V_1)) \rangle_1}, \quad (3.23)$$

where $M(x) = \min(1, \exp(-x))$ is the Metropolis function. Now, by running a single Monte-Carlo simulation with an extra trial move which switches the potentials every fixed number of steps, one gets a sampling in a mixed ensemble, and there is no need to samples both ensembles μ_0 and μ_1 anymore.

However, the use of equation (3.23) as an estimator of the ratio $Q_{(0 \rightarrow 1)}$ has several limitations. In fact, both acceptance probabilities $\langle M(\beta(V_1 - V_0)) \rangle_0$ and $\langle M(\beta(V_0 - V_1)) \rangle_1$ should be large enough to be determined with proper statistical accuracy in a single, reasonably short Monte-Carlo simulation. If one acceptance probability is significantly smaller than the other, one can

increase it by shifting the origin of one of the two potentials with an arbitrary constant. If both are too small, this means that there is not enough overlap between the two potentials V_0 and V_1 , and one might consider intermediate potentials and intermediate ratios to compute $Q_{(0 \rightarrow 1)}$. These problems led to the optimisation of a ratio estimator, where the origin of one of the potentials is shifted by a constant $C > 0$, and the weight function is given by the Fermi-Dirac function

$$f(x) = \frac{1}{1 + e^{\beta x}}, \quad \forall x \in \mathbb{R}.$$

The optimised ratio is given as follows:

$$Q_{(0 \rightarrow 1)} = \frac{\langle f(V_1 - V_0 - C) \rangle_0}{\langle f(V_0 - V_1 + C) \rangle_1} e^{-C},$$

and we refer to C. H. Bennett's work [10, Part IIb] for its proper derivation.

3.4.3 λ -dynamics and Orthogonal Space Random Walk sampling methods

3.4.3.1 λ -dynamics

The usual methods used to compute free energy differences sometimes can be burdensome, especially when the system is metastable. In 1996, X. Kong and C. L. Brooks III pointed out that the FEP method could yield different results depending on the initial condition considered [49]. As for US, in some cases like the study of the transition of the α -helix into the 3_{10} -helix, the calculations strongly differ from the experimental data [82]. To tackle this problem in the scope of alchemical transitions, X. Kong, C. L. Brooks and colleagues [49, 34, 47] designed λ -dynamics, whose key idea is to treat the transition coordinate λ (which is initially a simple scalar coordinate evolving continuously between 0 and 1) as a fictitious particle of fictitious mass $\mathbf{m} > 0$. As a consequence, one can now consider the extended microstate $(q, p; \lambda, A) \in \mathcal{T}^* \mathcal{D} \times [0, 1] \times \mathbb{R}$, where $A = \mathbf{m} \dot{\lambda}$ is the fictitious particle momenta and $\dot{\lambda}$ is the time-derivative. One then works with the following extended Hamiltonian:

$$H_{ext}(q, p; \lambda) = E_k^x(p) + E_k^\lambda(A) + V_{ext}(q; \lambda), \quad (3.24)$$

where E_k^x is the kinetic energy which depends solely on the atomic coordinates, $E_k^\lambda(A) = \frac{1}{2} \mathbf{m}^{-1} A^2$ is the kinetic energy of the coordinate λ and $V_{ext}(q; \lambda)$ is the potential energy of the system. Historically, it was given as follows:

$$V_{ext}(q; \lambda) = (1 - f(\lambda))V_0(q) + f(\lambda)V_1(q) \quad (3.25)$$

where V_0 (resp. V_1) is the potential energy of the initial (resp. final), unextended state, and f is a function in $\mathcal{C}^1([0, 1], \mathbb{R})$. The simplest choice for f is $f(\lambda) = \lambda$ for all $\lambda \in [0, 1]$, in which case the free energy difference between state $\lambda = 0$ and $\lambda = 1$ is then given by the FEP method:

$$\Delta_{(0 \rightarrow 1)} A = A_1 - A_0 = -\beta^{-1} \ln \left\langle e^{-\beta \lambda (V_1 - V_0)} \right\rangle_0.$$

Remark 14. *One notices that if the fluctuation of the potential energy $\Delta_{(0 \rightarrow 1)} V = V_1 - V_0$ is large, then basic statistical estimators of $\Delta_{(0 \rightarrow 1)} A$ will suffer from high variance, leading to slow convergence.*

Let us now consider a more general potential energy $V_{ext}(q; \lambda)$ for the extended Hamiltonian.

Since both equilibrium states $\lambda = 0$ and $\lambda = 1$ can be connected by different thermodynamic paths [49], one may consider a fixed discretisation of the path, given by the set of coordinates $\boldsymbol{\lambda} = (\lambda_i)_{i \in \llbracket 1, n \rrbracket} \in [0, 1]$ such that:

$$\begin{cases} \lambda_1 = 0 & \text{et} & \lambda_n = 1 \\ \lambda_i \in (0, 1), & \forall i \in \llbracket 2, n-1 \rrbracket. \end{cases}$$

Note that the choice of the discretisation $\boldsymbol{\lambda} = (\lambda_i)_{i \in \llbracket 1, n \rrbracket} \in [0, 1]$ has a strong impact on the speed of the configuration space sampling. Numerically, one launches a simulation and counts how many times the system visits the cells $[\lambda_i, \lambda_{i+1}[$ in order to evaluate ensemble averages and compute the free energy values $(A(\lambda_i))_{i \in \llbracket 1, n \rrbracket}$. Eventually, the total free energy difference is obtained from:

$$\Delta_{0 \rightarrow 1} A = \sum_{i=0}^{n-1} A(\lambda_{i+1}) - A(\lambda_i) = -\beta^{-1} \sum_{i=0}^{n-1} \ln \frac{\int_{\mathcal{D}} e^{-\beta H_{ext}(q, p; \lambda_{i+1})} dq dp}{\int_{\mathcal{D}} e^{-\beta H_{ext}(q, p; \lambda_i)} dq dp}.$$

In other words, the Hamiltonian being separable:

$$\Delta_{0 \rightarrow 1} A = \sum_{i=0}^{n-1} A(\lambda_{i+1}) - A(\lambda_i) = -\beta^{-1} \sum_{i=0}^{n-1} \ln \frac{\int_{\mathcal{D}} e^{-\beta V_{ext}(q; \lambda_{i+1})} dq}{\int_{\mathcal{D}} e^{-\beta V_{ext}(q; \lambda_i)} dq}.$$

By allowing the collective variable to evolve dynamically, the λ -dynamics method enlarges the size of the sampled space, and in the manner of US, one may also bias the dynamics in order to enhance the configuration space sampling. The multidimensional approach of λ -dynamics is an asset in relatively complex problems: it has been used for the efficient determination of ligand binding affinity by putting ligands in competition for a given receptor [34, 49]. In order to remain concise, we will detail the inherent features of λ -dynamics and its applications in Section 6.3. Now, let us highlight several issues that will be answered in the scope of this thesis.

Problem 1 – Note that the TI method would yield a free energy difference given by

$$\Delta_{0 \rightarrow 1} A = -\beta^{-1} \ln \frac{\int_{\mathcal{D}} e^{-\beta V_{ext}(q; 1)} dq}{\int_{\mathcal{D}} e^{-\beta V_{ext}(q; 0)} dq} = \int_0^1 \langle \partial_\lambda V(\cdot; \lambda) \rangle_\lambda d\lambda, \quad (3.26)$$

where $\langle \cdot \rangle_\lambda$ denotes the ensemble average with respect to the measure $\mu_\lambda(dq) \propto \exp(-\beta V_{ext}(q; \lambda)) dq$. The expression (3.26) motivates the need to obtain a sufficiently smooth profile of the potential of mean force (PMF) $\nabla A(\lambda)$, built with the different values of the PMF $(\nabla A(\lambda_i))_{i \in \llbracket 1, n \rrbracket} = (\langle \partial_\lambda V_{ext}(\cdot; \lambda_i) \rangle_\lambda)_{i \in \llbracket 1, n \rrbracket}$ [83]. However in some cases, the system numerically explodes: adding or subtracting a particle in the system can generate a singularity, and the force $\nabla A(\lambda)$ can explode when the collective variable λ reaches the end states 0 or 1.

A typical example is the solvation of an ion in a box of solvent, when one starts from the initial state “A: the ion is located at the middle of the box and does not interact with any of the solvent molecules”, to gradually reach the final state “B: the ion is fully interacting with the

neighbouring solvent molecules". In this specific case, the extended potential energy $V_{ext}(q; \lambda)$ given by equation (3.25) is naive. As a matter of fact, the potential energy (3.25) does not make the difference between the short-range and long-range interactions: the van der Waals and electrostatics forces are lit up simultaneously as λ evolves between 0 and 1. If in the initial configuration solvent molecules would happen to be too close to the ion, the repulsive term of the van der Waals interactions would make the system explode. Consequently one would greatly benefit from using a more intricate extended potential. The idea is to use *softcore* potentials, whose goal is to obtain finite pair-interaction energies, while insuring that we obtain a sufficiently smooth PMF profile [36, 12, 83]. One can then modify the extended Hamiltonian (3.24) by replacing the potential energy (3.25) with a potential whose dependency in λ is non-linear. We will introduce in detail the concept of softcore potentials in Section 6.4, with the goal of numerically implementing an efficient λ -dynamics in the Tinker-HP and Collective Variables module.

Problem 2 – Just as for the US method, the thermodynamic path λ is chosen ahead of the simulation. As a consequence, there is no reason whatsoever for the parameter λ to be able to capture the slow movement of the system and avoid metastability! To overcome this issue and to enhance the sampling of the phase space, one may bias the dynamics in the direction of the variable λ , as in the US method, and play with the values of the fictitious mass m_λ . In the scope of the numerical implementation of the λ -dynamics, we used the Adaptive Biasing Force method introduced in Section 3.3.2.4 to enhance the sampling of the phase space. Furthermore, around the end states $\lambda = 0$ and $\lambda = 1$, one may encounter substantial numerical errors. The FEP method bypasses this issue, but it is not the case for the λ -dynamics method: this led us to consider several options for the boundary conditions, as we will discuss in Sections 6.4 and 6.5.

3.4.3.2 The Orthogonal Space Random Walk method

λ -dynamics is an efficient method to compute free energy differences in the alchemical setting, but as said above, it is not perfect. In this section, we introduce one of the possible reasons why metastability is not avoided by the λ -dynamics, and present a possible way of resolving this issue. Let us briefly consider both conformational and alchemical transitions, where the system evolves from an initial state Γ_0 to a final state Γ_1 via the evolution of a transition coordinate ξ . The free energy difference computed with thermodynamic integration then reads:

$$\begin{aligned} \Delta A_{\Gamma_0 \rightarrow \Gamma_1} &= \int_{\xi=z_0}^{\xi=z_1} \left. \frac{dA}{d\xi} \right|_z dz \\ &= \int_{\xi=z_0}^{\xi=z_1} \left\langle \frac{dV_{ext}}{d\xi} - RT \frac{d \ln |J|}{d\xi} \right\rangle_z dz, \end{aligned}$$

where $|J|$ is the Jacobian matrix associated to the parameter ξ and

$$F_\xi := \frac{dV_{ext}}{d\xi} - RT \frac{d \ln |J|}{d\xi} \quad (3.27)$$

is the generalized force applied to ξ . Note that in the alchemical case, $\xi \equiv \lambda$ is a scalar quantity, and consequently one has $|J| = 0$ and $F_\xi = \frac{dV_{ext}}{d\xi}$.

A key problem of thermodynamic integration is the choice of the thermodynamic path between the two end states. One needs to sample all the intermediate values of ξ between its initial value z_0 and final value z_1 in order to properly sample the values ϕ of the generalized force F_ξ . This choice, as said in the previous section, is prior to the simulation, and *a priori* does not avoid metastability, leading to high computation times. Furthermore, the computation of F_ξ is intrinsically linked to the relaxation of the system after each move of the order parameter ξ : the generalized force sometimes has difficulties reaching the expected region for a given value of ξ . This is the *Hamiltonian lagging* phenomena [72]. An idea, suggested by Wei Yang and his colleagues [87, 88, 66, 89] is to use another order parameter $h(\cdot, \xi)$ being coupled to the evolution of ξ , and designed to capture the metastability in the space orthogonal to that of ξ . Then, an enhanced sampling method, called the *Orthogonal Space Random Walk* (OSRW) method, is used to reduce metastability in both the ξ and $h(\cdot, \xi)$ -spaces. A natural choice for the function $h(\cdot, \xi)$ is the generalized force F_ξ as defined in (3.27). To our knowledge, the OSRW method has led to few applications and is not easily reproducible.

In Section 6.1, we detail the original construction of the OSRW method. We will then discuss its limitations and proceed to motivate our intuition to rely on the choice of reaction coordinate $h(q, \lambda)$ suggested in [87, 88, 66, 89] and [1] in order to perform **alchemical** simulations using an enhanced sampling method such as the Adaptive Biasing Force method, which is known to be robust [57]. In order to test our hypothesis, we implemented the Adaptive Biasing Force algorithm and λ -dynamics in the Tinker-HP software for alchemical transitions. Doing so will allow us to obtain a two-dimensional free energy profile, which to our knowledge is new.

Contributions of the thesis

This thesis is structured into two parts. Each part is dedicated to a specific problem, that we will quickly present.

Study of the robustness of the Adaptive Biasing Force method with non-conservative forces

The Adaptive Biasing Force (ABF) and Projected Adaptive Biasing Force (PABF) algorithms introduced in Section 3.3.2 are importance sampling methods. They are known to enhance the sampling of a given system's configuration space, provided the interaction force \mathcal{F} between the particles is *conservative*, namely it can be written as the gradient of a potential energy V . Let us recall the dynamics given by the ABF method:

$$\begin{cases} dq_t &= (-\nabla V(q_t) + B_t \circ \xi(q_t)) \nabla \xi(q_t) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t(z) &= \mathbb{E}[F(q_t) | \xi(q_t) = z] \quad \forall z \in \mathcal{M}, \end{cases} \quad (4.1)$$

where the adaptive B_t is given by

$$B_t(z) = \mathbb{E}[F(q_t) | \xi(q_t) = z] = \int_{\Sigma_z} F(q) \pi_t^\xi(dq|z) \quad \forall z \in \mathcal{M},$$

with $\pi_t^\xi(\cdot|z)$ being the conditional measure at a fixed value $\xi(q) = z$ of the reaction coordinate, obtained from the instantaneous law π_t of the process $(q_t)_{t \geq 0}$. The Fokker-Planck equation satisfied by the law is the following:

$$\partial_t \pi_t = \beta^{-1} \Delta \pi_t + \nabla \cdot (\nabla V \cdot \pi_t).$$

One can check that the measure $\mu_A \propto \exp(-\beta(V - A \circ \xi))$ is a fixed point of the above equation. In other words, if $q_0 \sim \mu_A$, then $q_t \sim \mu_A$ for all $t \geq 0$ and $(q_t)_{t \geq 0}$ is exactly the diffusion (3.14).

For any given initial distribution and under mild assumptions, it has been shown [57], using entropy estimates and functional inequalities that:

- (1) The relative entropy $\mathcal{H}(\pi_t | \mu_A)$ of the law π_t with respect to μ_A converges exponentially

fast in the long-time limit.

- (2) The adaptive bias B_t converges in the long-time limit to the gradient of the system's free energy ∇A .
- (3) The flat histogram property is satisfied: the relative entropy $\mathcal{H}(\pi_t^\xi|\lambda)$ of the density π_t^ξ (namely, the law of $(\xi(q_t))_{t \geq 0}$) with respect to the Lebesgue measure converges exponentially fast in the long-time limit. In other words, the energy profile is flattened in the direction of the reaction coordinate ξ .

One may wonder if these above results still hold for the following Projected Adaptive Biasing Force algorithm:

$$\begin{cases} dX_t &= (-\nabla V(X_t) + B_t(\xi(X_t)) \nabla \xi(X_t)) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t &= \mathbf{P}_{L^2(\lambda)}(G_t) \\ G_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z] \quad \forall z \in \mathcal{M}, \end{cases} \quad (4.2)$$

where $\mathbf{P}_{L^2(\lambda)}(f)$ stands for the Helmholtz projection with respect to the Lebesgue measure λ . In 2015, H. Alrachid and T. Lelièvre [2] answered positively by proving (1)–(3). However, the algorithm considered was but a variant of the above algorithm, where the bias B_t is defined with the Helmholtz projection in the weighted space $L^2(\pi_t^\xi)$ instead of the Helmholtz projection in the $L^2(\lambda)$ space. This was motivated by some cancellations in the computations of the proofs. Nevertheless, as already noted in [2], the classical Helmholtz projection is used in practice. As a consequence, a first question arises:

Open problem: Can one prove (1)–(3) for the classical PABF method (4.2)?

As mentioned in Section 3.3.2.6, some force field models can lead to *hysteresis*, namely, there is a violation of the conservation of the Hamiltonian system's energy, and the interaction force \mathcal{F} is *a priori* no longer conservative. Using the ABF and PABF algorithms in this case may be an issue, as most of the aforementioned proofs rely on the fact that \mathcal{F} can be written as a gradient. Before using either one of the two methods, the following problems should be solved:

Open problem: In case one considers either dynamics (4.1) or (4.2) with a generic, possibly non-conservative interaction force \mathcal{F} :

- (i) Does the flat histogram property hold?
- (ii) Is there a stationary measure for the Fokker-Planck equation satisfied by the law of the process? Does the law of the process converge in the long-time limit to said stationary measure?
- (iii) Does there exist a stationary bias, to which the adaptive bias converges in the long-time limit?
- (iv) If the non-conservative force \mathcal{F} is a perturbation of a reference conservative force $-\nabla V$, can one bound the error made while estimating the free energy of the system and the error made while estimating an observable's canonical mean with the error made in the force fields?

In Part II we will provide answers to the above questions. Chapter 5 stems from the paper entitled "*The Adaptive Biasing Force algorithm with non-conservative force and related topics*", which has been submitted and is currently available as a preprint on the page arXiv:2102.09957. Results are explicitly presented in Section 5.2, and the proofs can be found in Section 5.3 for the flat histogram property, Section 5.4 for the existence of a stationary measure and stationary bias, and Section 5.5 for the long-time convergence of both algorithms.

Study of the Orthogonal Space Random Walk sampling method in the case of alchemical transitions

The ABF method relies on a reaction coordinate ξ , but the optimal choice of said reaction coordinate is still an open problem to this day. As pointed out in Section 3.2 and 3.4, choosing the right reaction coordinate will allow us to determine free energy differences, which are key quantities in biochemistry and pharmacology. In the second part of this thesis, we intend to address the question of choosing a transition coordinate in a very specific case, that of alchemical transitions, introduced in Section 3.2.2. Let us recall that during an alchemical transition, the system's evolution from an initial state $\lambda = 0$ towards a final state $\lambda = 1$ is indexed by a parameter $\lambda \in [0, 1]$. An example of alchemical transitions is the transformation of a ligand L_1 into another ligand L_2 : this allows to determine which one of the two is more prone to link itself with a target receptor. When a system undergoes an alchemical transition, the key quantity to estimate is the free energy difference $\Delta_{(0 \rightarrow 1)}A = A_1 - A_0$. Several methods are available to do so, as seen in Section 3.4.2. Among them is the λ -dynamics method, which allows for the parameter λ to evolve dynamically, enhancing the sampling in the λ -space. However, there is no reason for the λ -dynamics method to avoid metastability. Furthermore, some metastability may remain hidden in the space orthogonal to the λ -space. In several papers Wei Yang (Florida State University) and colleagues [87, 88, 66, 89] suggested the use of another order parameter $h(\cdot, \lambda)$ being coupled to the evolution of λ , and designed to capture the metastability in the space orthogonal to that of λ . Then, an enhanced sampling method, called the *Orthogonal Space Random Walk* (OSRW) method, is used to reduce metastability in both the λ and $h(\cdot, \lambda)$ -spaces. A natural choice for the function $h(\cdot, \lambda)$ is the generalized force F_λ as defined in (3.27). Note that as presented in Section 3.4.3.2, the OSRW method can also be used for conformational transitions. To our knowledge, the OSRW method has led to few applications and is not easily reproducible. One idea would then be to implement an alternative algorithm, say the ABF method, relying on the OSRW choice of reaction coordinate, and compare it to the original method. One would then need to answer the following questions:

The ABF method relies on a reaction coordinate ξ , but the optimal choice of said reaction coordinate is still an open problem to this day. As pointed out in Section 3.2 and 3.4, choosing the right reaction coordinate will allow us to determine free energy differences, which are key quantities in biochemistry and pharmacology. In the second part of this thesis, we intend to apprehend the question of choosing a transition coordinate in a very specific case, that of alchemical transitions, introduced in Section 3.2.2. Let us recall that during an alchemical transition, the system's evolution from an initial state $\lambda = 0$ towards a final state $\lambda = 1$ is indexed by a parameter $\lambda \in [0, 1]$. An example of alchemical transitions is the transformation of a ligand L_1 into another ligand L_2 : this allows to determine which one of the two is more prone to link itself with a target receptor. When a system undergoes an alchemical transition, the key quantity to estimate is the free energy difference $\Delta_{(0 \rightarrow 1)}A = A_1 - A_0$. Several methods are available to do so, as seen in Section 3.4.2. Among them is the λ -dynamics method, which allows for the parameter λ to evolve dynamically, enhancing the sampling in the λ -space. However, there is no reason for

the λ -dynamics method to avoid metastability. Furthermore, some metastability may remain hidden in the space orthogonal to the λ -space. In several papers Wei Yang (Florida State University) and colleagues [87, 88, 66, 89] suggested the use of another order parameter $h(q, \lambda)$ being coupled to the evolution of λ , and designed to capture the metastability in the space orthogonal to that of λ . Then, an enhanced sampling method, called the *Orthogonal Space Random Walk* (OSRW) method, is used to reduce metastability in both the λ and $h(q, \lambda)$ -spaces. A natural choice for the function $h(q, \lambda)$ is the generalized force F_λ as defined in (3.27). Note that as presented in Section 3.4.3.2, the OSRW method can also be used for conformational transitions. To our knowledge, the OSRW method has led to few applications and is not easily reproducible. One idea would then be to implement an alternative algorithm, say the ABF method, relying on the OSRW choice of reaction coordinate, and compare it to the original method. One would then need to answer the following questions: »»»> 9be8a9c708503eb6529abdd994c0c9295f6ffdf

Open problem:

- (i) Is the Adaptive Biasing Force method (3.18) applied with the reaction coordinate (λ, F_λ) successful in enhancing the sampling in both the λ space and its orthogonal space?
- (ii) Is (λ, F_λ) the optimal choice of reaction coordinate, or is the use of two coupled coordinates generally more efficient?

Part III is dedicated to the study of the Orthogonal Space Random Walk method. More precisely, we present a possible implementation of the OSRW method. We first recall the theoretical background of the λ -dynamics, its eventual applications in the alchemical setting along with its eventual limitations in Section 6.3, which will lead to the introduction of the concept of softcore potentials. Section 6.4 is dedicated to the implementation of softcore potentials in the Tinker-HP software, along with the implementation of the λ -dynamics in the Tinker-HP and Collective Variables module codes. This laid the foundations of a proxy between both softwares, expanding the field of possibilities for MD simulations. Eventually, we present our first numerical results in Section 6.4.5, where we test the efficiency of the implemented λ -dynamics to compute free energy differences for several toy models. Then, we proceed in Section 6.5 to compare the OSRW method with the two-dimensional ABF method: we will first present the implementation of the method in the Tinker-HP and Collective Variables module codes, and discuss the numerical challenges faced doing so. To finish, we will discuss how said challenges could be overcome, what remains to be done up to today, and present the future of the OSRW method.

Part II

Robustness of the Adaptive Biasing Force method with non-conservative forces and related topics

Et ainsi on comprend ce qu'il y a de vivant, de précaire aussi, dans nos connaissances et on évite de tomber dans le dogmatisme, qui est quelque peu décourageant comme toute chose définitive et morte.

PAUL LANGEVIN,
La valeur éducative de l'histoire des sciences, (1926)

Robustness of the Adaptive Biasing Force algorithm

with T. Lelièvre and P. Monmarché

The following chapter stems from the paper untitled "*The Adaptive Biasing Force algorithm with non-conservative force and related topics*", which has been submitted and is currently available as a preprint on the page arXiv:2102.09957. Some of its paragraphs may be redundant, as some content has already been introduced in the general introduction. For the sake of clarity, and to allow for this chapter to be read independently of the rest of this corpus, we decided to keep them as it is. If one already went through the introductory chapters 1, 2 and 3, we suggest a quick diagonal reading of Section 5.1, where we introduce the notations used throughout this paper.

We here propose a study of the Adaptive Biasing Force method's robustness under generic (possibly non-conservative) forces. We first ensure the flat histogram property is satisfied in all cases. We then introduce a fixed point problem yielding the existence of a stationary state for both the Adaptive Biasing Force and Projected Adapted Biasing Force algorithms, relying on generic bounds on the invariant probability measures of homogeneous diffusions. Using classical entropy techniques, we prove the exponential convergence of both biasing force and law as time goes to infinity, for both the Adaptive Biasing Force and the Projected Adaptive Biasing Force methods.

5.1 Introduction

After presenting in Sections 5.1.1, 5.1.2 and 5.1.3 the motivation and well-known results on the Adaptive Biasing Force (ABF) method applied to the overdamped Langevin dynamics with conservative forces, we present in Section 5.1.4 the dynamics we are interested in, namely the ABF method applied to the overdamped Langevin dynamics with non-conservative forces.

5.1.1 Setting

Let us work within the so-called *canonical ensemble* (or *NVT ensemble*), where a system of N particles is contained in a fixed volume \mathcal{V} , and is in contact with a thermostat of constant temperature T . Denote by $q = (q_1, \dots, q_N) \in \mathcal{D}$ the positions, $p = (p_1, \dots, p_N) \in \mathbb{R}^{dN}$ the momenta, and $(m_1, \dots, m_N) \in \mathbb{R}^N$ the masses of the particles, where \mathcal{D} is the configuration space and $d \in \{1, 2, 3\}$ is the space dimension. Usually, \mathcal{D} is an open subset of \mathbb{R}^{dN} (or \mathbb{T}^{dN} , where the dN -dimensional torus is viewed as the cube $[0, 1]^{dN}$ with opposite sides identified, in other words, $\mathbb{T}^{dN} = \mathbb{R}^{dN}/\mathbb{Z}^{dN}$). Interactions between particles are taken into account via a potential function $V : \mathcal{D} \rightarrow \mathbb{R}$, so that the system's total energy is given by the following Hamiltonian:

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

with $M = \text{diag}(m_1 I_d, \dots, m_N I_d)$ being the mass matrix. Since this Hamiltonian is *separable*, the positions and the momenta are independent variables in the canonical ensemble, namely under the probability distribution $Z^{-1} e^{-\beta H(q, p)} dq dp$ where $\beta = 1/(k_B T)$, with k_B being the Boltzmann constant, and $Z = \int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H(q, p)} dq dp$ is the normalization constant, or *partition function*. The momenta p being distributed according to a Gaussian measure, the main issue resides in sampling the positions q , which are distributed according to the *Boltzmann-Gibbs measure*:

$$\mu(dq) = Z_\mu^{-1} e^{-\beta V(q)} dq, \quad Z_\mu = \int_{\mathcal{D}} e^{-\beta V(q)} dq.$$

Thermodynamic properties are obtained by averaging functions of the microstate q which are called *observables*. Given an observable ψ , one would like to compute the following thermodynamic quantity:

$$\mathbb{E}_\mu[\psi] = \int_{\mathcal{D}} \psi d\mu.$$

One of the simplest dynamics to sample the Boltzmann-Gibbs measure is the *overdamped Langevin dynamics*:

$$dX_t = -\nabla V(X_t) dt + \sqrt{2\beta^{-1}} dW_t, \quad (5.1)$$

where $(W_t)_{t \geq 0}$ is a dN -dimensional standard Brownian motion, and $-\nabla V : \mathcal{D} \rightarrow \mathbb{R}^{dN}$ is the *interaction force*. Notice that here, the interaction force is *conservative*, namely it is the gradient of a function (here, minus the gradient of the potential energy V). Under reasonable assumptions on the potential V (see [57] for more details), the process $(X_t)_{t \geq 0}$ is ergodic with respect to μ . In other words, for any observable $\psi \in \mathcal{C}_0^\infty(\mathcal{D})$, the average over a trajectory of the process converges to the canonical average:

$$\lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_0^\tau \psi(X_t) dt = \mathbb{E}_\mu[\psi]. \quad (5.2)$$

5.1.2 Metastability, reaction coordinate and free-energy profiles

Computing thermodynamic averages can be troublesome, as microscopic and macroscopic timescales can violently differ. Typical microscopic phenomena occur on timescales of the order of 10^{-15} s, while macroscopic ones can take up to 1 h [58]. Furthermore, N needs to be sufficiently large so that the targeted macroscopic phenomena can emerge from the collective, microscopic behaviour of the system.

Such timescales differences are linked to the system's *metastability*: low-energy regions of the

configuration space are separated by either high-energy or high-entropy barriers. These regions are called metastable: the process (5.1) remains trapped in a metastable region and occasionally jumps to another one after a long period of time. From a probabilistic point of view, metastability is linked to the *multimodality* of the measure μ : likely regions are separated by low probability regions. The exploration of the state space by the process and the convergence of the trajectorial averages (5.2) can thus take a considerably long time.

One way of avoiding metastability is to capture some slow components of the dynamics $(X_t)_{t \geq 0}$. To do so, we consider *transition coordinates* (also called reaction coordinates or collective variables), namely mappings $\xi : \mathcal{D} \rightarrow \mathcal{M}$, where \mathcal{M} is a manifold of dimension $m \ll dN$. Transition coordinates are designed to provide a *coarse-grained information* on the system's state (for example, the dihedral angle of a molecule, in which case $\mathcal{M} = \mathbb{T}$, or the signed distance to a hypersurface of \mathcal{D} , in which case $\mathcal{M} = \mathbb{R}$). In other words, $\xi(x) \in \mathcal{M}$ is the macroscopic state of a microscopic state $x \in \mathcal{D}$. Designing a good reaction coordinate is a difficult problem, that will not be discussed further in the present work.

Decomposing

$$\mathcal{D} = \bigsqcup_{z \in \mathcal{M}} \Sigma_z = \bigsqcup_{z \in \mathcal{M}} \{q \in \mathcal{D} | \xi(q) = z\},$$

and denoting by σ_{Σ_z} the measure on Σ_z induced by the Lebesgue measure on \mathcal{D} , one can define the measure $\delta_{\xi(q)-z}(dq)$ by

$$\delta_{\xi(q)-z}(dq) = \frac{1}{\sqrt{\det G(q)}} \sigma_{\Sigma_z}(dq),$$

where $G = (\nabla \xi)^\top \nabla \xi$, in other words,

$$G_{i,j} = \nabla \xi_i \cdot \nabla \xi_j, \quad \text{for all } (i,j) \in \llbracket 1, m \rrbracket^2.$$

The free energy associated to ξ is then expressed as follows: for every $z \in \mathcal{M}$,

$$A(z) = -\frac{1}{\beta} \ln(Z_{\Sigma_z}), \quad Z_{\Sigma_z} = \int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq), \quad (5.3)$$

assuming V and ξ are such that $Z_{\Sigma_z} < +\infty$. As can be seen using the co-area formula [57], this definition ensures that the image of μ by ξ is given by

$$\xi \star \mu(dz) = \frac{e^{-\beta A(z)} dz}{\int_{\mathcal{M}} e^{-\beta A(u)} du}. \quad (5.4)$$

5.1.3 The Adaptive Biasing Force method

Introducing a reaction coordinate allows us to construct a less metastable dynamics, the idea being to substitute the potential V in (5.1) for a *biased potential* $V - A \circ \xi$. The new equilibrium measure is then

$$\mu_A(dq) = Z_{\mu_A}^{-1} e^{-\beta(V - A \circ \xi)(q)} dq, \quad (5.5)$$

where $Z_{\mu_A} = \int_{\mathcal{D}} e^{-\beta(V - A \circ \xi)(q)} dq$. Given the expression (5.4), the image of μ_A by ξ is the uniform measure: $\xi \star \mu_A = \lambda(\mathcal{M})^{-1} \mathbb{1}_{\mathcal{M}}$, with $\lambda(\mathcal{M})$ being the Lebesgue measure of \mathcal{M} (which is here assumed to be compact). Since, contrary to the initial probability measure $\xi \star \mu$, the uniform measure is no longer multimodal, we expect a faster sampling of the phase space, provided ξ is

well chosen so that μ_A is less multimodal than μ .

Although this change of potential can accelerate the phase space sampling, the free-energy A is *a priori* unknown. The main idea to get round this issue will be to approximate on the fly A , or ∇A , its derivative with respect to the reaction coordinate. To do so, we will consider the *Adaptive Biasing Force (ABF) algorithm* [24, 39]:

$$\begin{cases} dX_t &= (-\nabla V(X_t) + B_t(\xi(X_t)) \nabla \xi(X_t)) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z] \quad \forall z \in \mathcal{M}, \end{cases} \quad (5.6)$$

where $-\nabla V$ is the conservative interaction force, and F is the so-called *local mean force*, which is the vector with components $(F_i)_{i \in [1, m]}$ given by:

$$F_i = \sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \cdot \nabla V - \beta^{-1} \operatorname{div} \left(\sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \right),$$

where $G_{i,j}^{-1}$ denotes the (i, j) -component of the inverse of the matrix G defined above. This process is motivated by the fact that the aforementioned free energy satisfies:

$$\nabla A(z) = \mathbb{E}[F(X) | \xi(X) = z], \quad \forall z \in \mathcal{M} \quad \text{if } X \sim \mu_A,$$

so that μ_A is a fixed point of the Fokker-Planck equation associated to the process. In other words, if $X_0 \sim \mu_A$, then $X_t \sim \mu_A$ for all $t \geq 0$ and $(X_t)_{t \geq 0}$ is exactly the diffusion (5.1) with the biased potential $V - A \circ \xi$.

Starting from another initial distribution, using entropy estimates and functional inequalities, it has been proven in [57], under mild assumptions, that this fixed point is in fact an attractor of the dynamics, in the sense that B_t converges to ∇A in the long-time limit, and the law of X_t converges to μ_A .

Remark 15.

▷ In some cases \mathcal{M} is not bounded, for example when ξ is a distance. If so, an additional confining potential $W \circ \xi$ is needed in the drift [57].

▷ As discussed in [57], the algorithm (5.6) can be modified in order to obtain a diffusive behaviour for the law of $\xi(X_t)$. Additional terms depending on ξ are added to obtain the following variant:

$$\begin{cases} dX_t &= (-\nabla V + B_t \circ \xi - \nabla W \circ \xi + \beta^{-1} \nabla \ln(|\nabla \xi|^{-2})) |\nabla \xi|^{-2}(X_t) dt + \sqrt{2\beta^{-1}} |\nabla \xi|^{-1}(X_t) dW_t \\ B_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z], \quad \forall z \in \mathcal{M}. \end{cases}$$

In this case the long-time convergence of B_t towards ∇A is stronger than in the case of (5.6), in that it requires fewer hypotheses.

We might also consider a variant of the ABF method, namely the *Projected Adaptive Biasing Force (PABF) algorithm*, introduced in [2]:

$$\begin{cases} dX_t &= (-\nabla V(X_t) + B_t(\xi(X_t)) \nabla \xi(X_t)) dt + \sqrt{2\beta^{-1}} dW_t \\ B_t &= \mathbf{P}_{L^2(\lambda)}(G_t) \\ G_t(z) &= \mathbb{E}[F(X_t) | \xi(X_t) = z] \quad \forall z \in \mathcal{M}, \end{cases}$$

where $\mathbf{P}_{L^2(\lambda)}(f)$ stands for the Helmholtz projection with respect to the Lebesgue measure λ of a vector field f on an open bounded set $\mathcal{M} \subset \mathbb{R}^{dN}$ with Lipschitz boundary $\partial\mathcal{M}$ [3]. In other words, it is the gradient of the minimizer on $\{g \in H^1(\mathcal{M}), \int_{\mathcal{M}} g dx = 0\}$ of

$$g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 dx.$$

More generally, if ν is a continuous positive measure on \mathcal{M} , the Helmholtz projection with respect to ν is the minimizer on $\{g \in H^1(\mathcal{M}), \int_{\mathcal{M}} g dx = 0\}$ of $g \mapsto \int_{\mathcal{M}} |f(x) - \nabla g(x)|^2 \nu(dx)$.

5.1.4 The non-conservative case

From now on, we only consider periodic boundary conditions and reaction coordinates that are Euclidean coordinates of the system, namely $\mathcal{D} = \mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ for some $n \in \mathbb{N}^*$, $\mathcal{M} = \mathbb{T}^m$ for $m \in \mathbb{N}^*$ such that $m \leq n$ and $\xi(x, y) = x$, where we decompose $(x, y) \in \mathcal{D}$ with $x \in \mathbb{T}^m$ and $y \in \mathbb{T}^{n-m}$. This latter restriction may seem quite narrow: nevertheless, it is the generic case used for alchemical reactions [49]. Besides, more general reaction coordinates can be reduced to this setting by adding extended variables [31]. Here, such restriction is made only for the sake of clarity: most arguments could be extended (at the price of heavier computations) to the general case $\xi(x, y) \in \mathcal{M}$.

We are interested in the case where the force in (5.1) is not necessarily conservative, namely is not the gradient of some potential energy V . There are several motivations for this approach, one of them being that the numerical computation of conservative forces $-\nabla V$ sometimes relies on approximations which make the force *a priori* not conservative, in particular in the context of *ab initio* molecular dynamics, see e.g. [74, 69, 20]. In this case, one is interested in knowing if, by controlling the error made on the force $-\nabla V$, one can deduce an estimation of the error made on the system's free energy. The robustness of a diffusion's invariant measure with respect to the perturbation of its drift is a classical problem (see e.g. Section 5.4.3), but note that in the ABF case, the adaptive procedure makes the question more subtle. Moreover, the convergence of the ABF method in such a context cannot be deduced from the aforementioned convergence analysis. We consequently consider the ABF algorithm in the case where $-\nabla V$ is replaced by a general force field $\mathcal{F} \in \mathcal{C}^1(\mathcal{D}, \mathbb{R}^n)$ that we rewrite as $\mathcal{F}(x, y) = (\mathcal{F}_1(x, y), \mathcal{F}_2(x, y)) \in \mathbb{R}^m \times \mathbb{R}^{n-m}$. The local mean force is simply $F = -\mathcal{F}_1$, and the corresponding process is thus, for all $t \geq 0$:

$$\begin{cases} dX_t &= \mathcal{F}_1(X_t, Y_t)dt + B_t(X_t)dt + \sqrt{2\beta^{-1}}dW_t^1 \\ dY_t &= \mathcal{F}_2(X_t, Y_t)dt + \sqrt{2\beta^{-1}}dW_t^2 \end{cases} \quad (5.7)$$

where $W = (W^1, W^2)$ is a standard Brownian motion on $\mathbb{T}^m \times \mathbb{T}^{n-m}$, and, given the average mean force

$$G_t(x) = -\mathbb{E}[\mathcal{F}_1(X_t, Y_t) | X_t = x], \quad \forall t \geq 0, \forall x \in \mathbb{T}^m,$$

one has for all $t \geq 0$ and $x \in \mathbb{T}^m$, in the case of the ABF method,

$$B_t(x) = G_t(x),$$

or, in the case of the PABF method,

$$B_t(x) = P_{L^2(\lambda)}(G_t)(x) := \nabla H_t(x).$$

In either case, denoting by π_t the law of $Z_t = (X_t, Y_t)$ and $\pi_t^\xi(x) = \int_{\mathbb{T}^{n-m}} \pi_t(x, y) dy$ the density of $X_t = \xi(Z_t)$, then

$$G_t(x) = \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x, y) \frac{\pi_t(x, y)}{\pi_t^\xi(x)} dy,$$

so that π_t is a weak solution of the Fokker-Planck equation associated to (5.7), that is

$$\begin{cases} \partial_t \pi_t &= \beta^{-1} \Delta \pi_t - \nabla \cdot (\mathcal{F} \pi_t) - \nabla_x \cdot (B_t \pi_t) \\ B_t &= \begin{cases} G_t & \text{in the ABF case} \\ \nabla H_t = P_{L^2(\lambda)}(G_t) & \text{in the PABF case} \end{cases} \\ G_t(x) &= \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x, y) \frac{\pi_t(x, y)}{\pi_t^\xi(x)} dy \quad \forall x \in \mathbb{T}^m. \end{cases} \quad (5.8)$$

For a given initial condition π_0 , the existence of the process and the proof that it admits a density with respect to the Lebesgue measure, being a strong solution of (5.8), can be established by fixed point arguments or by the convergence of an interacting particles system [45]. We will not address this question here. As a consequence, we would like to emphasize that our arguments will be partially formal, in the sense that we work under the assumption that a density π_t that solves (5.8) exists and is sufficiently regular so that the algebraic computations in the proofs are valid.

Let us emphasize that the bias B_t in (5.8) (i.e. either G_t or $\nabla H_t = P_{L^2(\lambda)}(G_t)$) depends on π_t , which makes (5.8) a non-linear PDE.

Remark 16. *In the conservative case, where $\mathcal{F} = -\nabla V$, and $\mu \propto e^{-\beta V}$, up to an additive constant, the free energy A is characterized by either one of these properties:*

1. $\xi_* \mu \propto e^{-\beta A}$ (distribution of the reaction coordinate at equilibrium).
2. $\nabla A = \mathbb{E}[\nabla_1 V(Z) | \xi(Z) = \cdot]$ with $Z \sim \mu$ (average local mean force at equilibrium).
3. $\nabla A = \mathbb{E}[\nabla_1 V(Z) | \xi(Z) = \cdot]$ with $Z \sim \mu_A$ (fixed point of the ABF algorithm).

In the non-conservative case, there is no reason for these various definitions to coincide. Besides, $x \mapsto \mathbb{E}[-\mathcal{F}_1(Z) | \xi(Z) = x]$ is a priori not a gradient. Denoting by $\mu_{\mathcal{F}}$ the invariant measure of the

non-biased, out-of-equilibrium dynamics $\partial_t \pi_t = \beta^{-1} \Delta \pi_t - \nabla \cdot (\mathcal{F} \pi_t)$, we are then led to consider the (in general different) functions H_1 , H_2 and H_3 given, up to an additive constant, by

1. $\xi \star \mu_{\mathcal{F}} \propto e^{-\beta H_1}$.
2. $\nabla H_2 = \mathbf{P}_{L^2(\lambda)} (\mathbb{E}[-\mathcal{F}_1(Z)|\xi(Z) = \cdot])$ with $Z \sim \mu_{\mathcal{F}}$.
3. $\nabla H_3 = \mathbf{P}_{L^2(\lambda)} (\mathbb{E}[-\mathcal{F}_1(Z)|\xi(Z) = \cdot])$ with $Z \sim \pi_{\infty}^{\mathcal{F}}$ an equilibrium of the (P)ABF algorithm.

In other words, in the non-conservative case, an equilibrium of an adaptive algorithm yields an alternative generalization of the notion of free energy that does not coincide in general with the log-density of the law of the reaction coordinates at (unbiased) equilibrium, and whose gradient is not in general the average local mean force at (unbiased) equilibrium.

Outline of this paper. Section 5.2 introduces several preliminary notions, before stating the main results. Section 5.3 focuses on the law π_t^{ξ} of the process $(X_t)_{t \geq 0} = (\xi(Z_t))_{t \geq 0}$. More precisely, we show that π_t^{ξ} satisfies a particular Fokker-Planck equation, which differs depending on the method considered, and that π_t^{ξ} converges in the long-time limit to the Lebesgue measure λ . Section 5.4 then states several results on the invariant measure of a generic diffusion, in order to address the issue of the existence of both stationary measure and stationary bias to equation (5.8), and later handles the robustness of the conservative equilibrium to non-conservative perturbations. Eventually, Section 5.5 is devoted to the long-time convergence of both the ABF and PABF methods, in the conservative case, with a force $\mathcal{F} = -\nabla V$ (generalizing in particular results from [2]), and in the non-conservative case, with a generic force \mathcal{F} .

5.2 Main results

Before stating the main results of this paper, we encourage the reading of Section 3.1.2.2, where we recall the definitions of the relative entropy of two measures along with several functional inequalities.

5.2.1 Precise statements of the results

In all this section, π_t satisfies (5.8). First of all, let us consider the equation satisfied by the density π_t^{ξ} in the general case where \mathcal{F} is either conservative or non-conservative.

Lemma 1. *The density π_t^{ξ} satisfies the following Fokker-Planck equation:*

$$\partial_t \pi_t^{\xi} = \beta^{-1} \Delta \pi_t^{\xi} - \nabla \cdot \left((B_t - G_t) \pi_t^{\xi} \right). \quad (5.9)$$

Proof. Take a test function $\varphi \in \mathcal{C}^\infty(\mathbb{T}^m)$. Then, using an integration by parts,

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{T}^m} \varphi \pi_t^\xi &= \frac{d}{dt} \int_{\mathbb{T}^n} \varphi(x) \pi_t(x, y) dx dy \\ &= \int_{\mathbb{T}^n} (\beta^{-1} \Delta_x \varphi(x) + (\mathcal{F}_1(x, y) + B_t(x)) \nabla_x \varphi(x)) \pi_t(dx, dy) \\ &= \int_{\mathbb{T}^m} \left(\beta^{-1} \Delta_x \varphi(x) \pi_t^\xi(x) + \left(\int_{\mathbb{T}^{n-m}} \mathcal{F}_1(x, y) \frac{\pi_t(x, y)}{\pi_t^\xi(x)} dy + B_t(x) \pi_t^\xi(x) \right) \nabla_x \varphi(x) \right) dx \\ &= \int_{\mathbb{T}^m} (\beta^{-1} \Delta_x \varphi + (B_t - G_t) \nabla_x \varphi) \pi_t^\xi. \end{aligned}$$

□

Remark that in [2, Proposition 2], the Helmholtz projection is done in $L^2(\pi_t^\xi)$, so that $\nabla \cdot ((B_t - G_t) \pi_t^\xi) = 0$ and one ends up with the heat equation. Here, we get the heat equation in the ABF case ($B_t = G_t$) and, in the PABF case ($B_t = P_{L^2(\lambda)}(G_t)$), an additional time-dependent divergence-free drift.

Remark 17. Since the density π_t^ξ , as well as constants, satisfies the Fokker-Planck equation (5.9) which preserves positivity, provided there exists $m_0^\xi > 0$ such that $\pi_0^\xi \geq m_0^\xi$, one can show that $\pi_t^\xi \geq m_0^\xi$ for all $t \geq 0$ on the torus \mathbb{T}^m . Note that if π_0^ξ was to be zero at some points or not sufficiently smooth, the conditional mean G_0 given in (5.8) might not be well defined.

In view of Remark 17, from now on, assume the following:

Assumption 4. The initial condition π_0 admits a smooth density with respect to the Lebesgue measure, such that π_0^ξ is positive.

As a consequence, the conditional means G_t are well defined for all $t \geq 0$, along with the entropy $\mathcal{H}(\pi_0 | \lambda)$, which is ensured to be finite. Furthermore, π_0^ξ belongs to $L^2(\mathbb{T}^m)$.

Both the ABF and PABF algorithms are designed in order to ensure that all the values of the transition coordinate have been visited. In other words, the density of $\xi(X_t, Y_t)$ should converge to a flat histogram, namely the Lebesgue measure λ . In the conservative case, this is known to hold in both the ABF case [56] and the PABF case [2]. We now extend the flat histogram property to the general –possibly non-conservative– case.

Proposition 3. For both the ABF and PABF algorithm, under Assumption 4, π_t^ξ converges towards the Lebesgue measure as $t \rightarrow \infty$. More precisely, for all $t \geq 0$:

$$\mathcal{H}(\pi_t^\xi | \lambda) \leq e^{-8\beta^{-1}\pi^2 t} \mathcal{H}(\pi_0^\xi | \lambda).$$

Furthermore, the entropic convergence of the density can be strengthened to an L^∞ one, that will prove useful in the rest of the study:

Proposition 4. For both the ABF and PABF algorithm, under Assumption 4, there exists $C > 0$ such that for all initial distribution $\pi_0^\xi \in L^2(\mathbb{T}^m)$, for all $t \geq 1$:

$$\|\pi_t^\xi - 1\|_\infty \leq C e^{-4\beta^{-1}\pi^2 t} \|\pi_0^\xi - 1\|_2.$$

As detailed in [57, 2], in the conservative case $\mathcal{F} = -\nabla V$, $\pi_\infty = \mu_A$ is a stationary state of (5.8). In the non-conservative case, the existence of such a stationary state may be unclear, and this issue will be treated in Theorem 1 below, which will be proved in Section 5.4.2. For now, let us consider the following assumption:

Assumption 5. *The interaction force \mathcal{F} is in $\mathcal{C}^1(\mathbb{T}^n, \mathbb{R}^n)$, and we denote by $M > 0$ a constant such that for all $y \in \mathbb{T}^{n-m}$, $x \mapsto \mathcal{F}_1(x, y)$ is M -Lipschitz.*

Theorem 1. *For both the ABF and PABF algorithms, under Assumption 5, there exists a couple of stationary measure and bias $(\pi_\infty^{\mathcal{F}}, B_\infty^{\mathcal{F}})$ to (5.8), such that $\pi_\infty^{\mathcal{F}} \in \mathcal{C}^0(\mathbb{T}^n)$ is strictly positive. As a consequence,*

- (i) $\pi_\infty^{\mathcal{F}}$ satisfies a log-Sobolev inequality for some constant $R > 0$,
- (ii) the conditional density $y \mapsto \pi_{\infty, x}^{\mathcal{F}}(y) := \pi_\infty^{\mathcal{F}}(x, y) / \pi_\infty^{\mathcal{F}, \xi}(x)$ satisfies a log-Sobolev inequality for some constant $\rho > 0$, for all $x \in \mathbb{T}^m$.

Remark 18. *Note that there is no reason whatsoever for $\pi_\infty^{\mathcal{F}}$ to be the same in both the ABF and PABF case. Nevertheless, as shown in Proposition 3, $\pi_\infty^{\xi} \equiv 1$ in all cases.*

Remark 19. *An important remark is that, at small temperatures (i.e. $\beta \gg 1$), the optimal log-Sobolev constant of a probability measure with density proportional to $\exp(\beta W)$ for some W , roughly scales like $\exp(\beta d_W)$ where d_W is the so-called critical depth of W [63] (the critical depth is the highest energy barrier to overcome in order to reach a global minimum of W). If the transition coordinate is well-chosen, the metastability in the orthogonal space should be small, meaning that for all $x \in \mathbb{T}^m$ the critical depth of $W(x, \cdot)$ should be small with respect to the critical depth of W . As a consequence, as a function of β , ρ is expected to be much larger than the log-Sobolev constant of $\mu \propto e^{-\beta V}$, which is the convergence rate to equilibrium of the original (unbiased) dynamics (5.1).*

The following result deals with the robustness of the conservative equilibrium to non-conservative perturbations, and will be proved in Section 5.4.3.

Proposition 5. *For the PABF algorithm, under Assumption 4 and Assumption 5, for all $V \in \mathcal{C}^2(\mathbb{T}^n)$ and $p \geq 1$, there exists $K_V > 0$ and $K_p > 0$ such that the following holds. Denote by A the free energy associated to V (see equation (5.3) for the definition of A). For all $\mathcal{F} \in \mathcal{C}^1(\mathbb{T}^n)$ satisfying $\|\mathcal{F} + \nabla V\|_\infty \leq 1$, for all equilibrium measure $\pi_\infty^{\mathcal{F}}$ of (5.8), considering the corresponding bias $\nabla H_\infty^{\mathcal{F}}$, one has*

$$\|\nabla A - \nabla H_\infty^{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \leq K_V K_p \|\mathcal{F} + \nabla V\|_\infty,$$

and, for all $\psi \in L^\infty(\mathbb{T}^n)$, considering

$$\hat{I}_\psi := \frac{\int_{\mathbb{T}^n} \psi(x, y) e^{-\beta H_\infty^{\mathcal{F}}(x)} \pi_\infty^{\mathcal{F}}(x, y) dx dy}{\int_{\mathbb{T}^n} e^{-\beta H_\infty^{\mathcal{F}}(x)} \pi_\infty^{\mathcal{F}}(x, y) dx dy},$$

one has

$$\left| \int_{\mathbb{T}^n} \psi d\mu - \hat{I}_\psi \right| \leq K_V \|\psi\|_\infty \|\mathcal{F} + \nabla V\|_\infty.$$

The first point of Proposition 5 states that, if the error on the forces ∇V is small, then the bias of the free energy estimation is small. The second point states that similarly, the bias on the computations of averages with respect to μ is small if the error on the forces is small. Indeed, in practice, in order to compute averages with respect to the initial target law μ from the biased trajectory, two strategies are available: either standard importance sampling re-weighting, or estimation of the conditional expectations given $\xi(X, Y) = x$ and then average with respect to $\exp(-H_{\mathcal{F}}(x))$. In both cases, if $-\nabla V$ is replaced by \mathcal{F} due to some numerical errors and the process converges in large time towards an equilibrium $\pi_{\infty}^{\mathcal{F}}$, then a quantity of the form $\int_{\mathbb{T}^n} \psi d\mu$ is approximated by an estimator that converges in large time towards the quantity \hat{I}_{ψ} defined in Proposition 5.

Finally, we turn to the long-time convergence of the density π_t on the whole space. The first theorem concerns the classical, conservative case, whereas the second concerns the general case, where the force \mathcal{F} can be non conservative. These will respectively be proved in sections 5.5.2 and 5.5.3.

Theorem 2. *Let us consider (π_t, B_t) solution of (5.8) for either the ABF or PABF algorithm, under Assumption 4 and Assumption 5. Let us suppose moreover that $\mathcal{F} = -\nabla V$, with $V \in \mathcal{C}^2(\mathbb{T}^n)$. Then, there exists $K > 0$ such that, for all $\varepsilon > 0$ and for all $t \geq 0$:*

$$\mathcal{H}(\pi_t | \mu_A) \leq K \left(1 + \frac{1}{\varepsilon^2}\right) e^{-(\Lambda - \varepsilon)t},$$

with μ_A being given by (5.5), $\Lambda = (8\pi^2 \wedge 2\rho) \beta^{-1}$ in the ABF case, $\Lambda = (4\pi^2 \wedge 2\rho) \beta^{-1}$ in the PABF case, and ρ is the log-Sobolev constant of the conditional density $y \mapsto \mu_{A,x}(y) := \mu_A(x, y) / \mu_A^{\xi}(x)$. Furthermore, (5.8) consequently admits a unique stationary state: using the notations of Theorem 1, $(\pi_{\infty}^{-\nabla V}, B_{\infty}^{-\nabla V}) = (\mu_A, \nabla A)$.

This extends [57, Theorem 1], which is restricted to the ABF algorithm with $m = 1$. Besides, for the PABF algorithm, [2, Theorem 1] is a similar convergence result but for a variant of the algorithm where the classical Helmholtz projection in $L^2(\lambda)$ is replaced by the Helmholtz projection in the weighted space $L^2(\pi_t^{\xi})$. This variant is motivated in [2] by some cancellations in the computations of the proofs. Nevertheless, as already noted in [2], the classical Helmholtz projection is used in practice. Theorem 2 in the PABF case is thus a new result which fills a gap between the existing theoretical convergence results and the practical algorithm.

Remark 20. *For $t > 0$, applying Theorem 2 with $\varepsilon = 1/t$ yields*

$$\mathcal{H}(\pi_t | \mu_A) \leq K e^1 (1 + t^2) e^{-\Lambda t}.$$

The next results address the general –possibly non-conservative– case, and as such are new.

Theorem 3. *Let us consider (π_t, B_t) solution of (5.8) for either the ABF or PABF algorithm, under Assumption 4 and Assumption 5. Let $\pi_\infty^{\mathcal{F}}, R, \rho$ be a stationary measure for (5.8) and the two associated constants, as introduced in Theorem 1. Suppose moreover that $M\beta < 2\rho$. Then there exists $K \geq 0$ such that, for all $t \geq 0$:*

$$\mathcal{H}(\pi_t | \pi_\infty^{\mathcal{F}}) \leq K e^{-\Lambda t},$$

with $\Lambda = 2R(1 - \frac{M\beta}{2\rho})\beta^{-1}$. As a consequence, the dynamics (5.8) admits a unique stationary state.

Eventually, one has the following result, which will be proved in Section 5.5.4.

Corollary 1. *Under the settings of either Theorem 2 or 3, there exists a unique stationary state $(\pi_\infty^{\mathcal{F}}, B_\infty^{\mathcal{F}})$ for the dynamics (5.8). Furthermore, there exists $K \geq 0$ such that for all $t \geq 0$,*

$$\int_{\mathbb{T}^m} |B_t - B_\infty^{\mathcal{F}}|^2 dx \leq K e^{-\Lambda t},$$

where Λ is given by either Theorem 2 (where $\mathcal{F} = -\nabla V$) or 3 (where \mathcal{F} is general).

Remark 21. *A direct consequence of the Csizàr-Kullback inequality (3.3) combined with either Theorem 2 or Theorem 3 is that for all $t \geq 0$*

$$\|\pi_t - \pi_\infty^{\mathcal{F}}\|_{TV} \leq \sqrt{2K} e^{-\frac{1}{2}\Lambda t},$$

where $K, \Lambda \geq 0$ are given by either Theorem 2 (where $\mathcal{F} = -\nabla V$) or 3 (where \mathcal{F} is general).

Theorem 3 shows the exponential convergence to a unique stationary state for the ABF and PABF algorithms even for non-conservative forces. Notice that the rate of convergence obtained in Theorem 2 for conservative forces is better than the rate of convergence in Theorem 3. It would be interesting to further investigate the sharpness of these rates.

The rest of this paper is devoted to the proofs of the results stated in this section. From now on, and without loss of generality, we will assume that $\beta = 1$. Note that the assumption of Theorem 3 now becomes $M < 2\rho$. An adequate change of variable to then deduce the results for $\beta \neq 1$ is: $\tilde{t} = \beta^{-1}t$, $\tilde{\mathcal{F}}(x, y) = \beta\mathcal{F}(x, y)$, $\tilde{W}^1(x) = \beta W^1(x)$, $\tilde{W}^2(y) = \beta W^2(y)$, and $\tilde{\pi}_t(x, y) = \pi_t(x, y)$, for all $t \geq 0$ and for all $(x, y) \in \mathbb{T}^n$.

5.3 Law of the transition coordinate

After proving in Section 5.3.1 the long-time entropic convergence of the density π_t^ξ towards the Lebesgue measure λ , we prove in Section 5.3.2 its long-time L^∞ -convergence, by relying on a Nash inequality on the n -dimensional torus and on the proof of [5, Theorem 6.3.1].

5.3.1 Proof of Proposition 3

Proof. One has:

$$\frac{d}{dt} \mathcal{H}(\pi_t^\xi | \lambda) = \frac{d}{dt} \int_{\mathbb{T}^m} \pi_t^\xi \ln \pi_t^\xi.$$

Considering $\mathcal{L}_t \mu = \nabla \cdot (\nabla \mu - (B_t - G_t) \mu)$:

$$\frac{d}{dt} \int_{\mathbb{T}^m} \pi_t^\xi = \int_{\mathbb{T}^m} \mathcal{L}_t \pi_t^\xi = 0.$$

One gets, using integration by parts,

$$\begin{aligned} \frac{d}{dt} \mathcal{H}(\pi_t^\xi | \lambda) &= \int_{\mathbb{T}^m} \ln(\pi_t^\xi) \mathcal{L}_t \pi_t^\xi + \int_{\mathbb{T}^m} \mathcal{L}_t \pi_t^\xi \\ &= - \int_{\mathbb{T}^m} \frac{|\nabla \pi_t^\xi|^2}{\pi_t^\xi} + \int_{\mathbb{T}^m} (B_t - G_t) \nabla \pi_t^\xi \\ &= - \int_{\mathbb{T}^m} \frac{|\nabla \pi_t^\xi|^2}{\pi_t^\xi} \quad (\text{since } \nabla \cdot (B_t - G_t) = 0) \\ &= - \int_{\mathbb{T}^m} \left| \nabla \ln \left(\frac{\pi_t^\xi}{\lambda} \right) \right|^2 \pi_t^\xi \\ &= -I(\pi_t^\xi | \lambda). \end{aligned} \tag{5.10}$$

Since the Lebesgue measure λ satisfies a log-Sobolev inequality of constant $4\pi^2$ [5, Proposition 5.7.5(ii)], we have:

$$\partial_t \mathcal{H}(\pi_t^\xi | \lambda) \leq -2(4\pi^2) \mathcal{H}(\pi_t^\xi | \lambda),$$

which concludes the proof of Proposition 3, denoting by $\pi_\infty^\xi \equiv \lambda$ the long-time limit of π_t^ξ . \square

5.3.2 Proof of Proposition 4

We first state a Nash inequality on the n -dimensional torus.

Lemma 2. *For all $n \in \mathbb{N}^*$, there exists $a = a(n) > 0$ such that for all functions $u \in H^1(\mathbb{T}^n)$:*

$$\|u\|_2^2 \leq 2\|u\|_1^2 + a\|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}}. \tag{5.11}$$

Proof. Let us recall that $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$. We consider $L^2(\mathbb{T}^n)$ equipped with the inner product $\langle u, v \rangle := \int_{\mathbb{T}^n} u(x) \bar{v}(x) dx$. The sequence $\{e^{2\pi i k x}\}_{k \in \mathbb{Z}^n}$ is an orthonormal basis of $L^2(\mathbb{T}^n)$. Now given a function $u \in L^2(\mathbb{T}^n)$ and its Fourier coefficients

$$c_k = \int_{\mathbb{T}^n} u(x) e^{-2\pi i k x} dx, \quad \forall k \in \mathbb{Z}^n,$$

denoting by $k = (k_1, \dots, k_n)$ a vector in \mathbb{Z}^n , and $|k| = \sqrt{\sum_{j=1}^n |k_j|^2}$, the Parseval identity yields

$$\|u\|_2^2 = \sum_{k \in \mathbb{Z}^n} |c_k|^2, \quad \|\nabla u\|_2^2 = \sum_{k \in \mathbb{Z}^n} |k|^2 |c_k|^2.$$

Let $\rho > 0$ to be fixed later on. One has, considering $\|k\|_\infty = \max_{j \in \llbracket 1, n \rrbracket} \{|k_j|\}$:

$$\begin{aligned} \|u\|_2^2 &= \sum_{k \in \mathbb{Z}^n} |c_k|^2 = \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \sum_{\|k\|_\infty > \rho} |c_k|^2 \\ &\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \sum_{\|k\|_\infty > \rho} \|k\|_\infty^2 |c_k|^2 \\ &\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \sum_{\|k\|_\infty > \rho} |k|^2 |c_k|^2 \\ &\leq \sum_{\|k\|_\infty \leq \rho} |c_k|^2 + \frac{1}{\rho^2} \|\nabla u\|_2^2. \end{aligned}$$

And:

$$\sum_{\|k\|_\infty \leq \rho} |c_k|^2 = \sum_{\|k\|_\infty \leq \rho} \left| \int_{\mathbb{T}^n} u(x) e^{-2\pi i k x} dx \right|^2 \leq \|u\|_1^2 \sum_{\|k\|_\infty \leq \rho} 1 \leq (2\rho + 1)^n \|u\|_1^2.$$

Consequently:

$$\|u\|_2^2 \leq 3^n (\rho \vee 1)^n \|u\|_1^2 + \frac{1}{\rho^2} \|\nabla u\|_2^2. \quad (5.12)$$

We now distinguish between two cases:

(i) If $3^n \|u\|_1^2 \leq \|\nabla u\|_2^2$, by choosing

$$\rho = 3^{-\frac{n}{n+2}} \frac{\|\nabla u\|_2^{\frac{2}{n+2}}}{\|u\|_1^{\frac{2}{n+2}}} \geq 1,$$

inequality (5.12) yields:

$$\begin{aligned} \|u\|_2^2 &\leq 3^n 3^{-\frac{n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}} + 3^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}} \\ &= 2 \cdot 3^{\frac{2n}{n+2}} \|\nabla u\|_2^{\frac{2n}{n+2}} \|u\|_1^{\frac{4}{n+2}}. \end{aligned} \quad (5.13)$$

(ii) If $3^n \|u\|_1^2 \geq \|\nabla u\|_2^2$, one wishes to rely on the Poincaré-Wirtinger inequality on the torus \mathbb{T}^n . The optimal Poincaré constant in $H_0^1(\mathbb{T}^n)$ being equal to λ_1^{-1} , where $\lambda_1 = 4\pi^2$ is the first non trivial eigenvalue of the negative Laplacian $-\Delta$, one can consider the following Poincaré-Wirtinger inequality:

$$\|u - \bar{u}\|_2^2 \leq \frac{1}{4\pi^2} \|\nabla u\|_2^2, \quad \forall u \in H^1(\mathbb{T}^n), \quad (5.14)$$

where $\bar{u} = \int_{\mathbb{T}^n} u(x) dx$. One consequently gets:

$$\begin{aligned}
\|u\|_2^2 &\leq 2\bar{u}^2 + 2\|u - \bar{u}\|_2^2 \\
&\leq 2\|u\|_1^2 + 2\frac{1}{4\pi^2}\|\nabla u\|_2^2 \\
&= 2\|u\|_1^2 + \frac{1}{2\pi^2}\|\nabla u\|_2^{\frac{2n}{n+2}}\|\nabla u\|_2^{\frac{4}{n+2}} \\
&\leq 2\|u\|_1^2 + \frac{1}{2\pi^2}3^{\frac{2n}{n+2}}\|\nabla u\|_2^{\frac{2n}{n+2}}\|u\|_1^{\frac{4}{n+2}}.
\end{aligned} \tag{5.15}$$

Combining (5.13) and (5.15), one obtains:

$$\|u\|_2^2 \leq 2\|u\|_1^2 + 3^{\frac{2n}{n+2}} \left(\frac{1}{2\pi^2} \vee 2 \right) \|\nabla u\|_2^{\frac{2n}{n+2}} \|\nabla u\|_1^{\frac{4}{n+2}}$$

which yields (5.11), with $a = 2 \cdot 3^{\frac{2n}{n+2}}$. \square

We are now in position to prove Proposition 4.

Proof of Proposition 4. We will rely on the idea of the proof of [5, Theorem 6.3.1]. Let us start with two preliminary results. Let $\varphi \in C^\infty(\mathbb{T}^m)$ be a test function and consider:

$$\forall z \in \mathbb{T}^m, \quad \varphi_t(z) = \mathbb{E}_z[\varphi(Z_t)] = \mathbb{E}[\varphi(Z_t) | Z_0 = z],$$

where $(Z_t)_{t \geq 0}$ satisfies the following dynamics:

$$dZ_t = (B_t - G_t)(Z_t)dt + \sqrt{2}dW_t,$$

where $(W_t)_{t \geq 0}$ is a n -dimensional Brownian motion and $\nabla \cdot (B_t - G_t) = 0$. Let ν_Z be the invariant measure of this dynamics, $\mathcal{L} = (B_t - G_t) \cdot \nabla + \Delta$ its infinitesimal generator, and $\mathcal{L}^* = -\nabla \cdot (B_t - G_t) + \Delta$ its adjoint in $L^2(\nu_Z)$. Using Itô calculus, φ_t satisfies:

$$\varphi_0 = \varphi, \quad \partial_t \varphi_t = \Delta \varphi_t + (B_t - G_t) \cdot \nabla \varphi_t \tag{5.16}$$

which is equivalent to

$$\varphi_0 = \varphi, \quad \partial_t \varphi_t = \Delta \varphi_t + \nabla \cdot ((B_t - G_t)\varphi_t).$$

Given the result of Lemma 1, $\pi_t^\xi - 1$ satisfies:

$$\partial_t (\pi_t^\xi - 1) = \Delta (\pi_t^\xi - 1) - \nabla \cdot ((B_t - G_t)(\pi_t^\xi - 1)). \tag{5.17}$$

For a fixed $t > 0$, one as, for all $0 \leq s \leq t$:

$$\frac{d}{ds} \int_{\mathbb{T}^n} \varphi_{t-s} (\pi_s^\xi - 1) = - \int_{\mathbb{T}^n} \mathcal{L} \varphi_{t-s} (\pi_s^\xi - 1) + \int_{\mathbb{T}^n} \varphi_{t-s} \mathcal{L}^* (\pi_s^\xi - 1) = 0.$$

Integrating between $s = 0$ and $s = t$ yields

$$\int_{\mathbb{T}^m} \varphi_t (\pi_0^\xi - 1) = \int_{\mathbb{T}^m} \varphi (\pi_t^\xi - 1), \quad \forall t \geq 0. \tag{5.18}$$

Second, for all $t \geq 0$,

$$\|\varphi_t\|_1 \leq \|\varphi\|_1. \quad (5.19)$$

Indeed, one has on the torus \mathbb{T}^m :

$$\|\varphi_t\|_1 \leq \int_{\mathbb{T}^m} \psi(t, z) \, dz$$

where, for all $t \geq 0$ and $z \in \mathbb{T}^n$, $\psi(t, z) = \mathbb{E}[|\varphi(Z_t)| | Z_0 = z] \geq 0$ satisfies (5.16) with initial condition $\psi(0, \cdot) = |\varphi(\cdot)| \geq 0$ on \mathbb{T}^m . Integrating by parts and using that $\nabla \cdot (B_t - G_t) = 0$ one can check that $\frac{d}{dt} \int_{\mathbb{T}^m} \psi = 0$, so that:

$$\int_{\mathbb{T}^m} \psi(t, z) \, dz = \int_{\mathbb{T}^m} \psi(0, z) \, dz = \|\varphi\|_1, \quad \forall t \geq 0,$$

hence the result.

Step 1: Now let us show that there exists $\mathcal{C} > 0$ such that, for all $t > 0$,

$$\|\varphi_t\|_2^2 \leq (\mathcal{C}t^{-\frac{m}{2}} + 2) \|\varphi\|_1^2.$$

To do so, consider for all $t \geq 0$,

$$\Lambda(t) = \int_{\mathbb{T}^m} |\varphi_t|^2.$$

Since $\nabla \cdot (B_t - G_t) = 0$ one can show from (5.16) that:

$$\Lambda'(t) = -2 \int_{\mathbb{T}^m} |\nabla \varphi_t|^2.$$

Knowing that $\|\varphi_t\|_1 \leq \|\varphi\|_1$ for all time $t \geq 0$, we use the inequality (5.11) given by Lemma 2 to obtain:

$$\Lambda(t) \leq 2\|\varphi\|_1^2 + a \left[-\frac{1}{2} \Lambda'(t) \right]^{\frac{m}{m+2}} \|\varphi\|_1^{\frac{4}{m+2}}.$$

Consider for all $t \geq 0$, $g(t) = \Lambda(t) - 2\alpha$, where $\alpha = \|\varphi\|_1^2$. By construction, g is decreasing on \mathbb{R}^+ . We distinguish between three cases:

(i) Assume that $g(0) \leq 0$. In this case, $g(t) \leq 0$ for all $t \geq 0$ and, for all $t \geq 0$:

$$\|\varphi_t\|_2^2 \leq 2\|\varphi\|_1^2.$$

(ii) Assume that $g(t) > 0$ for all $t \geq 0$. Then:

$$\begin{aligned} g(t) &\leq a\alpha^{\frac{2}{m+2}} \left[-\frac{1}{2}g'(t) \right]^{\frac{m}{m+2}} \Leftrightarrow g(t)^{\frac{m+2}{m}} \leq -\frac{1}{2}a^{\frac{m+2}{m}}\alpha^{\frac{2}{m}}g'(t) \\ &\Leftrightarrow g'(t) \leq -2 \cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}g(t)^{\frac{m+2}{m}} \\ &\Leftrightarrow g'(t)g(t)^{-\frac{m+2}{m}} \leq -2 \cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}} \\ &\Leftrightarrow -\frac{m}{2} \frac{d}{dt} \left(g(t)^{-\frac{2}{m}} \right) \leq -2 \cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}} \\ &\Leftrightarrow \frac{d}{dt} \left(g(t)^{-\frac{2}{m}} \right) \geq \frac{4}{m} \cdot a^{-\frac{m+2}{m}}\alpha^{-\frac{2}{m}}. \end{aligned}$$

Integrating between 0 and t yields:

$$\begin{aligned}
g(t)^{-\frac{2}{m}} &\geq g(0)^{-\frac{2}{m}} + \frac{4}{m} \cdot a^{-\frac{m+2}{m}} \alpha^{-\frac{2}{m}} t \Leftrightarrow g(t)^{\frac{2}{m}} \leq \frac{1}{g(0)^{-\frac{2}{m}} + \frac{4}{m} \cdot a^{-\frac{m+2}{m}} \alpha^{-\frac{2}{m}} t} \\
&\Leftrightarrow g(t)^{\frac{1}{m}} \leq \frac{1}{\left(g(0)^{-\frac{2}{m}} + \frac{4}{m} \cdot a^{-\frac{m+2}{m}} \alpha^{-\frac{2}{m}} t\right)^{\frac{1}{2}}} \\
&\Rightarrow g(t)^{\frac{1}{m}} \leq \frac{1}{\left(\frac{4}{m} \cdot a^{-\frac{m+2}{m}} \alpha^{-\frac{2}{m}} t\right)^{\frac{1}{2}}} \\
&\Leftrightarrow g(t) \leq 2^{-m} a^{\frac{m+2}{2}} m^{\frac{m}{2}} \alpha t^{-\frac{m}{2}}.
\end{aligned}$$

Eventually for all $t \geq 0$:

$$\|\varphi_t\|_2^2 \leq (\mathcal{C}t^{-\frac{m}{2}} + 2) \|\varphi\|_1^2$$

with $\mathcal{C} = 2^{-m} a^{\frac{m+2}{2}} m^{\frac{m}{2}} > 0$.

(iii) Assume that $g(0) > 0$ and let us assume that $t^* > 0$ is the smallest time t such that $g(t^*) \leq 0$. In this case, using the above reasonings, one obtains:

a) For all $t \geq t^*$, $g(t) \leq g(t^*) < 0$ and thus

$$\|\varphi_t\|_2^2 \leq 2\|\varphi\|_1^2.$$

b) For all $t \in [0, t^*[$, $g(t) > 0$ and thus:

$$\|\varphi_t\|_2^2 \leq (\mathcal{C}t^{-\frac{m}{2}} + 2) \|\varphi\|_1^2.$$

Hence, for all $t \geq 0$, $\|\varphi_t\|_2^2 \leq (\mathcal{C}t^{-\frac{m}{2}} + 2) \|\varphi\|_1^2$.

Step 2: Now, for all $t \geq 0$, equation (5.18) yields:

$$\left| \int_{\mathbb{T}^m} \varphi(\pi_t^\xi - 1) \right|^2 = \left| \int_{\mathbb{T}^m} \varphi_t(\pi_0^\xi - 1) \right|^2.$$

Hence, for all $t \geq 0$:

$$\begin{aligned}
\left| \int_{\mathbb{T}^m} \varphi(\pi_t^\xi - 1) \right|^2 &\leq \|\varphi_t\|_2^2 \|\pi_0^\xi - 1\|_2^2 \\
&\leq (\mathcal{C}t^{-\frac{m}{2}} + 2) \|\varphi\|_1^2 \|\pi_0^\xi - 1\|_2^2 \quad (\text{using Inequality (5.19)})
\end{aligned}$$

Since this is true for any function $\varphi \in L^1(\mathbb{T}^m)$, by duality, for all $t \geq 0$:

$$\|\pi_t^\xi - 1\|_\infty \leq \sqrt{(\mathcal{C}t^{-\frac{m}{2}} + 2)} \|\pi_0^\xi - 1\|_2. \quad (5.20)$$

Step 3: Considering the equation satisfied by π_t^ξ given in Lemma 1, with initial condition π_s^ξ with $s \geq 0$, and using inequality (5.20) over the time interval $[s, s+1]$, there exists $\mathcal{K} = \mathcal{K}(m) > 0$ such that:

$$\|\pi_{s+1}^\xi - 1\|_\infty \leq \mathcal{K} \|\pi_s^\xi - 1\|_2.$$

Denote by $H_0^1(\mathbb{T}^n)$ the closure of the space $\mathcal{C}_0^\infty(\mathbb{T}^n)$ of indefinitely differentiable functions with compact support, with respect to the Sobolev norm $\|\cdot\|_{H^1}$. Using the same reasoning as in the proof of Proposition 3, since $\int_{\mathbb{T}^m} (\pi_t^\xi - 1) = 0$, $(\pi_t^\xi - 1)$ belongs in $H_0^1(\mathbb{T}^n)$, and, using equation (5.17) and the Poincaré-Wirtinger inequality (5.14), one has:

$$\|\pi_t^\xi - 1\|_2 \leq \|\pi_0^\xi - 1\|_2 e^{-4\pi^2 t}, \quad \forall t \geq 0. \quad (5.21)$$

Eventually, for all $t \geq 1$:

$$\|\pi_t^\xi - 1\|_\infty \leq \mathcal{K} \|\pi_{t-1}^\xi - 1\|_2 \leq \mathcal{K} e^{-4\pi^2(t-1)} \|\pi_0^\xi - 1\|_2,$$

which concludes the proof with $\mathcal{C} = \mathcal{K} e^{4\pi^2}$. \square

Remark 22. Note that one could use the maximum principle for times $t \in [0, 1]$ in order to replace the right-hand term $\|\pi_0^\xi - 1\|_2$ by the L^∞ -norm $\|\pi_0^\xi - 1\|_\infty$. Indeed, since by Assumption 4, π_0^ξ is continuous on \mathbb{T}^m , one has a uniform bound on $\pi_0^\xi - 1$. Nevertheless, considering an L^2 -bound highlights the fact that the uniform bound at time 0 is not essential to the proof of Proposition 4, which could be useful for possible generalizations to non-bounded state space cases.

5.4 Existence of a stationary measure

In Section 5.4.1 we state and prove preliminary estimates on the invariant probability measures of homogeneous diffusions. We then proceed in Section 5.4.2 to prove Theorem 1, which gives the existence of a stationary state to (5.8) in the general case, where the force \mathcal{F} can be non-conservative. Eventually, one can find in Section 5.4.3 the proof of Proposition 5 where one establishes bounds on the bias of the free energy estimation and on the bias on the computations of averages with respect to μ .

5.4.1 Preliminary estimates for homogeneous diffusions

The next section is concerned with the sensitivity of the equilibrium measure of a diffusion with respect to its drift, when this drift is in L^p for some p . Consider the following process on \mathbb{T}^n , with $n \geq 1$:

$$dX_t = a(X_t)dt + \sqrt{2}dW_t \quad (5.22)$$

with $(W_t)_{t \geq 0}$ a classical n -dimensional Brownian motion on the torus \mathbb{T}^n and $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ for $p \geq 2$ with $p > n$. We refer to [50] for a probabilistic study of this SDE (existence, strong Markov and Feller properties, existence and Hölder continuity of the transition kernel, etc.). In the following we take a PDE point of view, namely we are interested in the existence, uniqueness and properties of a solution ν in $H^1(\mathbb{T}^n)$ such that $\int_{\mathbb{T}^n} \nu(x) dx = 1$ of the following equation:

$$\forall \varphi \in H^1(\mathbb{T}^n), \quad \int_{\mathbb{T}^n} (a(z) \cdot \nabla \varphi(z) \nu(z) - \nabla \varphi(z) \cdot \nabla \nu(z)) dz = 0. \quad (5.23)$$

This implies in particular that $\int_{\mathbb{T}^n} (\mathcal{L}\varphi)\nu = 0$ for all $\varphi \in \mathcal{C}^2(\mathbb{T}^n)$ with \mathcal{L} being the generator of (5.22).

Remark 23. Note that the Sobolev embedding $H^1 \hookrightarrow L^q$ for some q such that $\frac{1}{q} > \frac{1}{2} - \frac{1}{n}$ and the assumption that $p > n$ ensure that the integrals in (5.23) are well defined for all ν, φ in $H^1(\mathbb{T}^n)$ and all $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$.

Proposition 6. *Let $M > 0$ and $p > n \geq 1$ with $p \geq 2$. There exists $C > 0$ which depends solely on M, p and n , such that the following holds. For all $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ such that $\|a\|_{L^p(\mathbb{T}^n)} \leq M$, there exists a unique probability density $\nu_a \in H^1(\mathbb{T}^n)$ that solves (5.23), and which is such that*

$$\|\nu_a\|_\infty + \|1/\nu_a\|_\infty + \|\nu_a\|_{H^1(\mathbb{T}^n)} \leq C.$$

Moreover, if ν_b is the solution of (5.23) with a replaced by $b \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ with $\|b\|_{L^p(\mathbb{T}^n)} \leq M$, then

$$\|\nu_a - \nu_b\|_{L^2(\mathbb{T}^n)} \leq C\|a - b\|_{L^2(\mathbb{T}^n)}.$$

Remark 24. *In the case of a gradient drift $a = -\nabla \mathcal{A}$, the invariant measure ν_a is explicit: for all $z \in \mathbb{T}^n$,*

$$\nu_a(z) = \frac{1}{Z_{\mathcal{A}}} e^{-\mathcal{A}(z)}, \quad Z_{\mathcal{A}} = \int_{\mathbb{T}^n} e^{-\mathcal{A}(z)} dz,$$

and the L^∞ -bound of Proposition 6 amounts to the continuous injection given by Morrey's inequality [17, Theorem IX.12],

$$W^{1,p}(\mathbb{T}^n) \hookrightarrow L^\infty(\mathbb{T}^n), \quad \forall p > n. \quad (5.24)$$

Indeed, if $\mathcal{A} \in W^{1,p}(\mathbb{T}^n)$, then $\mathcal{A} \in L^\infty(\mathbb{T}^n)$ and ν_a is bounded from above and below (and conversely if ν_a is bounded above and below then \mathcal{A} is bounded). In particular, since this injection is false for $p \leq n$, we see that the condition $p > n$ is necessary in Proposition 6.

Proof. Step 1: First assume that $a \in C^\infty(\mathbb{T}^n, \mathbb{R}^n)$. By [25, Theorem 5.11], there exists a Markov process $(\tilde{X}_t)_{t \geq 0}$ on \mathbb{R}^n whose transition probability density is given by the fundamental solution of the equation $\partial_t \tilde{f}_t = -\operatorname{div}(a \tilde{f}_t - \nabla \tilde{f}_t)$, where a is seen as a 1-periodic function on \mathbb{R}^n . Note that by [26, Theorem 0.5 and Condition 0.24.A1], the density \tilde{f}_t is strictly positive and depends continuously on the initial condition. Moreover, [25, Theorems 11.4 and 11.5] yield that $(\tilde{X}_t)_{t \geq 0}$ solves the stochastic differential equation (5.22) on \mathbb{R}^n . Now, consider $(X_t)_{t \geq 0}$ the image of $(\tilde{X}_t)_{t \geq 0}$ by the canonical projection from \mathbb{R}^n to \mathbb{T}^n . Since a is periodic, $(X_t)_{t \geq 0}$ solves (5.22) as an equation on \mathbb{T}^n , and thus, using Itô's formula, it is a Markov process (the proof is the same as [25, Theorems 11.5] in \mathbb{R}^n). Denote by $(P_t)_{t \geq 0}$ the associated Markov semigroup on $L^\infty(\mathbb{T}^n)$. The positivity and continuity in the initial condition of \tilde{f}_t implies that, for all $t > 0$, there exists $r_t > 0$ such that for all $x \in \mathbb{T}^n$ and all Borel set A of \mathbb{T}^n , $\mathbb{P}_x(X_t \in A) \geq r_t \lambda(A)$, namely the process satisfies a uniform Doeblin condition. In particular, for a fixed $t > 0$, the Markov chain with transition operator P_t is recurrent and irreducible and thus, by [65, Theorem 10.0.1], it admits a unique invariant measure ν_a . Now, for $s \geq 0$, $(\nu_a P_s) P_t = (\nu_a P_t) P_s = \nu_a P_s$, which means that $\nu_a P_s$ is an invariant measure for P_t . Hence by uniqueness, $\nu_a P_s = \nu_a$ for all $s \geq 0$. In other words, ν_a is the unique invariant measure for the semigroup $(P_t)_{t \geq 0}$.

Now, let $\varphi \in C^2(\mathbb{T}^n)$. Denoting by $\mathcal{L} = a \cdot \nabla + \Delta$ the infinitesimal generator of (5.22) and using Itô's formula, one gets for all $t \geq 0$

$$0 = \nu_a(P_t(\varphi) - \varphi) = \int_0^t \nu_a P_s \mathcal{L} \varphi ds = t \nu_a(\mathcal{L} \varphi).$$

In other words, ν_a is a solution of the weak equation

$$\forall \varphi \in \mathcal{C}^2(\mathbb{T}^n), \quad \nu_a(\mathcal{L}\varphi) = 0. \quad (5.25)$$

By elliptic regularity (e.g. [42] applied to ν_a seen as a periodic measure on \mathbb{R}^n), ν_a has then a \mathcal{C}^∞ density (that we still denote by ν_a) and, integrating by parts, we can write (5.25) as

$$\int_{\mathbb{T}^n} (a(z) \cdot \nabla \varphi(z) \nu_a(z) - \nabla \varphi(z) \cdot \nabla \nu_a(z)) dz = 0$$

for all $\varphi \in \mathcal{C}^2(\mathbb{T}^n)$ and thus, by density, for all $\varphi \in H^1(\mathbb{T}^n)$. This is (5.23).

Define $\tilde{\nu}_a$ on \mathbb{R}^n by $\tilde{\nu}_a(x+k) = \nu_a(x)$ for all $k \in \mathbb{Z}^n$ and $x \in \mathbb{T}^n$ (seen as $[0, 1]^n$). It is such that

$$\forall \varphi \in H^1(\mathbb{R}^n), \quad \int_{\mathbb{R}^n} (a(z) \cdot \nabla \varphi(z) \tilde{\nu}_a(z) - \nabla \varphi(z) \cdot \nabla \tilde{\nu}_a(z)) dz = 0,$$

where, again, a is seen as a 1-periodic function. Since $p > n$, using the notations of [14] and applying the Harnack inequality [14, Corollary 1.7.2], with the operator $L_{I_n, a, 0}$ (I_n being the identity matrix of size n) and the domain $\Omega = [-1, 2]^n$ which strictly contains $[0, 1]^n$, we get that there exists $C_1 > 0$ depending only on M , p and n such that:

$$\sup_{z \in [0, 1]^n} \tilde{\nu}_a(z) \leq C_1 \inf_{z \in [0, 1]^n} \tilde{\nu}_a(z).$$

Using that $\int_{\mathbb{T}^n} \nu_a = 1$, this implies that

$$1 \leq \sup_{z \in \mathbb{T}^n} \nu_a \leq C_1 \inf_{z \in \mathbb{T}^n} \nu_a \leq C_1. \quad (5.26)$$

Taking $\varphi = \nu_a$ in (5.23) and using the Cauchy-Schwarz inequality yields

$$\int_{\mathbb{T}^n} |\nabla \nu_a|^2 = \int_{\mathbb{T}^n} a \cdot \nabla \nu_a \nu_a \leq \|\nu_a\|_\infty \|a\|_{L^2(\mathbb{T}^n)} \|\nabla \nu_a\|_{L^2(\mathbb{T}^n)},$$

hence $\|\nabla \nu_a\|_{L^2(\mathbb{T}^n)} \leq MC_1$. Consequently, using the Poincaré-Wirtinger inequality (5.14), $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leq C_2$ for some $C_2 > 0$ that depends only on M, p, n .

Step 2: Now we consider $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$, with $\|a\|_{L^p(\mathbb{T}^n)} \leq M$, and proceed to prove the existence of a solution ν_a to equation (5.23). Let $(a_k)_{k \in \mathbb{N}}$ be a sequence of \mathcal{C}^∞ functions that converges to a in $L^p(\mathbb{T}^n)$ and such that $\|a_k\|_{L^p(\mathbb{T}^n)} \leq M$ for all $k \in \mathbb{N}$. Let $(\nu_{a_k})_{k \in \mathbb{N}}$ be the associated solutions of (5.23) given in Step 1. From Step 1, $(\nu_{a_k})_{k \in \mathbb{N}}$ is bounded in $H^1(\mathbb{T}^n)$, and thus we can consider a subsequence that converges weakly in H^1 and strongly in L^2 to some $\nu_a \in H^1(\mathbb{T}^n)$. The weak convergence in H^1 implies that ν_a solves (5.23) and $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leq C_2$. The L^2 -convergence implies that ν_a is a probability density.

Step 3: Let us now consider any solution and establish bounds similar to the previous step and a Poincaré inequality. For $a \in L^p(\mathbb{T}^n, \mathbb{R}^n)$, let $\nu_a \in H^1(\mathbb{T}^n)$ be any probability density solution of (5.23). Using again [14, Corollary 1.7.2] and the fact that the mass of ν_a is 1, we get that $1/C_1 \leq \nu_a \leq C_1$ with the same constant C_1 . From this, as in Step 1, we also get that $\|\nu_a\|_{H^1(\mathbb{T}^n)} \leq C_2$, with the same constant C_2 . The Poincaré-Wirtinger inequality (5.14), together with the lower and upper bounds on ν_a classically yields a Poincaré inequality for ν_a . Indeed,

for any $\varphi \in H^1(\nu_a)$, $\int_{\mathbb{T}^n} \varphi \nu_a$ is the minimizer in \mathbb{R} of $c \mapsto \int_{\mathbb{T}^n} (\varphi - c)^2 \nu_a$, so that

$$\begin{aligned} \int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi \nu_a \right)^2 \nu_a &\leq \int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi \right)^2 \nu_a \leq C_1 \int_{\mathbb{T}^n} \left(\varphi - \int_{\mathbb{T}^n} \varphi \right)^2 \\ &\leq \frac{C_1}{4\pi^2} \int_{\mathbb{T}^n} |\nabla \varphi|^2 \\ &\leq \frac{C_1^2}{4\pi^2} \int_{\mathbb{T}^n} |\nabla \varphi|^2 \nu_a. \end{aligned} \quad (5.27)$$

Step 4: We now proceed to the proof of the last part of the proposition, from which the uniqueness of ν_a immediately follows. Let $a, b \in L^p(\mathbb{T}^n, \mathbb{R}^n)$ with L^p norms less than M and $\nu_a, \nu_b \in H^1(\mathbb{T}^n)$ be probability densities solutions of (5.23) (with respective drift a and b). From the L^∞ -bounds on $\nu_a, 1/\nu_a, \nu_b$ and $1/\nu_b$ obtained in Step 3, we get that ν_b/ν_a and $(\nu_b/\nu_a)^2$ are in $H^1(\mathbb{T}^n)$. Applying (5.23) for b with $\varphi = \nu_b/\nu_a$ as a test function,

$$\begin{aligned} 0 &= \int_{\mathbb{T}^n} \left(b \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b - \nabla \left(\frac{\nu_b}{\nu_a} \right) \cdot \nabla \nu_b \right) \\ &= \int_{\mathbb{T}^n} \left(b \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b - \nabla \left(\frac{\nu_b}{\nu_a} \right) \cdot \nabla \left(\frac{\nu_b}{\nu_a} \nu_a \right) \right) \\ &= \int_{\mathbb{T}^n} \left(b \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b - \left| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right|^2 \nu_a - \frac{1}{2} \nabla \left(\left(\frac{\nu_b}{\nu_a} \right)^2 \right) \cdot \nabla \nu_a \right) \\ &= \int_{\mathbb{T}^n} \left(b \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b - \left| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right|^2 \nu_a - \frac{1}{2} a \cdot \nabla \left(\left(\frac{\nu_b}{\nu_a} \right)^2 \right) \nu_a \right) \\ &= \int_{\mathbb{T}^n} \left(b \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b - \left| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right|^2 \nu_a - a \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b \right) \end{aligned}$$

where the last term of the above equality stems from (5.23) with drift a and test function $\varphi = (\nu_b/\nu_a)^2/2$. As a consequence, using the Cauchy-Schwarz's inequality and the uniform bounds on ν_a and ν_b , one gets:

$$\left\| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right\|_{L^2(\nu_a)}^2 = \int_{\mathbb{T}^n} (b - a) \cdot \nabla \left(\frac{\nu_b}{\nu_a} \right) \nu_b \leq C_1^2 \|b - a\|_{L^2(\mathbb{T}^n)} \left\| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right\|_{L^2(\nu_a)},$$

i.e

$$\left\| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right\|_{L^2(\nu_a)} \leq C_1^2 \|b - a\|_{L^2(\mathbb{T}^n)}. \quad (5.28)$$

Now, since

$$\|\nu_b - \nu_a\|_{L^2(\mathbb{T}^n)}^2 \leq \|\nu_a\|_\infty \int_{\mathbb{T}^n} \left| \frac{\nu_b}{\nu_a} - 1 \right|^2 \nu_a \leq C_1 \left\| \frac{\nu_b}{\nu_a} - 1 \right\|_{L^2(\nu_a)}^2,$$

using the Poincaré inequality (5.27) with $\varphi = \nu_b/\nu_a$ (so that $\int_{\mathbb{T}^n} \varphi \nu_a = 1$) yields:

$$\begin{aligned} \|\nu_b - \nu_a\|_{L^2(\mathbb{T}^n)}^2 &\leq \frac{C_1^3}{4\pi^2} \left\| \nabla \left(\frac{\nu_b}{\nu_a} \right) \right\|_{L^2(\nu_a)}^2 \\ &\leq \frac{C_1^7}{4\pi^2} \|b - a\|_{L^2(\mathbb{T}^n)}^2 \quad (\text{using (5.28)}). \end{aligned}$$

Hence

$$\|\nu_b - \nu_a\|_{L^2(\mathbb{T}^n)} \leq \frac{C_1^{\frac{7}{2}}}{2\pi} \|b - a\|_{L^2(\mathbb{T}^n)}.$$

In the particular case $a = b$, we get that there is only one probability density $\nu \in H^1(\mathbb{T}^n)$ that solves (5.23). \square

5.4.2 Proof of Theorem 1

Proof. Let us recall that one can assume, without loss of generality, that $\beta = 1$ (see the change of variables at the beginning of Section 5.5). From now on, let us fix $p = n + 1$. Consider \mathcal{P}^+ the set of probability densities on \mathbb{T}^n that are lower bounded by a positive constant. Given a probability measure $\pi \in \mathcal{P}^+$, let

$$G_\pi(x) = \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x, y) \frac{\pi(x, y)}{\pi^\xi(x)} dy, \quad \forall x \in \mathbb{T}^m,$$

where $\pi^\xi(\cdot) = \int_{\mathbb{T}^{n-m}} \pi(\cdot, y) dy$. In the ABF case, set $B_\pi = G_\pi$ and, in the PABF case, consider the Helmholtz projection

$$B_\pi = \nabla H_\pi = \mathbf{P}_{L^2(\lambda)}(G_\pi).$$

In both cases, given [3, Lemma 15.13], for all $p \geq 2$, there exists a constant $c^* > 0$ such that,

$$\|B_\pi\|_{L^p(\mathbb{T}^m)} \leq c^* \|G_\pi\|_{L^p(\mathbb{T}^m)} \leq c^* \|\mathcal{F}\|_\infty, \quad (5.29)$$

in other words, for every $\pi \in \mathcal{P}^+$, B_π belongs to the L^p ball $E = \{f \in L^p(\mathbb{T}^m), \|f\|_{L^p(\mathbb{T}^m)} \leq c^* \|\mathcal{F}\|_\infty\}$. In return, given $B \in E$, consider the infinitesimal generator $\mathcal{L}_B = (\mathcal{F} + B) \cdot \nabla + \Delta$ and denote by π_B its invariant measure, such as given in Proposition 6 (in particular $\pi_B \in \mathcal{P}^+$). Composing these two steps, we obtain an application from E to itself,

$$\begin{array}{ccc} T & : & E \longrightarrow E \\ & & f \longmapsto B_{\pi_f} \end{array}.$$

The link with Theorem 1 is that a probability measure π is a stationary state for the non-linear dynamics (5.8) if and only if the associated bias B_π is a fixed point of T . Proving Theorem 1 is thus equivalent to prove that T admits a fixed point. This will be established thanks to the Schauder's fixed point theorem [28, Part 9.2.2 Theorem 3]. One thus have to prove that T is continuous on $(E, \|\cdot\|_{L^p(\mathbb{T}^m)})$ and that the family $T(E) := \{T(B), B \in E\}$ has compact closure in L^p . We have already seen that $T(E) \subset E$, which is a bounded subset of L^p . From the Fréchet-Kolmogorov theorem [17, Theorem IV.25], compactness follows from the following condition :

$$\sup_{z \in \mathbb{R}^n, |z| \leq \delta} \sup_{f \in T(E)} \|\tau_z f - f\|_{L^p(\mathbb{T}^m)} \xrightarrow{\delta \rightarrow 0} 0, \quad (5.30)$$

where τ_z is the translation operator, namely $\tau_z f(x) = f(x+z)$ for all $x \in \mathbb{T}^m$.

Let us recall that from Proposition 6 there exists a constant $C > 0$ such that for all $B \in E$,

$$\|\pi_B\|_{H^1(\mathbb{T}^n)} + \|\pi_B\|_\infty + \|1/\pi_B\|_\infty \leq C \quad (5.31)$$

and for all $B_1, B_2 \in E$,

$$\|\pi_{B_1} - \pi_{B_2}\|_{L^2(\mathbb{T}^n)} \leq C\|B_1 - B_2\|_{L^2(\mathbb{T}^m)}. \quad (5.32)$$

Continuity of T . Let $B_1, B_2 \in E$ and to alleviate notations, denote by $\pi_1 = \pi_{B_1}$, $\pi_2 = \pi_{B_2}$ the associated invariant measures. In both the ABF and PABF cases, using the same arguments as in (5.29) one gets:

$$\|T(B_1) - T(B_2)\|_{L^p(\mathbb{T}^m)} \leq c^* \|G_{\pi_1} - G_{\pi_2}\|_{L^p(\mathbb{T}^m)}.$$

Moreover, relying on inequalities (5.31) and (5.32), one has, for all $x \in \mathbb{T}^m$,

$$\begin{aligned} |G_{\pi_1}(x) - G_{\pi_2}(x)| &\leq \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \left| \frac{\pi_1(x, y)}{\pi_1^\xi(x)} - \frac{\pi_2(x, y)}{\pi_2^\xi(x)} \right| dy \\ &\leq \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \frac{|\pi_1(x, y) - \pi_2(x, y)|}{\pi_1^\xi(x)} + \frac{\pi_2(x, y)|\pi_1^\xi(x) - \pi_2^\xi(x)|}{\pi_1^\xi(x)\pi_2^\xi(x)} dy \\ &\leq \|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi_1(x, y) - \pi_2(x, y)| + |\pi_1^\xi(x) - \pi_2^\xi(x)| dy \\ &\leq 2\|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi_1(x, y) - \pi_2(x, y)| dy \\ &\leq 2\|\mathcal{F}\|_\infty C^3 \|\pi_1 - \pi_2\|_{L^2(\mathbb{T}^n)} \\ &\leq 2\|\mathcal{F}\|_\infty C^4 \|B_1 - B_2\|_{L^2(\mathbb{T}^m)}. \end{aligned}$$

As a consequence, since $p \geq 2$, by Sobolev embedding,

$$\|T(B_1) - T(B_2)\|_{L^p(\mathbb{T}^m)} \leq c^* \|G_{\pi_1} - G_{\pi_2}\|_{L^p(\mathbb{T}^m)} \leq 2c^* \|\mathcal{F}\|_\infty C^4 \|B_1 - B_2\|_{L^p(\mathbb{T}^m)},$$

which proves that T is a Lipschitz function on $(E, \|\cdot\|_{L^p(\mathbb{T}^m)})$.

Remark 25. *In the particular case where $\|\mathcal{F}\|_\infty$ is small enough so that $2c^*\|\mathcal{F}\|_\infty C^4 < 1$, we directly get that T is a contraction of the L^p -norm, which yields the existence and uniqueness of a fixed-point.*

Compactness. Fix $B \in E$ and let $\pi = \pi_B$ to alleviate notations. For $z \in \mathbb{R}^m$, τ_z commutes with the Helmholtz projection so that, using [3, Lemma 15.13],

$$\|\tau_z \mathbf{P}_{L^2(\lambda)}(G_\pi) - \mathbf{P}_{L^2(\lambda)}(G_\pi)\|_{L^p(\mathbb{T}^m)} = \|\mathbf{P}_{L^2(\lambda)}(\tau_z G_\pi - G_\pi)\|_{L^p(\mathbb{T}^m)} \leq c^* \|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^m)}.$$

Hence, in both the ABF and PABF cases, for all $z \in \mathbb{R}^m$,

$$\|\tau_z T(B) - T(B)\|_{L^p(\mathbb{T}^m)} \leq c^* \|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^m)}.$$

Now, for all $x \in \mathbb{T}^m$ and $z \in \mathbb{R}^m$, using the same argument as in the proof of the continuity of

T ,

$$\begin{aligned}
|G_\pi(x+z) - G_\pi(x)| &= \left| \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x+z, y) \frac{\pi(x+z, y)}{\pi^\xi(x+z)} dy - \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x, y) \frac{\pi(x, y)}{\pi^\xi(x)} dy \right| \\
&\leq \left| \int_{\mathbb{T}^{n-m}} (-\mathcal{F}_1(x+z, y) + \mathcal{F}_1(x, y)) \frac{\pi(x+z, y)}{\pi^\xi(x+z)} dy \right| \\
&\quad + \left| \int_{\mathbb{T}^{n-m}} -\mathcal{F}_1(x, y) \left(\frac{\pi(x+z, y)}{\pi^\xi(x+z)} - \frac{\pi(x, y)}{\pi^\xi(x)} \right) dy \right| \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + \|\mathcal{F}\|_\infty \int_{\mathbb{T}^{n-m}} \left| \frac{\pi(x+z, y)}{\pi^\xi(x+z)} - \frac{\pi(x, y)}{\pi^\xi(x)} \right| dy \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^3 \int_{\mathbb{T}^{n-m}} |\pi(x+z, y) - \pi(x, y)| dy \\
&\leq |z| \|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^3 \|\tau_z \pi - \pi\|_{L^2(\mathbb{T}^n)},
\end{aligned}$$

where C stems from (5.31) and (5.32). To bound the last term, write

$$\begin{aligned}
\int_{\mathbb{T}^n} |\pi(x+z, y) - \pi(x, y)|^2 dx dy &= \int_{\mathbb{T}^n} \left| \int_0^1 z \cdot \nabla_x \pi(x + sz, y) ds \right|^2 dx dy \\
&\leq \int_0^1 \int_{\mathbb{T}^n} |z|^2 |\nabla_x \pi(x + sz, y)|^2 dx dy ds \\
&= |z|^2 \|\nabla_x \pi\|_2^2 \\
&\leq |z|^2 \|\nabla \pi\|_2^2.
\end{aligned}$$

As a conclusion, using (5.31):

$$\|\tau_z G_\pi - G_\pi\|_{L^p(\mathbb{T}^n)} \leq |z| (\|\nabla \mathcal{F}\|_\infty + 2\|\mathcal{F}\|_\infty C^4),$$

so that (5.30) holds.

Consequently, there exists an equilibrium measure $\pi_\infty^{\mathcal{F}}$ which is continuous and positive, along with an associated bias $B_\infty^{\mathcal{F}}$. By Proposition 6, one has positive upper and lower bounds on $\pi_\infty^{\mathcal{F}}$ and, relying on the Holley-Stroock perturbation result [5, Proposition 5.1.6], $\pi_\infty^{\mathcal{F}}$ satisfies $LSI(R)$ for some $R > 0$ and the conditional densities $y \mapsto \pi_{\infty, x}^{\mathcal{F}}(y) := \pi_\infty^{\mathcal{F}}(x, y) / \pi_\infty^{\mathcal{F}, \xi}(x)$ satisfy $LSI(\rho)$ with some $\rho > 0$ uniform with respect to $x \in \mathbb{T}^m$. \square

5.4.3 Proof of Proposition 5

Let us conclude Section 5.4 with the proof of Proposition 5.

Proof. Let us consider the PABF algorithm. Again, without loss of generality, we suppose that $\beta = 1$. Fix $V \in \mathcal{C}^2(\mathbb{T}^n)$, and define

$$\mathfrak{F} = \{(\mathcal{F}, \pi_\infty^{\mathcal{F}}) \in \mathcal{C}^1(\mathbb{T}^n, \mathbb{R}^n) \times \mathcal{P}(\mathbb{T}^n) \mid \|\mathcal{F} + \nabla V\|_\infty \leq 1, \pi_\infty^{\mathcal{F}} \text{ stationary state for (5.8)}\}.$$

In particular, for $(\mathcal{F}, \pi_\infty^{\mathcal{F}}) \in \mathfrak{F}$, $\pi_\infty^{\mathcal{F}}$ is the invariant measure of the diffusion (5.22) on \mathbb{T}^n with drift $a = \mathcal{F} + \nabla(H_{\mathcal{F}} \circ \xi)$. Moreover,

$$\begin{aligned}
\|\mathcal{F} + \nabla V\|_\infty \leq 1 &\Rightarrow \|\mathcal{F}\|_\infty \leq 1 + \|\nabla V\|_\infty \\
&\Rightarrow \|G_{\mathcal{F}}\|_\infty \leq 1 + \|\nabla V\|_\infty.
\end{aligned}$$

By [3, Lemma 15.13], for all $p \geq 2$, there exists $c^* > 0$ such that

$$\|\nabla H_{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \leq c^* \|G_{\mathcal{F}}\|_{L^p(\mathbb{T}^m)} \leq c^* (1 + \|\nabla V\|_{\infty}),$$

which yields, by Minkowski's inequality, for all $p \geq 2$

$$\|\mathcal{F} + \nabla(H_{\mathcal{F}} \circ \xi)\|_{L^p(\mathbb{T}^n)} \leq (c^* + 1) (1 + \|\nabla V\|_{\infty}). \quad (5.33)$$

Note on the other hand, that for all $p \geq 2$

$$\|-\nabla V + \nabla(H_{\mathcal{F}} \circ \xi)\|_{L^p(\mathbb{T}^n)} \leq (1 + c^*) \|\nabla V\|_{\infty} + c^*. \quad (5.34)$$

Denote by $\nu_{\mathcal{F}}$ the invariant measure of the diffusion (5.22) on \mathbb{T}^n with drift $a = -\nabla V + \nabla(H_{\mathcal{F}} \circ \xi)$, in other words

$$\nu_{\mathcal{F}}(x, y) = \frac{1}{Z_{\nu_{\mathcal{F}}}} e^{-V(x, y) + H_{\mathcal{F}}(x)}, \quad Z_{\nu_{\mathcal{F}}} = \int_{\mathbb{T}^n} e^{-V(u, v) + H_{\mathcal{F}}(u)} du dv.$$

In the rest of the proof $(\mathcal{F}, \pi_{\infty}^{\mathcal{F}}) \in \mathfrak{F}$ is fixed and we are careful to give bounds which are uniform over \mathfrak{F} . Besides, to alleviate notations, we simply denote by $\pi = \pi_{\infty}^{\mathcal{F}}$, $\nu = \nu_{\mathcal{F}}$, $H = H_{\mathcal{F}}$ and $G = G_{\mathcal{F}}$.

Given the bounds (5.33) and (5.34), one can apply Proposition 6 with a drift a equal to either $\mathcal{F} + \nabla(H_{\mathcal{F}} \circ \xi)$ or $-\nabla V + \nabla(H_{\mathcal{F}} \circ \xi)$, which are both bounded in $L^p(\mathbb{T}^n)$ for all $p \geq 1$ as shown above. As a consequence, there exists a constant $C > 0$ such that for all $(\mathcal{F}, \pi) \in \mathfrak{F}$,

$$\|\nu\|_{\infty} + \|1/\nu\|_{\infty} + \|\nu\|_{H^1(\mathbb{T}^n)} + \|\pi\|_{\infty} + \|1/\pi\|_{\infty} + \|\pi\|_{H^1(\mathbb{T}^n)} \leq C,$$

and

$$\|\pi - \nu\|_{L^2(\mathbb{T}^n)} \leq C \|\mathcal{F} + \nabla V\|_{L^2(\mathbb{T}^n)} \leq C \|\mathcal{F} + \nabla V\|_{\infty}. \quad (5.35)$$

Notice that ν has the same conditional laws (given x) than the Gibbs measure μ , so that

$$\nabla A(x) = \frac{\int_{\mathbb{T}^{n-m}} \nabla_x V(x, y) e^{-V(x, y)} dy}{\int_{\mathbb{T}^{n-m}} e^{-V(x, y)} dy} = \int_{\mathbb{T}^{n-m}} \nabla_x V(x, y) \frac{\nu(x, y)}{\nu^{\xi}(x)} dy.$$

As a consequence,

$$\begin{aligned} |\nabla A(x) - G(x)| &= \left| \int_{\mathbb{T}^{n-m}} \nabla_x V(x, y) \frac{\nu(x, y)}{\nu^{\xi}(x)} dy + \int_{\mathbb{T}^{n-m}} \mathcal{F}_1(x, y) \frac{\pi(x, y)}{\pi^{\xi}(x)} dy \right| \\ &\leq \|\mathcal{F} + \nabla V\|_{\infty} + \|\nabla V\|_{\infty} \int_{\mathbb{T}^{n-m}} \left| \frac{\nu(x, y)}{\nu^{\xi}(x)} - \frac{\pi(x, y)}{\pi^{\xi}(x)} \right| dy. \end{aligned}$$

Using the same argument as in the proof of the continuity of T in Theorem 1 and (5.35),

$$\begin{aligned} \int_{\mathbb{T}^{n-m}} \left| \frac{\nu(x, y)}{\nu^{\xi}(x)} - \frac{\pi(x, y)}{\pi^{\xi}(x)} \right| dy &\leq 2C^3 \int_{\mathbb{T}^{n-m}} |\nu(x, y) - \pi(x, y)| dy \\ &\leq 2C^3 \|\nu - \pi\|_{L^2(\mathbb{T}^n)} \\ &\leq 2C^4 \|\mathcal{F} + \nabla V\|_{\infty}. \end{aligned}$$

We have thus obtained that, uniformly over \mathfrak{F} ,

$$\|\nabla A - G\|_{L^p(\mathbb{T}^m)} \leq (1 + 2\|\nabla V\|_\infty C^4)\|\mathcal{F} + \nabla V\|_\infty.$$

Which, given [3, Lemma 15.13], yields the following:

$$\|\nabla A - \nabla H\|_{L^p(\mathbb{T}^m)} = \|\mathbf{P}_{L^2(\lambda)}(\nabla A - G)\|_{L^p(\mathbb{T}^m)} \leq c^*\|\nabla A - G\|_{L^p(\mathbb{T}^m)} \leq K_V\|\mathcal{F} + \nabla V\|_\infty. \quad (5.36)$$

with $K_V = c^*(1 + 2\|\nabla V\|_\infty C^4)$. This concludes the proof of the first point of Proposition 5. Concerning the second point, first note that

$$\hat{I}_\psi = \frac{\int_{\mathbb{T}^n} \psi(x, y) e^{-H(x)} \pi(x, y) dx dy}{\int_{\mathbb{T}^n} e^{-H(x)} \pi(x, y) dx dy} = \frac{\int_{\mathbb{T}^n} \psi(x, y) e^{-H(x)} \pi(x, y) dx dy}{\int_{\mathbb{T}^m} e^{-H(x)} dx},$$

where we used Proposition 3 to see that since π is a stationary state of (5.8), π^ξ is necessarily the uniform measure on \mathbb{T}^m . Notice that this expression is unchanged if H is replaced by $H + c$ for some constant $c > 0$. As a consequence, for the remainder of the proof and without loss of generality, we suppose that H is normalised so that $\int_{\mathbb{T}^m} e^{-H} = 1$.

Using that

$$\int_{\mathbb{T}^n} \psi d\mu = \frac{Z_\nu}{Z_\mu} \int_{\mathbb{T}^n} \psi e^{-H \circ \xi} d\nu,$$

we are led to

$$\begin{aligned} \left| \int_{\mathbb{T}^n} \psi d\mu - \hat{I}_\psi \right| &= \left| \frac{Z_\nu}{Z_\mu} \int_{\mathbb{T}^n} \psi e^{-H \circ \xi} d\nu - \int_{\mathbb{T}^n} \psi e^{-H \circ \xi} d\pi \right| \\ &\leq \|\psi e^{-H \circ \xi}\|_\infty \left(\left| \frac{Z_\nu}{Z_\mu} - 1 \right| + \|\nu - \pi\|_{L^1(\mathbb{T}^n)} \right) \\ &\leq \|\psi e^{-H \circ \xi}\|_\infty \left(\left| \frac{Z_\nu}{Z_\mu} - 1 \right| + \|\nu - \pi\|_{L^2(\mathbb{T}^n)} \right). \end{aligned} \quad (5.37)$$

Besides,

$$\frac{Z_\nu}{Z_\mu} = \frac{\int_{\mathbb{T}^n} e^{-(V(x,y)-H(x))} dx dy}{\int_{\mathbb{T}^n} e^{-V(x,y)} dx dy} = \frac{\int_{\mathbb{T}^m} e^{-(A(x)-H(x))} dx}{\int_{\mathbb{T}^m} e^{-A(x)} dx}.$$

Again, this expression is unchanged if A is replaced by $A + c$ for some constant $c > 0$. In the remaining of the proof and without loss of generality, we suppose that A is normalized so that $\int_{\mathbb{T}^m} A - H = 0$, keeping in mind that $\exp(-A) \neq 1$. As a consequence, by the Poincaré-Wirtinger inequality [28, Part 5.8.1 Theorem 1], there exists a constant $\bar{K} > 0$ (that depends only on m and p) such that:

$$\|A - H\|_{L^p(\mathbb{T}^m)} \leq \bar{K}\|\nabla A - \nabla H\|_{L^p(\mathbb{T}^m)}.$$

Thus, using (5.36):

$$\begin{aligned} \|A - H\|_{W^{1,p}(\mathbb{T}^m)} &= \|A - H\|_{L^p(\mathbb{T}^m)} + \|\nabla A - \nabla H\|_{L^p(\mathbb{T}^m)} \\ &\leq (\bar{K} + 1)K_V\|\mathcal{F} + \nabla V\|_\infty. \end{aligned}$$

Now, (5.24) yields the existence of $\mathcal{K} > 0$ such that $\|A - H\|_\infty \leq \mathcal{K}\|A - H\|_{W^{1,p}(\mathbb{T}^m)}$, hence

$$\|A - H\|_\infty \leq \tilde{K}_V\|\mathcal{F} + \nabla V\|_\infty,$$

where $\tilde{K}_V := \mathcal{K}(\bar{K} + 1)K_V = \mathcal{K}(\bar{K} + 1)c^*(1 + 2\|\nabla V\|_\infty C^4)$. Then, using that $|e^a - 1| \leq |a|e^{|a|}$ for all $a \in \mathbb{R}$, for all $x \in \mathbb{T}^m$,

$$|e^{-A(x)+H(x)} - 1| \leq \tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty e^{\tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty},$$

so that, using the fact that $\int_{\mathbb{T}^m} e^{-H} = 1$,

$$\begin{aligned} \left| \frac{Z_\nu}{Z_\mu} - 1 \right| &= \left| \frac{\int_{\mathbb{T}^m} e^{-A+H}}{\int_{\mathbb{T}^m} e^{-A}} - 1 \right| \leq \frac{|\int_{\mathbb{T}^m} e^{-A+H} - 1| + |\int_{\mathbb{T}^m} e^{-A} - 1|}{\int_{\mathbb{T}^m} e^{-A}} \\ &\leq \frac{\int_{\mathbb{T}^m} |e^{-A+H} - 1| + \int_{\mathbb{T}^m} e^{-H} |e^{-A+H} - 1|}{\int_{\mathbb{T}^m} e^{-H-|H-A|}} \\ &\leq 2\tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty e^{2\tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty} \\ &\leq 2\tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty e^{2\tilde{K}_V}. \end{aligned}$$

Combining this with (5.35) in (5.37), we have obtained that

$$\begin{aligned} \left| \int_{\mathbb{T}^n} \psi d\mu - \hat{I}_\psi \right| &\leq \frac{\|\psi e^{-H \circ \xi}\|_\infty}{\int_{\mathbb{T}^m} e^{-H}} \left(2\tilde{K}_V e^{2\tilde{K}_V} + C \right) \|\mathcal{F} + \nabla V\|_\infty \\ &\leq \frac{\|\psi e^{-A \circ \xi}\|_\infty}{\int_{\mathbb{T}^m} e^{-A}} e^{2\tilde{K}_V \|\mathcal{F} + \nabla V\|_\infty} \left(2\tilde{K}_V e^{2\tilde{K}_V} + C \right) \|\mathcal{F} + \nabla V\|_\infty \\ &\leq \frac{\|\psi\|_\infty \|e^{-A}\|_\infty}{\int_{\mathbb{T}^m} e^{-A}} e^{2\tilde{K}_V} \left(2\tilde{K}_V e^{2\tilde{K}_V} + C \right) \|\mathcal{F} + \nabla V\|_\infty, \end{aligned}$$

which yields the conclusion. □

5.5 Long-time convergence

In Section 5.5.1 one can find the proof of intermediate results that will prove useful for the proofs of Theorem 2, Theorem 3, and Corollary 1. Said proofs can respectively be found in Section 5.5.2, Section 5.5.3 and Section 5.5.4.

In all this section, to alleviate notations, we will denote by π_∞ (dropping the \mathcal{F} superscript) a stationary measure given by Theorem 1. First and foremost, let us introduce the concept of total entropy and its macroscopic-microscopic decomposition. We define the *total entropy* as:

$$E(t) = \mathcal{H}(\pi_t | \pi_\infty).$$

In the same manner, the entropy between the marginals in $x \in \mathbb{T}^m$ (called *macroscopic entropy* henceforth) is given by:

$$E_M(t) = \mathcal{H}(\pi_t^\xi | \pi_\infty^\xi).$$

Note that accordingly, one can define the *macroscopic Fisher information*:

$$I_M(t) = I(\pi_t^\xi | \pi_\infty^\xi).$$

The entropy between the conditional measures at a given $x \in \mathbb{T}^m$ (called *local entropy* in the following) is:

$$e_m(t, x) = \mathcal{H}(\pi_{t,x} | \pi_{\infty,x}),$$

where $\pi_{t,x}(\cdot) = \frac{\pi_t(x, \cdot)}{\pi_t^\xi(x)}$ and $\pi_{\infty,x}(\cdot) = \frac{\pi_\infty(x, \cdot)}{\pi_\infty^\xi(x)}$. Now, let us introduce the so-called *microscopic entropy*:

$$E_m(t) = \int_{\mathbb{T}^m} e_m(t, x) \pi_t^\xi(x) dx.$$

One has for all $t \geq 0$, $E(t) = E_m(t) + E_M(t)$ (see [57, Lemma 1]).

Note that we have the following bound on the microscopic entropy:

$$\begin{aligned} E_m(t) &= \int_{\mathbb{T}^m} e_m(t, x) \pi_t^\xi(x) dx = \int_{\mathbb{T}^m} \mathcal{H}(\pi_{t,x} | \pi_{\infty,x}) \pi_t^\xi(x) dx \\ &\leq \frac{1}{2\rho} \int_{\mathbb{T}^m} I(\pi_{t,x} | \pi_{\infty,x}) \pi_t^\xi(x) dx \quad (\text{using Theorem 1 (ii)}). \end{aligned}$$

Since $\nabla_y \ln \left(\frac{\pi_{t,x}}{\pi_{\infty,x}} \right) = \nabla_y \ln \left(\frac{\pi_t}{\pi_\infty} \right)$, this leads to

$$E_m(t) \leq \frac{1}{2\rho} \int_{\mathbb{T}^n} \left| \nabla_y \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t. \quad (5.38)$$

5.5.1 Intermediate results

The proofs of both Theorems 2 and 3 will rely on the following intermediate results. Assumptions 4 and 5 are enforced. Here, \mathcal{F} can be either conservative ($\mathcal{F} = -\nabla V$) or not, and $(\pi_\infty, B_\infty, G_\infty)$ denotes a stationary state of (5.8), with R, ρ the corresponding constants given by Theorem 1.

Lemma 3 (Bound on $G_t(x) - G_\infty(x)$). *For all $t \geq 0$ and $x \in \mathbb{T}^m$:*

$$|G_t(x) - G_\infty(x)| \leq M \sqrt{\frac{2}{\rho} e_m(t, x)}.$$

Proof. Note that, given Theorem 1 (ii), since $\pi_{\infty,x}$ satisfies a log-Sobolev inequality with constant $\rho > 0$, it also satisfies a Talagrand inequality with constant ρ . Now, let $x \in \mathbb{T}^m$ and $\nu_x \in \Pi(\pi_{t,x}, \pi_{\infty,x})$ be a coupling measure. Then, one has:

$$\begin{aligned} G_t(x) - G_\infty(x) &= \int_{\mathbb{T}^{n-m}} (-\mathcal{F}_1(x, y) + \mathcal{F}_1(x, y')) \nu_x(dy, dy') \\ &\leq M \int_{\mathbb{T}^{n-m}} |y - y'| \nu_x(dy, dy') \quad (\text{by Assumption 5}) \\ &\leq M \left(\int_{\mathbb{T}^{n-m}} |y - y'|^2 \nu_x(dy, dy') \right)^{\frac{1}{2}}. \end{aligned}$$

Taking the infimum over $\Pi(\pi_{t,x}, \pi_{\infty,x})$ yields:

$$\begin{aligned} G_t(x) - G_\infty(x) &\leq MW_2(\pi_{t,x}, \pi_{\infty,x}) \\ &\leq M\sqrt{\frac{2}{\rho}\mathcal{H}(\pi_{t,x}|\pi_{\infty,x})} \quad (\text{by the Talagrand inequality (3.4)}). \end{aligned}$$

This yields the conclusion, since $\mathcal{H}(\pi_{t,x}|\pi_{\infty,x}) = e_m(t, x)$. \square

Lemma 4 (Total entropy). *One has,*

$$\frac{dE}{dt} = -\int_{\mathbb{T}^n} |\nabla \ln\left(\frac{\pi_t}{\pi_\infty}\right)|^2 \pi_t + \int_{\mathbb{T}^n} (B_t - B_\infty)(x) \cdot \nabla_x \ln\left(\frac{\pi_t}{\pi_\infty}\right) \pi_t.$$

Proof. If \mathcal{L}_t denotes the infinitesimal generator of (5.7) and \mathcal{L}'_t its formal adjoint in $L^2(\mathbb{T}^n)$ then the Fokker-Planck equation (5.8) can be rewritten as follows:

$$\partial_t \pi_t = \mathcal{L}'_t(\pi_t).$$

Denote by $\mathcal{L}_\infty = \mathcal{F} \cdot \nabla + B_\infty \cdot \nabla_x + \Delta$ the infinitesimal generator associated to the stationary state (π_∞, B_∞) . One has:

$$\begin{aligned} \frac{dE}{dt} &= \int_{\mathbb{T}^n} \partial_t \pi_t + \int_{\mathbb{T}^n} \partial_t \pi_t \ln\left(\frac{\pi_t}{\pi_\infty}\right) \\ &= \int_{\mathbb{T}^n} \mathcal{L}'_t(\pi_t) \ln\left(\frac{\pi_t}{\pi_\infty}\right) \quad (\text{since } \int_{\mathbb{T}^n} \partial_t \pi_t = 0) \\ &= \int_{\mathbb{T}^n} \mathcal{L}_t \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right) \right) \pi_t \\ &= \int_{\mathbb{T}^n} (\mathcal{L}_\infty + \mathcal{L}_t - \mathcal{L}_\infty) \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right) \right) \pi_t \end{aligned}$$

i.e

$$\frac{dE}{dt} = \int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right) \right) \pi_t + \int_{\mathbb{T}^n} (\mathcal{L}_t - \mathcal{L}_\infty) \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right) \right) \pi_t.$$

Since \mathcal{L}_∞ is the infinitesimal generator of a diffusion, it follows that, for any given functions a and f :

$$\mathcal{L}_\infty(a(f)) = a'(f)\mathcal{L}_\infty(f) + a''(f)|\nabla f|^2,$$

as mentioned in [67, Part 2.3]. Applying this with $a(\cdot) = \ln(\cdot)$ and $f = \frac{\pi_t}{\pi_\infty}$ we respectively obtain:

$$\int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\ln\left(\frac{\pi_t}{\pi_\infty}\right) \right) \pi_t = \int_{\mathbb{T}^n} \left(\frac{\pi_\infty}{\pi_t} \cdot \mathcal{L}_\infty \left(\frac{\pi_t}{\pi_\infty} \right) - \left(\frac{\pi_\infty}{\pi_t} \right)^2 \cdot \left| \nabla \frac{\pi_t}{\pi_\infty} \right|^2 \right) \pi_t$$

$$\begin{aligned} \int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t &= \int_{\mathbb{T}^n} \mathcal{L}_\infty \left(\frac{\pi_t}{\pi_\infty} \right) \pi_\infty - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \\ &= - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \quad (\text{since } \pi_\infty \text{ is invariant for } \mathcal{L}_\infty) \end{aligned}$$

and

$$\int_{\mathbb{T}^n} (\mathcal{L}_t - \mathcal{L}_\infty) \left(\ln \left(\frac{\pi_t}{\pi_\infty} \right) \right) \pi_t = \int_{\mathbb{T}^n} (B_t - B_\infty) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t,$$

which concludes the proof. \square

5.5.2 Proof of Theorem 2

Let us prove the convergence of the ABF and PABF algorithms in the conservative case, namely when $\mathcal{F} = -\nabla V$. In that case $\pi_\infty = \mu_A$ is invariant by (5.8) (recall μ_A is given by (5.5)), with a corresponding $G_\infty = \nabla A$, so that $B_\infty = \nabla A$ in both the ABF and PABF case.

Lemma 5. *In the conservative case ($\mathcal{F} = -\nabla V$ and $\pi_\infty = \mu_A$), for all $t \geq 0$ and $x \in \mathbb{T}^m$:*

$$G_t(x) - \nabla A(x) = \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t(x, y)}{\pi_\infty(x, y)} \right) \frac{\pi_t(x, y)}{\pi_t^\xi(x)} dy - \nabla_x \ln \left(\frac{\pi_t^\xi(x)}{\pi_\infty^\xi(x)} \right).$$

Proof. Knowing that $\pi_\infty^\xi = 1$, one has, for a fixed x in \mathbb{T}^m :

$$\begin{aligned} & \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \frac{\pi_t}{\pi_t^\xi} dy - \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \\ &= \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_t}{\pi_t} \cdot \frac{\pi_t}{\pi_t^\xi} dy - \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_\infty}{\pi_\infty} \cdot \frac{\pi_t}{\pi_t^\xi} dy - \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} + \partial_x \ln(1) \\ &= \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} - \int_{\mathbb{T}^{n-m}} \frac{\nabla_x \pi_\infty}{\pi_\infty} \cdot \frac{\pi_t}{\pi_t^\xi} dy - \frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} \\ &= - \int_{\mathbb{T}^{n-m}} \nabla_x (-V(x, y) + A(x)) \cdot \frac{\pi_t}{\pi_t^\xi} dy \\ &= \int_{\mathbb{T}^{n-m}} \nabla_x V(x, y) \cdot \frac{\pi_t}{\pi_t^\xi} dy - \int_{\mathbb{T}^{n-m}} \nabla A(x) \cdot \frac{\pi_t}{\pi_t^\xi} dy \\ &= G_t(x) - \nabla A(x). \end{aligned}$$

\square

In the following proofs, an integral over \mathbb{T}^n is with respect to $(x, y) \in \mathbb{T}^m \times \mathbb{T}^{n-m}$, an integral over \mathbb{T}^m is with respect to $x \in \mathbb{T}^m$, and an integral over \mathbb{T}^{n-m} is with respect to $y \in \mathbb{T}^{n-m}$.

Proof of Theorem 2.

Step 1: Since for all $t \geq 0$, $E(t) = E_m(t) + E_M(t)$, using (5.10) and Lemma 4, one has:

$$\begin{aligned} \frac{dE_m}{dt} &= - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \int_{\mathbb{T}^n} (B_t - \nabla A) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t + \int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi \\ &= - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \int_{\mathbb{T}^n} (G_t - \nabla A) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t + \int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi + J_t \end{aligned}$$

where

$$J_t := \int_{\mathbb{T}^n} (B_t - G_t) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t.$$

Now, using Lemma 5, one gets:

$$\begin{aligned} \frac{dE_m}{dt} &= - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \int_{\mathbb{T}^n} \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \frac{\pi_t}{\pi_t^\xi} dy \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t \\ &\quad - \int_{\mathbb{T}^n} \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t + \int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi + J_t. \end{aligned}$$

On the one hand, using Cauchy-Schwarz's inequality, the first terms in the right-hand side can be bounded as follows:

$$\begin{aligned} \int_{\mathbb{T}^n} \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \frac{\pi_t}{\pi_t^\xi} dy \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t &= \int_{\mathbb{T}^m} \left| \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t \right|^2 \frac{1}{\pi_t^\xi} \\ &= \int_{\mathbb{T}^m} \left| \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t^{\frac{1}{2}} \cdot \pi_t^{\frac{1}{2}} dy \right|^2 \frac{1}{\pi_t^\xi} \\ &\leq \int_{\mathbb{T}^m} \left(\int_{\mathbb{T}^{n-m}} \left| \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t dy \right) \left(\int_{\mathbb{T}^{n-m}} \pi_t dy \right) \frac{1}{\pi_t^\xi} \\ &\leq \int_{\mathbb{T}^n} \left| \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t. \end{aligned}$$

On the other hand, factorising the two next terms in the right-hand side, and using again Lemma 5 gives:

$$\begin{aligned} &- \int_{\mathbb{T}^n} \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t + \int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi \\ &= \int_{\mathbb{T}^m} \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \cdot \left(\nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) - \int_{\mathbb{T}^{n-m}} \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \frac{\pi_t}{\pi_t^\xi} \right) \pi_t^\xi \\ &= \int_{\mathbb{T}^m} \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) (\nabla A - G_t) \pi_t^\xi. \end{aligned}$$

Using once again the Cauchy-Schwarz's inequality, one gets:

$$\frac{dE_m}{dt} \leq - \int_{\mathbb{T}^n} \left| \nabla_y \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \left(\int_{\mathbb{T}^m} |\nabla A - G_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi \right)^{\frac{1}{2}} + J_t.$$

Eventually, recalling that $\int_{\mathbb{T}^m} |\nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right)|^2 \pi_t^\xi = I_M(t)$ and relying on (5.38) and Lemma 3, one has, for all $\varepsilon > 0$ and all $t \geq 0$:

$$\begin{aligned} \frac{dE_m}{dt} &\leq -2\rho E_m(t) + M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)} \sqrt{I_M(t)} + J_t \\ &\leq -2\rho E_m(t) + 2\sqrt{\varepsilon \rho E_m(t)} \sqrt{\frac{M^2}{2\rho^2 \varepsilon} I_M(t)} + J_t \\ &\leq -(2-\varepsilon)\rho E_m(t) + \frac{M^2}{2\rho^2 \varepsilon} I_M(t) + J_t. \end{aligned} \quad (5.39)$$

Step 2: In order to set the idea of the proof, let us first treat the case of the ABF algorithm, where one simply has $J_t = 0$ for all $t \geq 0$. Inequality (5.39) yields, for all $\varepsilon > 0$ and all $t \geq 0$:

$$\frac{dE_m}{dt} \leq -(2-\varepsilon)\rho E_m(t) + \frac{M^2}{2\rho^2 \varepsilon} I_M(t),$$

and using Gronwall's lemma, one has, for all $\varepsilon > 0$ for all $t \geq 0$,

$$E_m(t) \leq E_m(0)e^{-(2-\varepsilon)\rho t} + \frac{M^2}{2\rho^2 \varepsilon} \int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)} ds.$$

Remark 26. Note that in the ABF case [57, Lemma 12] or the PABF case with a Helmholtz projection done with respect to the marginal density π_t^ξ [2, Corollary 1], one has the exponential convergence towards zero of the macroscopic Fisher information $I_M(t)$. This is not the case when one considers the classical Helmholtz projection with respect to the Lebesgue measure: indeed, the density π_t^ξ does not satisfy the heat equation anymore, but an elliptic equation (5.9) with a null-divergence drift. Having no additional information about the regularity of the drift, one cannot prove the convergence of $I_M(t)$ towards zero in the long-time limit as done in [2, 57].

By Proposition 3, for all $t \geq 0$, $E_M(t) \leq E_M(0)e^{-8\pi^2 t}$. Since $I_M(t) = -E'_M(t)$, one gets:

$$0 \leq F(t) := \int_t^\infty I_M(s) ds \leq E_M(t) \leq E_M(0)e^{-8\pi^2 t}, \quad \forall t \geq 0. \quad (5.40)$$

Consequently, relying on (5.40) one has, for all $\varepsilon > 0$, for all $t \geq 0$:

$$\begin{aligned} \int_0^t I_M(s)e^{-(2-\varepsilon)\rho(t-s)} ds &= -e^{-(2-\varepsilon)\rho t} \int_0^t F'(s)e^{(2-\varepsilon)\rho s} ds \\ &= e^{-(2-\varepsilon)\rho t} \left(\int_0^t F(s)(2-\varepsilon)\rho e^{(2-\varepsilon)\rho s} ds - \left[F(s)e^{(2-\varepsilon)\rho s} \right]_0^t \right) \\ &\leq e^{-(2-\varepsilon)\rho t} \left((2-\varepsilon)\rho E_M(0) \int_0^t e^{-(8\pi^2 - (2-\varepsilon)\rho)s} ds - F(t)e^{(2-\varepsilon)\rho t} + F(0) \right) \end{aligned}$$

i.e

$$\begin{aligned} \int_0^t I_M(s) e^{-(2-\varepsilon)\rho(t-s)} ds &\leq e^{-(2-\varepsilon)\rho t} \left((2-\varepsilon)\rho E_M(0) \int_0^t e^{-(8\pi^2-(2-\varepsilon)\rho)s} ds - F(t) e^{(2-\varepsilon)\rho t} + E_M(0) \right) \\ &\leq E_M(0) e^{-(2-\varepsilon)\rho t} \left((2-\varepsilon)\rho \int_0^t e^{-(8\pi^2-(2-\varepsilon)\rho)s} ds + 1 \right) \end{aligned}$$

We distinguish between two case:

(i) If $8\pi^2 = (2-\varepsilon)\rho$, one gets:

$$\int_0^t I_M(s) e^{-(2-\varepsilon)\rho(t-s)} ds \leq E_M(0) e^{-8\pi^2 t} (8\pi^2 t + 1).$$

Since for all $\delta > 0$ and all $t \geq 0$, one has $t \leq \frac{e^{-1}}{\delta} e^{\delta t}$, choosing $\delta = \varepsilon$ yields, for all $t \geq 0$:

$$\int_0^t I_M(s) e^{-(2-\varepsilon)\rho(t-s)} ds \leq E_M(0) \left(\frac{8\pi^2}{e\varepsilon} \vee 1 \right) e^{-(8\pi^2-\varepsilon)t}.$$

(ii) If $8\pi^2 \neq (2-\varepsilon)\rho$, one gets, in all cases ($8\pi^2 > (2-\varepsilon)\rho$ or $8\pi^2 < (2-\varepsilon)\rho$):

$$\int_0^t I_M(s) e^{-(2-\varepsilon)\rho(t-s)} ds \leq E_M(0) \left(\frac{(2-\varepsilon)\rho}{|8\pi^2 - (2-\varepsilon)\rho|} \vee 1 \right) e^{-(8\pi^2 \wedge (2-\varepsilon)\rho)t}.$$

Which yields,

$$\begin{aligned} E_m(t) &\leq E_m(0) e^{-(2-\varepsilon)\rho t} + \frac{M^2}{2\rho^2\varepsilon} \int_0^t I_M(s) e^{-(2-\varepsilon)\rho(t-s)} ds \\ &\leq \left(E_m(0) \vee \frac{M^2}{2\rho^2\varepsilon} E_M(0) \left(\frac{8\pi^2}{e\varepsilon} \vee \frac{(2-\varepsilon)\rho}{|8\pi^2 - (2-\varepsilon)\rho|} \vee 1 \right) \right) e^{-((8\pi^2-\varepsilon) \wedge (2-\varepsilon)\rho)t}. \end{aligned}$$

Conclusion: for the ABF algorithm, we have obtained that for all $\varepsilon > 0$, there exists $\mathcal{K} = \mathcal{K}(\varepsilon) > 0$ such that for all $t \geq 0$,

$$E_m(t) \leq \mathcal{K} e^{-((8\pi^2 \wedge 2\rho) - \varepsilon)t},$$

$$\text{where } \mathcal{K} = \left(E_m(0) \vee \frac{M^2}{2\rho\varepsilon} E_M(0) \left(\frac{8\pi^2\rho}{e\varepsilon} \vee \frac{(2\rho - \varepsilon)}{|8\pi^2 - (2\rho - \varepsilon)|} \vee 1 \right) \right).$$

Step 3: Let us now concentrate on the PABF case, and let us prove an upper bound on J_t . For $t \geq 0$, recall the notation $\nabla H_t := \mathbb{P}_{L^2(\lambda)}(G_t)$, so that $B_t = \nabla H_t$. Similarly, let us introduce, for all $t \geq 0$,

$$\nabla \tilde{H}_t := \mathbb{P}_{L^2(\pi_t^\xi)}(G_t).$$

Recall that $\mathbb{P}_{L^2(\nu)}(f)$ stands for the Helmholtz projection of a vector field f with respect to the measure ν . In the conservative case one has $\pi_\infty \propto e^{-V+A}$, so that $G_\infty = \nabla A$. Since G_∞ is a gradient, one has:

$$\nabla H_\infty = \mathbb{P}_{L^2(\lambda)}(G_\infty) = \nabla A = \mathbb{P}_{L^2(\pi_\infty^\xi)}(G_\infty) = \nabla \tilde{H}_\infty.$$

On the contrary, there is no reason for ∇H_t and $\nabla \tilde{H}_t$ to be equal at a fixed time $t > 0$. Let us

decompose

$$J_t = \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t + \int_{\mathbb{T}^n} (\nabla \tilde{H}_t - G_t) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t.$$

As proven in [2, Lemma 6], relying on the fact that since $\nabla_x \ln(\pi_\infty) = -\nabla(V - A)$, $\pi_\infty^\xi \equiv 1$ and $\nabla \tilde{H}_t = \mathbb{P}_{L^2(\pi_t^\xi)}(G_t)$, one can show that the last right-hand term is negative. One consequently has:

$$\begin{aligned} \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t &= \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln(\pi_t) \pi_t - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln(\pi_\infty) \pi_t \\ &= \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \pi_t^\xi - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln(\pi_\infty) \pi_t \\ &= \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) \cdot \left(\frac{\nabla_x \pi_t^\xi}{\pi_t^\xi} \pi_t^\xi - 0 \times \pi_t^\xi \right) - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln(\pi_\infty) \pi_t \\ &= \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \pi_t^\xi - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \cdot \nabla_x \ln(\pi_\infty) \pi_t. \end{aligned}$$

Hence, in the PABF case,

$$\begin{aligned} J_t &\leq \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \pi_t^\xi - \int_{\mathbb{T}^n} (\nabla H_t - \nabla \tilde{H}_t) \nabla_x \ln(\pi_\infty) \pi_t \\ &= \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \pi_t^\xi - \int_{\mathbb{T}^m} (\nabla H_t - \nabla \tilde{H}_t) (\nabla A - G_t) \pi_t^\xi \\ &\leq \left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \left(\left(\int_{\mathbb{T}^m} \left| \nabla_x \ln \left(\frac{\pi_t^\xi}{\pi_\infty^\xi} \right) \right|^2 \pi_t^\xi \right)^{\frac{1}{2}} + \left(\int_{\mathbb{T}^m} |\nabla A - G_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \right) \\ &\leq \left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \left(\sqrt{I_M(t)} + M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)} \right). \end{aligned} \quad (5.41)$$

Step 4: We will now consider times such that $t \geq 1$. Since $\nabla \tilde{H}_t = \mathbb{P}_{L^2(\pi_t^\xi)}(G_t)$, one has:

$$\int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 \pi_t^\xi = \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi + \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - \nabla H_t|^2 \pi_t^\xi,$$

which yields:

$$\begin{aligned}
\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi &= \int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 \pi_t^\xi - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla H_t - G_t|^2 - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 - \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \left(\left\| 1/\pi_t^\xi \right\|_\infty - 1 \right) \int_{\mathbb{T}^m} |\nabla \tilde{H}_t - G_t|^2 \pi_t^\xi \\
&\leq \|\pi_t^\xi\|_\infty \left(\left\| 1/\pi_t^\xi \right\|_\infty - 1 \right) \int_{\mathbb{T}^m} |G_t|^2 \pi_t^\xi
\end{aligned}$$

in other words,

$$\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi \leq \|\pi_t^\xi\|_\infty \left(\left\| 1/\pi_t^\xi \right\|_\infty - 1 \right) M^2,$$

where we used that, under Assumption 5, $\|G_t\|_\infty \leq \|\nabla_x V\|_\infty \leq M$. Now, from Proposition 4, there exists $C \geq 0$ such that, for all $t \geq 1$:

$$\|\pi_t^\xi\|_\infty \leq 1 + Ce^{-4\pi^2 t} \quad \text{and} \quad \left\| 1/\pi_t^\xi \right\|_\infty \leq 1 + Ce^{-4\pi^2 t}.$$

This yields the existence of a constant $\tilde{C} > 0$ such that, for all $t \geq 1$:

$$\left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \leq \tilde{C} e^{-2\pi^2 t},$$

and, for all $\varepsilon > 0$, for all $t \geq 1$:

$$\begin{aligned}
J_t &\leq \tilde{C} e^{-2\pi^2 t} \left(\sqrt{I_M(t)} + M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)} \right) \\
&\leq \tilde{C} e^{-2\pi^2 t} \left(\sqrt{I_M(t)} + 2 \sqrt{\frac{M^2}{2\rho^2\varepsilon}} \sqrt{\varepsilon \rho E_m(t)} \right) \\
&\leq \varepsilon \rho E_m(t) + I_M(t) + \left(\frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{2\rho^2\varepsilon} \right) e^{-4\pi^2 t}.
\end{aligned}$$

Hence one gets:

$$\frac{dE_m}{dt} \leq -(2 - 2\varepsilon)\rho E_m(t) + K_1 I_M(t) + K_2 e^{-4\pi^2 t}, \quad \forall t \geq 1,$$

with

$$K_1 = K_1(\varepsilon) = 1 + \frac{M^2}{2\rho^2\varepsilon}, \quad K_2 = K_2(\varepsilon) = \frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{2\rho^2\varepsilon}.$$

From now on, let us fix $\varepsilon \in (0, 1)$ and denote by $r_\varepsilon := 2(1 - \varepsilon)$. Using Gronwall's lemma

yields, for all $t \geq 1$:

$$E_m(t) \leq E_m(1)e^{r_\varepsilon \rho} e^{-r_\varepsilon \rho t} + \int_1^t \left(K_1 I_M(s)e^{-r_\varepsilon \rho(t-s)} + K_2 e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} \right) ds.$$

• Let us first consider, for all $t \geq 1$, $I_1 := K_1 \int_1^t I_M(s)e^{-r_\varepsilon \rho(t-s)}$. As done in Step 2, relying on (5.40), one has, for all $t \geq 1$:

$$\begin{aligned} I_1 &= -K_1 e^{-r_\varepsilon \rho t} \int_1^t F'(s)e^{r_\varepsilon \rho s} ds \\ &= K_1 e^{-r_\varepsilon \rho t} \left(r_\varepsilon \rho \int_1^t F(s)e^{r_\varepsilon s} ds - F(t)e^{r_\varepsilon \rho t} + F(1)e^{r_\varepsilon \rho} \right) \\ &\leq K_1 E_M(0)e^{-r_\varepsilon \rho t} \left(r_\varepsilon \rho \int_1^t e^{-(8\pi^2 - r_\varepsilon)s} ds + e^{-(8\pi^2 - r_\varepsilon \rho)} \right). \end{aligned}$$

We distinguish between two cases:

(i) If $8\pi^2 = r_\varepsilon \rho$, one gets, for all $t \geq 1$:

$$I_1 \leq K_1 E_M(0)e^{-8\pi^2 t} (8\pi^2 (t-1) + 1)$$

and, since $(t-1) \leq \frac{e^{-1-\delta}}{\delta} e^{\delta t}$ for all $\delta > 0$, considering $\delta = \varepsilon$, one gets that, for all $t \geq 1$:

$$\begin{aligned} I_1 &\leq K_1 E_M(0)e^{-8\pi^2 t} \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} e^{\varepsilon t} + 1 \right) \\ &\leq K_1 E_M(0) \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} \vee 1 \right) e^{-(8\pi^2 - \varepsilon)t}. \end{aligned}$$

(ii) If $8\pi^2 \neq r_\varepsilon \rho$, one gets, for all $t \geq 1$:

$$I_1 \leq K_1 E_M(0) \left(\frac{r_\varepsilon \rho}{|8\pi^2 - r_\varepsilon \rho|} \vee e^{-(8\pi^2 - r_\varepsilon \rho)} \right) e^{-(8\pi^2 \wedge r_\varepsilon \rho)t}.$$

In any case one has, for all $t \geq 1$

$$I_1 \leq \mathcal{K}_1 e^{-((8\pi^2 - \varepsilon) \wedge r_\varepsilon \rho)t},$$

where $\mathcal{K}_1 = \mathcal{K}_1(\varepsilon) = \left(1 + \frac{M^2}{2\rho^2 \varepsilon} \right) E_M(0) \left(8\pi^2 \frac{e^{-1-\varepsilon}}{\varepsilon} \vee \frac{r_\varepsilon \rho}{|8\pi^2 - r_\varepsilon \rho|} \vee e^{-(8\pi^2 - r_\varepsilon \rho)} \vee 1 \right) > 0$.

• Now consider, for all $t \geq 1$, $I_2 := K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} ds$. We distinguish between two cases:

(i) If $r_\varepsilon \rho \neq 4\pi^2$ then, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} ds \leq \frac{K_2}{|4\pi^2 - r_\varepsilon \rho|} e^{-(4\pi^2 \wedge r_\varepsilon \rho)t}.$$

(ii) If $r_\varepsilon \rho = 4\pi^2$ then, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} ds = K_2 e^{-4\pi^2 t} (t-1),$$

and, since $(t-1) \leq \frac{e^{-1-\delta}}{\delta} e^{\delta t}$ for all $\delta > 0$, considering $\delta = \varepsilon$, one gets that, for all $t \geq 1$:

$$K_2 \int_1^t e^{-4\pi^2 s - r_\varepsilon \rho(t-s)} ds \leq K_2 \frac{e^{-1-\varepsilon}}{\varepsilon} e^{-(4\pi^2 - \varepsilon)t}.$$

In any case one has, for all $t \geq 1$:

$$I_2 \leq \mathcal{K}_2 e^{-((4\pi^2 - \varepsilon) \wedge r_\varepsilon \rho)t},$$

where $\mathcal{K}_2 = \mathcal{K}_2(\varepsilon) = \left(\frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{2\rho^2 \varepsilon} \right) \left(\frac{1}{|4\pi^2 - r_\varepsilon \rho|} \vee \frac{e^{-1-\varepsilon}}{\varepsilon} \right) > 0$.

Hence, recalling that $r_\varepsilon = 2(1 - \varepsilon)$ one gets that for all $\varepsilon > 0$, for all $t \geq 1$,

$$\begin{aligned} E_m(t) &\leq E_m(1) e^{r_\varepsilon \rho} e^{-r_\varepsilon \rho t} + \mathcal{K}_1 e^{-((8\pi^2 - \varepsilon) \wedge r_\varepsilon \rho)t} + \mathcal{K}_2 e^{-((4\pi^2 - \varepsilon) \wedge r_\varepsilon \rho)t} \\ &\leq \mathcal{K}_3 e^{-((4\pi^2 \wedge 2\rho) - \varepsilon)t}, \end{aligned}$$

for some $\mathcal{K}_3 = \mathcal{K}_3(\varepsilon) = \left(E_m(1) e^{2\rho - \varepsilon} \vee \tilde{\mathcal{K}}_1 \vee \tilde{\mathcal{K}}_2 \right) > 0$, where

$$\begin{cases} \tilde{\mathcal{K}}_1 = \left(1 + \frac{M^2}{\rho \varepsilon} \right) E_M(0) \left(\frac{16\pi^2 \rho e^{-(1+\frac{\varepsilon}{2\rho})}}{\varepsilon} \vee \frac{(2\rho - \varepsilon)}{|8\pi^2 - (2\rho - \varepsilon)|} \vee e^{-(8\pi^2 - (2\rho - \varepsilon))} \vee 1 \right) \\ \tilde{\mathcal{K}}_2 = \left(\frac{\tilde{C}^2}{4} + \frac{M^2 \tilde{C}^2}{\rho \varepsilon} \right) \left(\frac{1}{|4\pi^2 - (2\rho - \varepsilon)|} \vee \frac{2\rho e^{-(1+\frac{\varepsilon}{2\rho})}}{\varepsilon} \right) \end{cases}.$$

Step 5: It remains to treat the case where $t \in [0, 1]$. We have:

$$\left(\int_{\mathbb{T}^m} |\nabla H_t - \nabla \tilde{H}_t|^2 \pi_t^\xi \right)^{\frac{1}{2}} \leq \|\pi_t^\xi\|_{L^2(\mathbb{T}^m)}^{\frac{1}{2}} \|\nabla H_t - \nabla \tilde{H}_t\|_{L^4(\mathbb{T}^m)}, \quad \forall t \in [0, 1].$$

From (5.21), there exists $C_2 > 0$ such that for all $t \in [0, 1]$, $\|\pi_t^\xi\|_{L^2(\mathbb{T}^m)}^{\frac{1}{2}} \leq C_2$, and, using [3, Lemma 15.13], there exists $C_4 > 0$ such that for all $t \in [0, 1]$,

$$\|\nabla H_t\|_{L^4(\mathbb{T}^m)} \leq C_4 \|G_t\|_{L^4(\mathbb{T}^m)} \leq C_4 \|\mathcal{F}\|_\infty \leq C_4 \|\nabla V\|_\infty < \infty.$$

Similarly, one has $\|\nabla \tilde{H}_t\|_{L^4(\mathbb{T}^m)} \leq C_4 \|\nabla V\|_\infty$. Hence inequality (5.41) becomes, for all $\varepsilon > 0$

and for all $t \in [0, 1]$:

$$\begin{aligned} J_t &\leq 2C_2C_4\|\nabla V\|_\infty \left(\sqrt{I_M(t)} + M\sqrt{\frac{2}{\rho}}\sqrt{E_m(t)} \right) \\ &\leq 2C_2C_4\|\nabla V\|_\infty \left(\sqrt{I_M(t)} + \sqrt{\frac{2M^2}{\varepsilon\rho^2}}\sqrt{\varepsilon\rho E_m(t)} \right) \\ &\leq \varepsilon\rho E_m(t) + I_M(t) + (C_2C_4\|\nabla V\|_\infty)^2 \left(1 + \frac{2M^2}{\varepsilon\rho^2} \right). \end{aligned}$$

It yields, from inequality (5.39), for all $\varepsilon > 0$ and for all $t \in [0, 1]$:

$$\frac{dE_m}{dt} \leq -r_\varepsilon\rho E_m(t) + K_1I_M(t) + K_2,$$

with

$$K_1 = K_1(\varepsilon) = 1 + \frac{M^2}{2\varepsilon\rho^2}, \quad K_2 = K_2(\varepsilon) = (C_2C_4\|\nabla V\|_\infty)^2 \left(1 + \frac{2M^2}{\varepsilon\rho^2} \right).$$

The Gronwall's lemma yields, for all $\varepsilon > 0$ and for all $t \in [0, 1]$:

$$E_m(t) \leq E_m(0)e^{-r_\varepsilon\rho t} + K_1 \int_0^t I_M(s)e^{-r_\varepsilon\rho(t-s)} ds + K_2 \int_0^t e^{-r_\varepsilon\rho(t-s)} ds,$$

i.e

$$\begin{aligned} E_m(t) &\leq E_m(0) + K_1e^0 \int_0^\infty I_M(s) ds + \frac{K_2}{r_\varepsilon\rho} (1 - e^{-r_\varepsilon\rho t}) \\ &\leq E_m(0) + K_1E_M(0) + \frac{K_2}{r_\varepsilon\rho}, \end{aligned}$$

where we used (5.40). Hence, for all $\varepsilon > 0$ and for all $t \in [0, 1]$

$$E_m(t)e^{((4\pi^2 \wedge 2\rho) - \varepsilon)t} \leq \left(E_m(0) + K_1E_M(0) + \frac{K_2}{r_\varepsilon\rho} \right) e^{((4\pi^2 \wedge 2\rho) - \varepsilon)t} < +\infty.$$

Conclusion: for the PABF algorithm, we have obtained that for all $\varepsilon > 0$, there exists $C = C(\varepsilon) > 0$ such that, for all $t \geq 0$,

$$E_m(t) \leq Ce^{-((4\pi^2 \wedge 2\rho) - \varepsilon)t}.$$

Recall that by Proposition 3, $E_M(t) \leq E_M(0)e^{-8\pi^2 t}$ for all $t \geq 0$. The decomposition $E(t) = E_m(t) + E_M(t)$ concludes the proof. \square

5.5.3 Proof of Theorem 3

Let us prove Theorem 3.

Proof. Using Lemma 4 one gets:

$$\begin{aligned} \frac{dE}{dt} &= - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \int_{\mathbb{T}^n} (B_t - B) \cdot \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \pi_t \\ &\leq - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \left(\int_{\mathbb{T}^m} |B_t - B|^2 \pi_t^\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{T}^n} \left| \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \right)^{\frac{1}{2}}. \end{aligned} \quad (5.42)$$

Step 1: Let us first consider $t \geq 1$. In the PABF case, since an orthogonal projection contracts the corresponding norm, for all $t \geq 1$:

$$\begin{aligned} \int_{\mathbb{T}^m} |\nabla H_t - \nabla H_\infty|^2 \pi_t^\xi &\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |\nabla H_t - \nabla H_\infty|^2 \\ &\leq \|\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \\ &\leq \|\pi_t^\xi\|_\infty \|1/\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \pi_t^\xi \\ &\leq (1 + Ce^{-4\pi^2 t}) \int_{\mathbb{T}^m} |G_t - G_\infty|^2 \pi_t^\xi, \end{aligned}$$

for some $C > 0$ according to Proposition 4. Together with Lemma 3 and the microscopic log-Sobolev inequality (5.38), we have thus obtained for all $t \geq 1$, in both the ABF case (where $B_t = G_t$ and $B_\infty = G_\infty$) and PABF case (where $B_t = \nabla H_t$ and $B_\infty = \nabla H_\infty$),

$$\left(\int_{\mathbb{T}^m} |B_t - B_\infty|^2 \pi_t^\xi \right)^{\frac{1}{2}} \leq \sqrt{1 + Ce^{-4\pi^2 t}} M \sqrt{\frac{2}{\rho}} \sqrt{E_m(t)},$$

in other words,

$$\left(\int_{\mathbb{T}^m} |B_t - B_\infty|^2 \pi_t^\xi \right)^{\frac{1}{2}} \leq \sqrt{1 + Ce^{-4\pi^2 t}} M \sqrt{\frac{2}{\rho}} \frac{1}{\sqrt{2\rho}} \left(\int_{\mathbb{T}^n} \left| \nabla_y \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \right)^{\frac{1}{2}}.$$

As a consequence,

$$\begin{aligned} \frac{dE}{dt} &\leq - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + \frac{M}{\rho} \sqrt{1 + Ce^{-4\pi^2 t}} \left(\int_{\mathbb{T}^n} \left| \nabla_y \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \right)^{\frac{1}{2}} \left(\int_{\mathbb{T}^n} \left| \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \right)^{\frac{1}{2}} \\ &\leq \left(-1 + \frac{M}{2\rho} + C'e^{-2\pi^2 t} \right) \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t. \end{aligned}$$

with $C' = M\sqrt{C}/(2\rho)$. Since we assumed $M < 2\rho$, there exists $t_0 \geq 1$ such that for all $t \geq t_0$, the right hand side is negative:

$$-1 + \frac{M}{2\rho} + C'e^{-2\pi^2 t} := -\alpha(t) \leq 0, \quad \forall t \geq t_0.$$

And, given the logarithmic-Sobolev inequality of constant $R > 0$ satisfied by π_∞ :

$$\frac{dE}{dt} \leq -2\alpha(t)RE(t) \quad \forall t \geq t_0.$$

Hence by Gronwall's lemma, for all $t \geq t_0$:

$$E(t) \leq E(t_0) \exp\left(-2R \int_{t_0}^t \alpha(s) ds\right) = E(t_0) \exp\left(-2R \left(1 - \frac{M}{2\rho}\right) (t - t_0) + \frac{C'R}{2\pi^2}\right)$$

Step 2: As for times $t \in [0, t_0]$, as in the third step of the proof of Theorem 2, there exists $C_2 > 0$ and $C_4 > 0$ such that for all $t \in [0, t_0]$:

$$\left(\int_{\mathbb{T}^m} |B_t - B_\infty|^2 \pi_t^\xi\right)^{\frac{1}{2}} \leq \sqrt{\|\pi_t^\xi\|_2} \|B_t - B_\infty\|_4 \leq 2C_2C_4 \|\mathcal{F}\|_\infty.$$

Inequality (5.42) becomes, for all $t \in [0, t_0]$:

$$\begin{aligned} \frac{dE(t)}{dt} &\leq - \int_{\mathbb{T}^n} \left| \nabla \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t + 2C_2C_4 \|\mathcal{F}\|_\infty \left(\int_{\mathbb{T}^n} \left| \nabla_x \ln \left(\frac{\pi_t}{\pi_\infty} \right) \right|^2 \pi_t \right)^{\frac{1}{2}} \\ &\leq C_2^2 C_4^2 \|\mathcal{F}\|_\infty^2. \end{aligned}$$

Hence, for all $t \in [0, t_0]$

$$E(t) \leq E(0) + (C_2C_4 \|\mathcal{F}\|_\infty)^2 t,$$

and

$$E(t) e^{2R(1-\frac{M}{2\rho})t} \leq \left(E(0) + (C_2C_4 \|\mathcal{F}\|_\infty)^2 t_0\right) e^{2R(1-\frac{M}{2\rho})t_0}$$

which concludes the proof, relying on the same argument as in the proof of Theorem 2. \square

5.5.4 Proof of Corollary 1

Proof. Similarly to the previous proofs, using Lemma 3 and Proposition 4, there exists $C > 0$ such that, for all $t \geq 1$:

$$\begin{aligned} \int_{\mathbb{T}^m} |G_t - G_\infty|^2 dx &\leq \|1/\pi_t^\xi\|_\infty \int_{\mathbb{T}^m} |G_t - G|^2 \pi_t^\xi \\ &\leq (1 + Ce^{-4\pi^2 t}) \frac{2M}{\rho} E_m(t) \\ &\leq (1 + Ce^{-4\pi^2 t}) \frac{2M}{\rho} Ke^{-\Lambda t}, \end{aligned}$$

where we used either Theorem 2 or 3. For $t \in [0, 1]$, we simply bound

$$\int_{\mathbb{T}^m} |G_t - G|^2 dx \leq 2 \|\mathcal{F}\|_\infty^2.$$

This concludes the ABF case, for which $B_t = G_t$ and $B_\infty = G_\infty$. Besides, the L^2 -norm is decreased by the Helmholtz projection, which concludes the PABF case. \square

5.6 What remains to be done

In this Chapter, we proved the long-time convergence of the ABF and PABF methods in the case of non-conservative forces. Both methods are adaptive biasing algorithms: so is the Adaptive

Biasing Potential method introduced in Section 3.3.2. With our current notations, one may ask the following question.

Would the following Adaptive Biasing Potential method:

$$\left\{ \begin{array}{l} dq_t = -\nabla(V - B_t \circ \xi)(q_t)dt + \sqrt{2\beta^{-1}}dW_t \\ \frac{dB_t(z)}{dt} = -\beta^{-1} \ln(\pi_t^\xi(z)), \quad \forall z \in \mathcal{M} \end{array} \right. ,$$

still be robust if the interaction force $\mathcal{F} = -\nabla V$ were to be replaced by a generic, *a priori* non-conservative force?

Furthermore, all three algorithms were designed relying on the overdamped Langevin dynamics (2.4) rather than on the Langevin dynamics (2.2), as it is known to be substantially easier to study. This motivates the need to provide answers to the following open problem.

Can the Adaptive Biasing Force method be used in practice with the Langevin dynamics? In the case of a conservative interaction force $\mathcal{F} = -\nabla V$, does the algorithm converge? If so, how about its behaviour in the case of non-conservative forces?

Part III

Alchemical free energies: the Orthogonal Space Random Walk sampling method

Plût à Dieu que ce fût un usage reçu, & que j'eusse des amis qui me rendissent ce dernier devoir, qui, dis-je, convertissent un jour mes os secs, & épuisés par de longs travaux, en cette substance diaphane, que la plus longue suite de siècles ne sauroit altérer, & qui conserve sa couleur générique, non la verdure des végétaux, mais cependant la couleur de lait du tremblant narcisse ; ce qui pourroit être exécuté en peu d'heures, &c.

JOHANN JOACHIM BECHER, cité dans CENDRES,
*Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des
métiers* eds. Denis Diderot and Jean le Rond d'Alembert

Study of the OSRW sampling method in the case of alchemical transitions

with J. Hémin, L. Lagardère, P. Monmarché and J.-P. Piquemal

Whenever a system undergoes a transition, be it configurational or alchemical, one can define its associated free energy A , which depends on the transition coordinate ξ at hand, as given in (3.9) and (3.11). One can then determine the free energy difference between the initial and final state of the transition. Free energy differences are key quantities used to deduce thermodynamic properties of chemical and biological systems. Many free energy calculation methods have been developed in the scope of molecular dynamics, in order to calculate hydration free energies of small molecules and ions, ligand binding affinities, protein stability or even pK_a predictions, as shown in Section 3.4. However, the currently available methods are limited, in the sense that they cannot *a priori* bypass the system's inherent metastability. This chapter is dedicated to the study of a particular method, called the *Orthogonal Space Random Walk* (OSRW) sampling method, in the scope of alchemical transitions. The OSRW method has two distinct steps: (1) first, it relies on the λ -dynamics method, presented in Section 3.4.3, to model the system's transition from an initial to a final state; (2) second, it uses an enhanced sampling method with a particular choice of reaction coordinates. Step (2) distinguishes the OSRW method from the other well-known enhanced sampling methods. Unfortunately, the OSRW method has to this day not been properly reproduced and tested. In this chapter, we will attempt to build an alternative algorithm to the historical OSRW method.

In Section 6.1, we will first highlight the limitations of the available free energy differences computation methods, taking the FEP and TI methods as an example. We then give a general definition of the OSRW method: we expose its potential limitations to motivate our intuition to rely on the choice of reaction coordinate suggested in [87, 88, 66, 89] and [1], and to use it with the ABF method, which we have shown to be robust. In order to implement our alternative algorithm

to the OSRW method, we need tools to follow steps (1) and (2): we introduce in Section 6.2 the Tinker-HP package and Collective Variables module. We will introduce in Section 6.2.3 our first implementation result, which was to establish a working proxy between the two programs. We then focus on the case of alchemical transitions. We will present in section 6.3, the λ -dynamics method, which has proved to be a powerful tool for ligand binding computations, as we will show in Section 6.3.2. Estimating a ligand's affinity for a given receptor is a long-term goal which motivates the work done in this chapter. However, the classical λ -dynamics method has shown some limitations: we shall state them in Section 6.3.3 and present how we could tackle them. Afterwards, we proceed to detail how we implemented the λ -dynamics in the Tinker-HP package and the Collective Variables Module in Section 6.4. This led us to implement softcore potentials, allowing for new kinds of potential functions to be used in the Tinker-HP software. We will distinguish between the treatment of van der Waals and electrostatic interactions in Sections 6.4.2 and 6.4.3. Eventually, Section 6.4.5 will be dedicated to the numerical results: we will compare free energy differences obtained with the newly implemented λ -dynamics method to free energy differences computed with the Bennett Acceptance Ratio (BAR) method. Section 6.5 is dedicated to the step (2) of the OSRW method, namely the use of the a special pair of reaction coordinates with an ABF method. We first describe the equations of motion of the system at hand in Section 6.5.1 and present how the Tinker-HP and Colvars softwares treat the propagation of the special reaction coordinates in Section 6.5.2. We eventually proceed to the first attempts made to implement the OSRW method in Section 6.5.3: we will give insights into what has been done up until now, the challenges we are currently facing, and how we intend to solve them.

⚠ In this chapter, in order to avoid any misunderstanding, we use a slight change of notations: unless otherwise stated, $\mathbf{r} = (x_1, x_2, x_3) \in \mathbb{R}^3$ is the notation used for a position, \mathbf{p} for a momentum, and q for a charge. Considering a potential V which depends on the distance $r > 0$ between two atoms, we will denote by $\partial_r V = \frac{\partial V}{\partial r}$ its derivative with respect to the distance r , and denote by $\nabla V = (\partial_r V \times \frac{x_1}{r}, \partial_r V \times \frac{x_2}{r}, \partial_r V \times \frac{x_3}{r})^\top$ its Cartesian gradient.

6.1 What is the OSRW method?

6.1.1 Limitations of free energy differences computation methods

Let us consider a transition that is either alchemical or conformational. In both cases, the system's evolution between the initial state Γ_0 and the final state Γ_1 is indexed by an order parameter ξ which evolves from its initial value z_0 to its final value z_1 . In the alchemical case, ξ is a mapping of the extended coordinates (\mathbf{r}, λ) where $\mathbf{r} = (\mathbf{r}_i)_{i \in [1, N]}$ is the position of the system's N particles, with $N \in \mathbb{N}^*$, and λ is a scalar variable between 0 and 1: namely, $\xi(\mathbf{r}, \lambda) = \lambda$, $z_0 = 0$ and $z_1 = 1$. On the other hand, in the conformational case, ξ is a reaction coordinate, in other words, a low dimensional function of the positions \mathbf{r} . Given the transition coordinate ξ , one can define the Helmholtz free energy A of the system and compute its difference between the two end states, namely,

$$\Delta_{\Gamma_0 \rightarrow \Gamma_1} A = A(z_1) - A(z_0). \quad (6.1)$$

Several methods are available to calculate the free energy difference (6.1), such as the Free Energy Perturbation (FEP) method and the Thermodynamic Integration (TI) method. The free energy difference between the initial state Γ_0 and the final state Γ_1 computed with the FEP

methods is given as follows:

$$\Delta G_{\Gamma_0 \rightarrow \Gamma_1} = -\beta^{-1} \ln \langle \exp(-\beta \lambda \Delta_{0 \rightarrow 1} V(r)) \rangle_0, \quad (6.2)$$

where $\Delta_{0 \rightarrow 1} V(r) = V_1(r) - V_0(r)$ is the potential energy fluctuation (V_1 and V_0 being the potential energy of the final and initial state respectively), $\beta = 1/k_B T$ is the thermodynamic beta, and T is the absolute temperature of the system, which is considered constant. The notation $\langle \cdot \rangle$ denotes the average against the Boltzmann-Gibbs distribution $\mu_0 \propto \exp(-\beta V_0)$. One notices that if the fluctuation of the potential energy $\Delta_{0 \rightarrow 1} V$ is large, then basic statistical estimators of (6.2) will suffer from a high variance, leading to a slow convergence. The calculation of the free energy difference can consequently take too much time. Furthermore, it has been pointed out that in certain cases, the FEP method could lead to different results depending on the initial condition considered [49].

On the other hand, for either the conformational or alchemical case, the free energy difference computed with thermodynamic integration then reads:

$$\begin{aligned} \Delta G_{\Gamma_0 \rightarrow \Gamma_1} &= \int_{\xi=z_0}^{\xi=z_1} \left. \frac{dG}{d\xi} \right|_z dz \\ &= \int_{\xi=z_0}^{\xi=z_1} \left\langle \frac{dV_{ext}}{d\xi} - RT \frac{d \ln |J|}{d\xi} \right\rangle_z dz, \end{aligned}$$

where $|J|$ is the Jacobian matrix associated to the parameter ξ and

$$F_\xi := \frac{dV_{ext}}{d\xi} - RT \frac{d \ln |J|}{d\xi}$$

is the generalized force applied on ξ . Note that in the alchemical case, $\xi \equiv \lambda$ is a scalar quantity, and consequently one has $\ln |J| = 0$ and $F_\xi = \frac{dV_{ext}}{d\xi}$.

A key problem of thermodynamic integration is the choice of the thermodynamic path between the two end states. One needs to sample all the intermediate values of ξ between its initial value z_0 and final value z_1 in order to properly sample the values ϕ of the generalized force F_ξ . This choice, which is prior to the simulation, sometimes does not avoid metastability, and the system may remain trapped in potential wells, leading to high computation times. Furthermore, the computation of F_ξ is intrinsically linked to the relaxation of the system after each move of the order parameter ξ : the generalized force sometimes has difficulties reaching the expected region for a given value of ξ . This is the *Hamiltonian lagging* phenomena [72].

As hinted above, there are inherent limitations to the methods used to compute free energy differences, the most important being that metastability in the direction of the reaction coordinate ξ is oftentimes inevitable. As a consequence, one may wish to enhance the sampling of ξ : there are many intuitive ways to do so. One idea to avoid the sampling issues of the FEP method would be to consider ξ as a collective variable, namely a fictitious particle, of fictitious mass m_ξ : this is the idea of the λ -dynamics, introduced by X. Kong, and C. L. Brooks III in 1996 [49, 34, 47]. The λ -dynamics method has led to efficient free energy differences calculations, from ion solvation free energies to protein-ligand binding affinity. Nevertheless, there is *a priori* no reasons for the reaction coordinate ξ to capture *all* of the slow movements of the system and avoid all metastability. Indeed, the space orthogonal to the reaction coordinate space might still contain metastable regions. If this is the case, there is to this day no obvious way to enhance the sampling in said orthogonal space. This motivates the need to define an explicit coordinate that

describes fluctuations in the orthogonal space. An idea, suggested by Wei Yang and co-workers is to use another order parameter $h(\cdot, \xi)$ which is coupled to the evolution of ξ and designed to capture the metastability in the space orthogonal to that of ξ . Then, an enhanced sampling method, called the *Orthogonal Space Random Walk* (OSRW) method, is used to reduce metastability in both the ξ and $h(\cdot, \xi)$ -spaces. However, to our knowledge, such method has led to few applications and is not easily reproducible.

6.1.2 Definition of the OSRW method

6.1.2.1 Adding a second reaction coordinate

The general question of designing good reaction coordinates has always drawn a high interest, in particular with recent advances in machine learning, see Section 3.2.3. In our case, we do not discuss the choice of the initial reaction coordinate ξ , which is supposed to be the coordinate of interest, and we are only interested in the additional coordinates $h(\cdot, \xi)$ used to enhance the sampling of the initial one (the auxiliary variables are *a priori* not meant to contribute to a low-dimensional description of the system, and the free energy associated to them is only estimated as an intermediary step within the adaptive algorithms). As a consequence, one considers the reaction coordinate given by $\tilde{\xi}(\mathbf{r}, \xi) := (\xi(\mathbf{r}), h(\mathbf{r}, \xi))$. In this context, it has been argued [87, 88] that a good candidate for $h(\cdot, \xi)$ (at least in the context of alchemical transformations) is the local mean force F_ξ associated to ξ , which is the vector with components $(F_i)_{i \in \llbracket 1, m \rrbracket}$ given by:

$$F_i = \sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \cdot \nabla V - \beta^{-1} \operatorname{div} \left(\sum_{j=1}^m G_{i,j}^{-1} \nabla \xi_j \right),$$

where $G_{i,j}^{-1}$ denotes the (i, j) -component of the inverse of the Gram matrix G defined by $G = (\nabla \xi)^\top \nabla \xi$, as in Section 3.2.1.

This has the advantage to give a systematic possibility of auxiliary variable, independently from the particular system or reaction coordinate. Moreover, along with the arguments of [87, Equation 3], let us remark the following. Denote by $\tilde{A}(z, \phi)$ the free energy associated to the extended reaction coordinate $\tilde{\xi} = (\xi, F_\xi)$, so that at equilibrium $\tilde{\xi}(x)$ is distributed according to the Gibbs law associated to \tilde{A} . Then, considering a system x at equilibrium, a standard computation shows that

$$\nabla_z A(z) = \mathbb{E} (F_\xi(x) | \xi(x) = z) = \frac{\int \phi e^{-\beta \tilde{A}(z, \phi)} d\phi}{\int e^{-\beta \tilde{A}(z, \phi)} d\phi}.$$

In other words, knowing \tilde{A} , the gradient $\nabla_z A$, and hence A , can be recovered with an exact deterministic integration.

6.1.2.2 Possible limitations

In the literature, the term Orthogonal Space Random Walk refers not only to the choice of F_ξ as an auxiliary variable, but also to some specific adaptive algorithms designed for this particular choice of extended reaction coordinate, involving this deterministic integration. Recall that the

extended Hamiltonian of the system is of the form:

$$H_{ext}(\mathbf{r}, \mathbf{p}; \xi, F_\xi) = E_k + V_{ext}(\mathbf{r}; \xi, F_\xi),$$

where E_k is the total kinetic energy of the extended system, and V_{ext} is the modified potential energy. The idea is then to adaptively bias the extended potential energy $V_{ext}(\mathbf{r}; \xi, F_\xi)$: let us quickly sketch it, keeping in mind the construction of other adaptive biasing methods as exposed in Section 3.3.2. The process considered is given by $(\mathbf{r}_t; \xi_t, F_{\xi,t})_{t \geq 0}$ and at each time $t \geq 0$, one adds two biases f_t and g_t so that the potential energy becomes:

$$V_{ext}(\mathbf{r}_t; \xi_t, F_{\xi,t}) + f_t(\xi_t) + g_t(\xi_t, F_{\xi,t}),$$

with f_t being designed to converge in the long-time limit towards the gradient of the free energy A_0 associated to the unbiased potential energy V_{ext} , and g_t being designed to converge in the long-time limit towards the gradient of the free energy A_1 associated with the potential energy $V_{ext} - A_0 \circ \xi$. The bias g_t can either be defined with metadynamics [88], which is known to not converge, or in a recursive manner [60, 80]. In the latter case, g_t is updated as follows: for a given value z of the collective variable ξ , one first determines the derivative $\delta_\phi g_t(z, \phi)$ for all values ϕ of the second collective variable F_ξ :

$$\delta_\phi g_t(z, \phi) = \mathbb{E} \left[\tilde{F}_z(\mathbf{r}_t) \mid F_{\xi,t} = \phi \right],$$

which is the conditional mean of the local mean force $\tilde{F}_z(\mathbf{r}_t)$ associated to the reaction coordinate $\mathbf{r} \mapsto F_\xi(\mathbf{r})$ with $\xi = z$ being fixed. One then has

$$g_t(z, \phi) = \int_{-\infty}^{\phi} \delta_\phi g_t(z, \tilde{\phi}) d\tilde{\phi} + C_{g_t}(z),$$

where the expression of C_{g_t} depends on g_t and is given by [60, Equation 5]. The bias g_t consequently needs to be differentiable: however, the above construction does not guarantee it. As such, the definition of the biases f_t and g_t does not seem to ensure the proper long-time convergence and well-posedness of the dynamics at hand. As a consequence, these specific adaptive schemes vary from one reference to the other (all called OSRW), and their motivation is not always clear. For this reason, in the remainder of this chapter, the name OSRW only refers to the choice of coordinate $\tilde{\xi} = (\xi, F_\xi)$ (possibly with an extended auxiliary variable). This choice can then be used with any standard sampling algorithm, such as those available in the Colvars module, whose consistency has been abundantly analysed, both empirically and theoretically. One of our initial motivations was to understand whether the specific schemes considered in previous works on OSRW were important or if only the choice of the auxiliary variable mattered. A second question is the following: in the presence of orthogonal metastability, one can expect that many choices of auxiliary coordinates enhance the orthogonal sampling (at the cost of increasing the dimension of the free energy to estimate). Hence, we would like in the near future to compare the choice F_ξ to other possibilities.

As a conclusion of this section, we can now introduce –with general notations– the specific scheme that is the main subject of the present work. We only consider alchemical transformations, in which case $F_\xi = \partial_\lambda V_\lambda$. Moreover, in order to avoid the cumbersome computation of high-order derivatives and to increase the smoothness of the free energy, and hence to reduce the variance of its estimation, we use an extended variable ϕ (only for the auxiliary OSRW coordinate reaction, since the initial reaction coordinate is simply $(\mathbf{r}, \lambda) \mapsto \lambda$, which is already a Cartesian

coordinate on the state). In other words, the state is $x = (\mathbf{r}, \lambda, \phi)$, the energy is $U(x) = V_\lambda(q) + \kappa(\phi - \partial_\lambda V_\lambda(q))^2$ for some $\kappa > 0$ and the extended reaction coordinate is $\xi_2(x) = (\lambda, \phi)$.

6.2 The Tinker-HP and Collective Variables module softwares

6.2.1 Tinker-HP

The Tinker software is a general, modular package for molecular mechanics and molecular dynamics, jointly developed by the Jay Ponder Lab (Department of Chemistry, Washington University in St. Louis, Missouri), the Pengyu Ren Lab (Department of Biomedical Engineering University of Texas in Austin, Texas) and by Jean-Philip Piquemal's research team (Laboratoire de Chimie Théorique, Department of Chemistry, Sorbonne University, Paris). Introduced in the 1990s, it is currently released as version 8 [76], and is mostly written in standard Fortran 95 with OpenMP extensions. Tinker supports a wide range of force fields models including the CHARMM, OPLS, AMBER99 or AMOEBA models. Several options are available when it comes to modelise water. One of the main purposes of the Tinker program is to provide a framework for implementing already existing potential energy functions, and designing new ones. As such, Tinker distinguishes two kind of potential energy terms: the intramolecular and intermolecular terms. Intramolecular energy terms include all energetic terms related to simple motions such as bond stretching or torsional rotation, along with bond potentials. Intermolecular energy terms can be divided into two subclasses. The first one describes the repulsion-dispersion and van der Waals interactions, and there are currently five different van der Waals functional forms available. Among them, we will focus only on the Lennard-Jones 6-12 potential. The second subclass, which is the most complex, describes Coulombic or electrostatic interactions. We shall quickly review the *Ewald summation method* used in the Tinker code in Section 6.4.3.1. We refer to [76] for more insight on the treatment of potential energy terms in the Tinker software.

The canonical Tinker8 software has two other branches: the Tinker-HP code is designed to use MPI-parallel distributed memory supercomputers, whereas the Tinker-OpenMM code is designed to use graphical processing units (GPUs). Tinker-OpenMM's code is based on both the original, canonical Tinker code, and on the OpenMM library: it offers a 200-fold acceleration compared to a regular single core CPU computation, which leads to accurate free energy simulations [52]. We will here focus on the Tinker-HP code, whose main goal is to build a massively parallel version of the Tinker program, in order to obtain a 1000-fold and more speedup of computations. Tinker-HP remains consistent with the Tinker and Tinker-OpenMM codes, while allowing simulations to be run on clusters and on multicore desktop stations. The key asset of the Tinker-HP is that it relies on a strong mathematical background. We refer to [52] for more details on the first version of the code, released in 2017.

The Tinker suite runs on Linux, macOS and Windows, and all of its source code is available on Github (<https://github.com/TinkerTools>). All of the informations on Tinker-HP is available at <http://tinker-hp.ip2ct.upmc.fr/>.

6.2.2 The Collective Variables module

As shown in Section 3.2, transition coordinates, or *collective variables* are key tools to avoid metastability in a MD simulation. In order to do so, one uses an adaptive biasing method as introduced in Section 3.3, which heavily relies on the collective variable at hand. Implementing an enhanced sampling method using a collective variable can turn out to be costly, as one needs

to implement all of the variable functions as well. The Collective Variable module (shortened as "Colvars") introduced by J. Hénin and G. Fiorin in 2013 [30] intends to solve this problem. Colvars is a portable software, written in C++, which allows the design of new collective variables from pre-existing ones in order to combine them with any chosen adaptive biasing algorithm. The module, first implemented within NAMD [73, 11] has been interfaced with the LAMMPS¹ and GROMACS² molecular dynamics codes along with the VMD molecular visualisation program³. There is a wide set of functionals available to be used as collective variables in the Colvars module. Each defined collective variable ξ can be coupled to an extended degree of freedom ϕ , which will then be treated as an extended coordinate of the system at hand. The available collective variables can be divided into six different classes, and we refer to [30] for an exhaustive review. Along with the collective variable functionals, several sampling algorithms are implemented in the module, such as the Metadynamics method, or the Adaptive Biasing Force method, which is at the core of this thesis.

The Collective Variable Module code is available on Github at the following adress: <https://colvars.github.io/>.

6.2.3 The Tinker-HP–Colvars interface

Up to now, there was no possibility for the Tinker code to rely on the Colvars module to run MD simulations using collective variables. One of the first work done in order to test the OSRW method was to establish an interface between the two softwares. In order to interface the Colvars module to the Tinker-HP code, one needs to derive a new class from the `colvarproxy` class⁴. The `colvarproxy.h` file declares the class dealing with the objects needed for any interface. The `colvarproxy` class contains multiple classes which are designed for specific tasks. For example, the `colvarproxy_atoms` class is dedicated to the pre-processing of atomic data, the `colvarproxy_io` class includes methods for input/output processing, and the `colvarproxy_system` class includes methods to access the simulation system and obtain informations about it, including the periodic boundary conditions, the integrator and the force fields that are being used. Until now, the Colvars code was interfaced with three other MD codes: LAMMPS, NAMD and VMD, with the `colvarpoxy_lammps`, `colvarpoxy_namd` and `colvarpoxy_vmd` classes respectively. In the same manner, a `colvarpoxy_Tinker-HP` class has been generated in order to handle the communication between the two codes: at each time step of the MD simulation, the interface is called. One could of course think of establishing communication between the two codes only every n timesteps, $n \in \mathbb{N}^*$ in order to use multi-time-step integration.

¹<https://www.lammps.org/>

²<https://www.gromacs.org/>

³<https://www.ks.uiuc.edu/Research/vmd/>

⁴See <https://colvars.github.io/doxygen/html/index.html> for details on the Colvars module's classes.

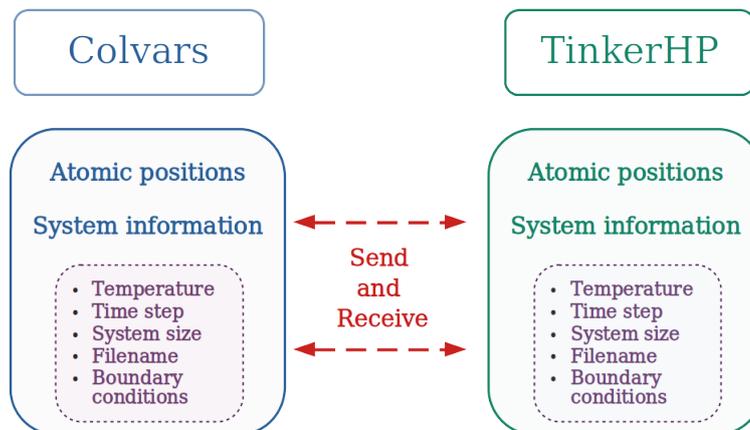


Figure 6.1 – Simple scheme of the communication between the Tinker-HP and Colvars module codes. The interface is responsible for communicating system information at *each time step* of the MD simulation run by Tinker-HP.

6.3 What is λ -dynamics?

We will from now on work only with alchemical transitions, where, let us recall, the system's transition from an initial state to a final state is indexed by a scalar collective variable λ in $[0, 1]$. We refer to Section 3.2.2 for reminders on alchemical transitions. We furthermore assume Assumption 3 of Section 3.2.2 is satisfied.

6.3.1 Some reminders on λ -dynamics

Let us recall the definition of the λ -dynamics method, briefly introduced in Section 3.4.3.1. In order to compute free energy differences in the case of alchemical transition, X. Kong and C. L. Brooks III introduced in 1996 the λ -dynamics method [49]. The key idea of λ -dynamics is to treat the collective variable (i.e, the transition coordinate) λ as a fictitious particle of fictitious mass $m_\lambda > 0$. As a consequence, one can now consider the extended microstate $(\mathbf{r}, \mathbf{p}; \lambda, A) \in \mathcal{T}^* \mathcal{D} \times [0, 1] \times \mathbb{R}$, where $A = m_\lambda \dot{\lambda}$ is the fictitious particle momenta and $\dot{\lambda}$ its time-derivative. One then work with the extended Hamiltonian given by equation (3.24):

$$H_{ext}(\mathbf{r}, \mathbf{p}; \lambda, A) = E_k^x(\mathbf{p}) + E_k^\lambda(A) + V_{ext}(\mathbf{r}; \lambda),$$

where E_k^x is the kinetic energy depending solely on the atomic coordinates, $E_k^\lambda(\lambda) = \frac{1}{2} m_\lambda \dot{\lambda}^2$ is the kinetic energy of the coordinate λ and $V_{ext}(\mathbf{r}; \lambda)$ is the potential energy of the system. The dynamics of the extended system is then given by:

$$\begin{cases} d\mathbf{r}_t = M^{-1} \mathbf{p}_t dt \\ d\mathbf{p}_t = -\nabla V_{ext}(\mathbf{r}_t; \lambda_t) dt - \gamma_1 M^{-1} \mathbf{p}_t dt + \sigma_1 dW_t^{\mathbf{p}} \\ d\lambda_t = m_\lambda^{-1} A_t dt \\ dA_t = -\partial_\lambda V_{ext}(\mathbf{r}_t; \lambda_t) dt - \gamma_2 m_\lambda^{-1} A_t dt + \sigma_2 dW_t^\lambda \end{cases} \quad (6.3)$$

where $(W_t^{\mathbf{P}})_{t \geq 0}$ (resp. $(W_t^\lambda)_{t \geq 0}$) is a dN -dimensional (resp. 1-dimensional) Brownian motion, M is the mass matrix of N original particles, and the pairs of positive constants (γ_1, σ_1) and (γ_2, σ_2) satisfy the fluctuation-dissipation condition (2.3).

6.3.2 A classical application: ligand binding affinity

Before making a first attempt at implementing the λ -dynamics, let us present one of its key applications, that of ligand binding. As highlighted in Section 3.4.1, estimating a given ligand's binding affinity with a receptor is a central problem in pharmacology, and should be one of the long-term tests that our implementation of the λ -dynamics and OSRW methods would undergo. We consequently wish to compare the binding affinities of two ligands L_1 and L_2 with the same receptor R : we put the two ligands in *competition*. Namely, it compares the binding free energy differences $\Delta G_{L_1}^{bind}$ (resp. $\Delta G_{L_2}^{bind}$) of the ligand L_1 (resp. L_2) with the receptor R , by estimating :

$$\Delta_{L_i \rightarrow L_j}(\Delta G^{bind}) = \Delta G_{L_j}^{bind} - \Delta G_{L_i}^{bind} = \Delta_{RL_i \rightarrow RL_j} G^{prot} - \Delta_{L_i \rightarrow L_j} G^{solv}, \quad i, j \in \{1, 2\}, \quad (6.4)$$

where $\Delta_{RL_i \rightarrow RL_j} G^{prot}$ is the binding free energy difference of the complexes RL_i and RL_j , and $\Delta_{L_i \rightarrow L_j} G^{solv}$ is the solvation free energy differences of the ligands L_i and L_j . The free energy differences $\Delta_{RL_i \rightarrow RL_j} G^{prot}$ and $\Delta_{L_i \rightarrow L_j} G^{solv}$ are habitually determined with an alchemical transition, where the ligand L_i is transformed in the ligand L_j , using λ -dynamics. The initial state $\lambda = 0$ thus corresponds to the ligand L_i , and the final state $\lambda = 1$ to the ligand L_j . In Figure 6.2, it is the ligand L_1 which is transformed into the ligand L_2 . We will denote by $\Delta_{L_1 \rightarrow L_2} G$ the free energy difference associated to our alchemical transition, where the ligands are either solvated along with the receptor, or bound to the receptor. The extended potential energy in this simple case is given by

$$V_{ext}(r, \lambda) = (1 - \lambda)V_0(r) + \lambda V_1(r), \quad (6.5)$$

where V_0 (resp. V_1) is the initial (resp. final) potential, where the ligand corresponds to L_1 (resp. L_2 .)

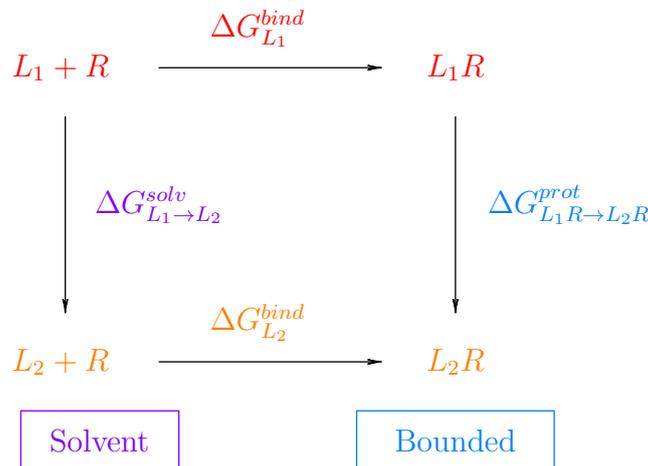


Figure 6.2 – Thermodynamic circle of the binding free energy comparison for two ligands L_1 and L_2 with a given receptor R (here, $i = 1, j = 2$).

By discretising the interval $[0, 1]$ with a discretisation step $\delta\lambda$ whose value is under 2 kcal.mol^{-1} ,

one obtains via the FEP method:

$$\begin{aligned}\Delta_{L_1 \rightarrow L_2} G &= \Delta_{\lambda=0 \rightarrow \lambda=1} G := \Delta_{0 \rightarrow 1} G = -\beta^{-1} \sum_{\lambda=0}^1 \ln \left\langle e^{-\beta(H(r,p;\lambda+\delta\lambda) - H(r,p;\lambda))} \right\rangle_{\lambda} \\ &= -\beta^{-1} \sum_{\lambda=0}^1 \ln \left(\frac{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta\delta\lambda(V_1 - V_0)} e^{-\beta H(r,p;\lambda)} dr dp}{\int_{\mathcal{D} \times \mathbb{R}^{dN}} e^{-\beta H(r,p;\lambda)} dr dp} \right)\end{aligned}$$

However, such a computation may quickly become demanding as the number of ligands put in competition increases, which is a very common situation in drug design. Indeed, one then needs to compare the ligand L_i to *all the other* ligands L_j , $j \neq i$, and run simulations for each discrete value of λ and for each pair $(L_i, L_j)_{i \neq j}$. Using the λ -dynamics method reduces simulation costs, making the target free energy available in just *one simulation*.

One may also rely on the *Generalized Ensemble Thermodynamic Integration* method introduced by Bitetti, Putzer et al.[13], which uses TI instead of the FEP method.

In the case where $L \in \mathbb{N}^*$ ligands are considered, we intend to estimate each binding free energy difference Δ ($\Delta_{L_i \rightarrow L_j} G^{bind}$) of the ligand pairs $(L_i, L_j)_{i,j \in \llbracket 1, L \rrbracket, i \neq j}$. Given equation (6.4), one needs to compute the differences $\Delta_{RL_i \rightarrow RL_j} G^{prot}$ and $\Delta_{L_i \rightarrow L_j} G^{solv}$, with the help of the λ -dynamics. The idea is to consider a set of collective variables $\boldsymbol{\lambda} := (\lambda_i)_{i \in \llbracket 1, L \rrbracket} \in [0, 1]$, so that λ_i scales the potential energy term associated to the ligand L_i for $i \in \llbracket 1, L \rrbracket$. The extended Hamiltonian is given by

$$H_{ext}((\mathbf{R}, \mathbf{r}), (\mathbf{P}, \mathbf{p}); \boldsymbol{\lambda}) = E_k^r + E_k^\lambda + V_{ext}(\mathbf{R}, \mathbf{x}, \boldsymbol{\lambda}),$$

where \mathbf{R} (resp. \mathbf{P}) stands for the Cartesian coordinates (resp. momenta) of the environment, $\mathbf{x} = \{x_i\}_{i \in \llbracket 1, L \rrbracket}$ stands for the Cartesian coordinates of the ligands $(L_i)_{i \in \llbracket 1, L \rrbracket}$, E_k^r and E_k^λ are the kinetic energies of the particles and collective variable respectively. As for the potential energy $V_{ext}(\mathbf{R}, \mathbf{x}, \boldsymbol{\lambda})$, it is given by:

$$V_{ext}(\mathbf{R}, \mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^L \lambda_i (V_i(\mathbf{R}, x_i) - F_i) + V_{env}(\mathbf{R}), \quad (6.6)$$

where V_{env} is the potential energy associated to the environment, and V_i is the potential energy associated to the ligand L_i . The biasing potential F_i associated to the ligand L_i is used to focus the sampling of a specific region of the phase space [47].

Now, denote by P_i the probability that the system is found in the state where the ligand L_i is predominant, for $i \in \llbracket 1, L \rrbracket$, namely if $P_i := \mathbb{P}(\lambda_i = 1, \lambda_j = 0, \forall j \in \llbracket 1, L \rrbracket \setminus \{i\})$. The *relative affinity* of ligands L_i and L_j is then given by:

$$\Delta(\Delta_{L_i \rightarrow L_j} G^{bind}) = -\beta^{-1} \ln \left(\frac{\mathbb{P}(\lambda_j = 1, \lambda_k = 0, \forall k \in \llbracket 1, L \rrbracket \setminus \{j\})}{\mathbb{P}(\lambda_i = 1, \lambda_k = 0, \forall k \in \llbracket 1, L \rrbracket \setminus \{i\})} \right) = -\beta^{-1} \ln \left(\frac{P_j}{P_i} \right). \quad (6.7)$$

The ratio $\frac{P_j}{P_i}$ is determined by estimating the time period during which the collective variables λ_i and λ_j are greater than a given cutoff $\lambda^* \in [0, 1]$. Typically, $\lambda^* \in [0.8, 0.9]$ [47]. Note however that the dependency of the relative affinity Δ ($\Delta_{L_i \rightarrow L_j} G^{bind}$) to the cutoff λ^* will depend on the free energy profile associated to λ_i and λ_j . X. Kong, C. L. Brooks III and colleagues have shown that the λ -dynamics led to a fast selection of the optimal ligand L among a set of competing ligands

for a given receptor [34, 49]. Knight and Brooks [47] observed that if the ligands $(L_i)_{i \in \llbracket 1, L \rrbracket}$ have a relative affinity (6.7) under 2 kcal.mol^{-1} , then the competition is balanced, and the end state $\lambda_i = 1$ is correctly sampled for all $i \in \llbracket 1, L \rrbracket$. On the other hand, a ligand L_k whose relative affinity with other ligands is greater than 3 kcal.mol^{-1} will be ruled out the competition by λ -dynamics, since the end state $\lambda_k = 1$ will never be sampled, in less than several picoseconds.

Remark 27.

- ▷ Note that the computational cost increases with the number of ligands L , and by consequence with the number of collective variables considered. One thus needs to couple the use of λ -dynamics with robust enhanced sampling methods.
- ▷ Until then, λ -dynamics has been used to put ligands in competition for a single binding site on the target receptor. Putting the ligands in competition for several binding sites on the target receptor would introduce λ -dynamics as a competitive method for the structure-activity relationship (SAR) analysis of a given molecule. Knowing the SAR of a molecule enables the determination of the moiety responsible for the biological activity of a molecule after its docking with a receptor. In 2017, R. L. Hayes and collaborators introduced Multisite λ -dynamics method [38], which showed promising results in tackling this question.
- ▷ According to Knight and Brooks [47] there exist two ways to incorporate the molecular topology⁵ in the case of alchemical transitions: either each ligand L_i , $i \in \llbracket 1, L \rrbracket$ is transformed in a single, hybrid reference ligand L^* which is allowed to not make sense physically (single hybrid representation), in which case the topology of L^* should be determined, or each ligand L_i is transformed in another ligand L_j for $i \neq j$ (multiple topology representation), in which case one needs to determine the topology of each ligand explicitly.

There are many other possible applications of the λ -dynamics method, including pK_a prediction, the study of pH-dependent protein folding and protein mutations. We refer to [47] for historical insights on the λ -dynamics' possible applications.

6.3.3 Limitations of the current λ -dynamics

Historically, the extended potential energy was given as in equation (6.5), its alternative expression for ligand binding competition being given by (6.6). Such potential proved to be limited, as highlighted in Paragraph 3.4.3.1, which we recall here. We first note that the TI method would yield a free energy difference given by

$$\Delta_{0 \rightarrow 1} A = -\beta^{-1} \ln \frac{\int_{\mathcal{D}} e^{-\beta V_{ext}(\mathbf{r}; 1)} d\mathbf{r}}{\int_{\mathcal{D}} e^{-\beta V_{ext}(\mathbf{r}; 0)} d\mathbf{r}} = \int_0^1 \langle \partial_\lambda V(\cdot; \lambda) \rangle_\lambda d\lambda,$$

where $\langle \cdot \rangle_\lambda$ denotes the ensemble average with respect to the measure $\mu_\lambda(d\mathbf{r}) \propto \exp(-\beta V_{ext}(\mathbf{r}; \lambda)) d\mathbf{r}$. This expression motivates the need to obtain a sufficiently smooth profile of the potential of mean force (PMF) $\nabla A(\lambda)$, built with the different values of the PMF $(\nabla A(\lambda_i))_{i \in \llbracket 1, n \rrbracket} = (\langle \partial_\lambda V_{ext}(\cdot; \lambda_i) \rangle_\lambda)_{i \in \llbracket 1, n \rrbracket}$ [83]. However in some cases, the system numerically explodes, namely, the simulation is stopped before the end because of an explosion of the system's energy: adding or

⁵By *molecular topology* we mean the representation of a given molecule's topology, where the molecule is composed of K atoms A_1, \dots, A_K . If we assume that for each pair of atoms $(A_i, A_j)_{i \neq j}$, it is possible to know all the bonds between the two atoms, then the topology of the molecule can be determined [35].

subtracting a particle in the system can generate a singularity, and the force $\nabla A(\lambda)$ can explode when the collective variable λ reaches the end states 0 or 1.

A typical example is the solvation of an ion in a box of solvent, when one starts from the initial state "*A: the ion is located at the middle of the box and do not interact with any of the solvent molecules*", to gradually reach the final state "*B: the ion is fully interacting with the neighbouring solvent molecules*". In this specific case, the extended potential energy $V_{ext}(\mathbf{r}; \lambda)$ given by equation (3.25) is naive. As a matter of fact, the potential energy (3.25) does not make the difference between the short-range and long-range interactions: the van der Waals and electrostatics forces are lit up simultaneously as λ evolves between 0 and 1. If in the initial configuration solvent molecules would happen to be too close to the ion, the repulsive term of the van der Waals interactions would make the system to explode. Consequently one would greatly benefit of using a more intricate extended potential.

The idea is then to use *softcore* potentials, whose goal is to obtain finite pair-interaction energies, while insuring that we obtain a sufficiently smooth PMF profile [36, 12, 83]. One can then modify the extended Hamiltonian (3.24) by replacing the potential energy (3.25) with a potential whose dependency in λ is non-linear. Namely, the extended Hamiltonian would then become:

$$V_{ext}(\mathbf{r}; \lambda) := V_{SC-vdW}(\mathbf{r}; \lambda) + V_{SC-e}(\mathbf{r}, q; \lambda), \quad (6.8)$$

where V_{SC-vdW} is the softcore potential term for van der Waals interactions, depending on both the positions of the particles and on λ , whereas $V_{SC-e}(\mathbf{r}, q; \lambda)$ is the softcore potential term for electrostatic interactions, depending on the positions and charges of the particles, along with λ . The associated extended Hamiltonian then reads:

$$H_{ext}(\mathbf{r}, \mathbf{p}; \lambda) = E_k^x(\mathbf{p}) + E_k^\lambda(\lambda) + V_{SC-vdW}(\mathbf{r}; \lambda) + V_{SC-e}(\mathbf{r}, q; \lambda). \quad (6.9)$$

Of course, if one couples the λ -dynamics to a given enhanced sampling method, one can also add biases to the extended Hamiltonian (6.9): this will be treated in a second phase dedicated to the implementation of the OSRW sampling method. For now, the following section is dedicated to the implementation of the λ -dynamics in the Tinker-HP and Colvars module codes, with the use of softcore potentials.

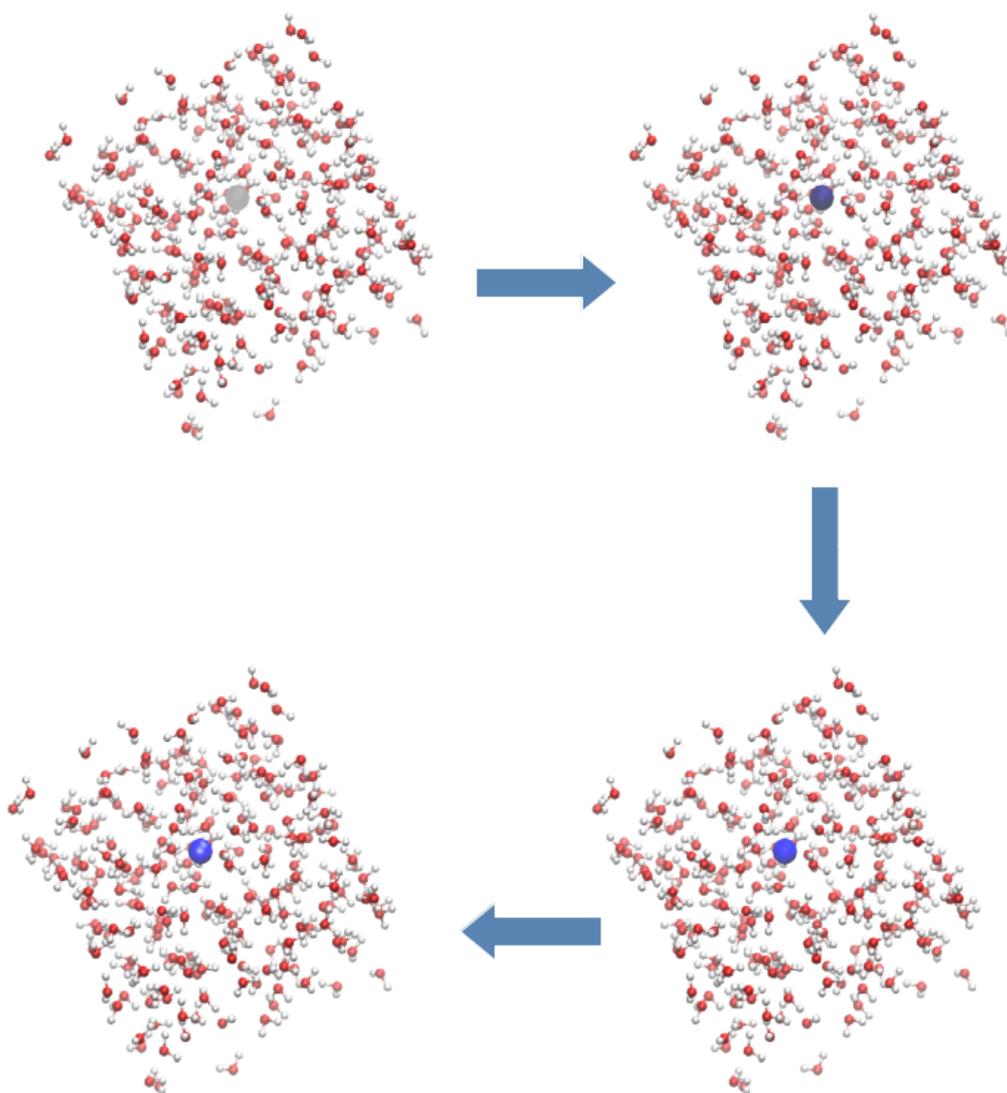


Figure 6.3 – The main idea behind the use of softcore potentials is to gradually light up the van der Waals and electrostatic interactions between the solute and the solvent. Here, a sodium cation is immersed in a cubic box of water. In order to determine its hydration free energy, one wishes to make the system go from an initial state $\lambda = 0$, where the ion does not exist (left upper corner) to a final state $\lambda = 1$, where the ion is fully interacting with the surrounding water molecules (left lower corner). To do so, one relies on λ -dynamics and softcore potentials, and the interactions between the ion and its surrounding are slowly lit up, so that the ion does not brutally interact with the solvent, avoiding numerical instability.

6.4 Implementation of the λ -dynamics with softcore potentials

6.4.1 Foreword

In order to implement the λ -dynamics in the Tinker-HP and Colvars module codes, one first need to treat the limitations of too simple, "hardcore" potentials like (6.5). We proceed to implement softcore potentials –also shortly called *softcores*– to treat the van der Waals and electrostatic interactions. In practice, softcores are built using the classical potentials' expressions. We will in this chapter rely on the softcores suggested by Beutler et al. in [12] and Steinbrecher et al. in [83] for both interactions. In order to do so, we will consider the theoretical, generic potential suggested by T. Halgren in 1992 [36] for the van der Waals interactions, and modify it in order to obtain a softcore potential. As for electrostatic interactions, we will use a softcore potential built on the *Particle Mesh Ewald* method. However, before doing anything, one must first decide how to treat the interactions during the λ -dynamics. Indeed, in the code, the softcore modification of the Lennard-Jones potential's repulsive term will be sufficiently weakened for low values of λ and allow oppositely charged particles to come close to each other. This will lead to numerical instabilities, which can be prevented in several manners [83]. We will focus on the simplest one, which is the *one step method*: the van der Waals and electrostatic interactions are simultaneously modified by introducing softcores for both Halgren and electrostatic potential energies.

6.4.1.1 Treatment of the collective variable λ

In order to use a one-step method, one needs to find a way to treat the dynamical collective variable λ , which evolves in $[0, 1]$. We will define two collective variables: $\lambda_e := \lambda_e(\lambda)$ (resp. $\lambda_v := \lambda_v(\lambda)$) which will light up electrostatic (resp. van der Waals) interactions. We will use two parameters b_e and b_v , whose values are necessarily between 0 and 1, and such that $0 < b_e \leq b_v \leq 1$. The values of b_e and b_v will be set by the user in the input `.key` file of the simulation. We then will define $\lambda_e(\lambda)$ and $\lambda_v(\lambda)$ as follows:

$$\text{for } \lambda \in [0, 1], \quad \begin{cases} \lambda_v(\lambda) = \begin{cases} \frac{1}{b_v}\lambda & \text{if } \lambda \in [0, b_v] \\ 1 & \text{else} \end{cases} \\ \lambda_e(\lambda) = \begin{cases} \frac{1}{1-b_e}\lambda - \frac{b_e}{1-b_e} & \text{if } \lambda \in [b_e, 1] \\ 0 & \text{else} \end{cases} \end{cases} ,$$

The variables λ_e , λ_v and the parameters b_e and b_v are declared in the `mutant` module of the Tinker code. The collective variable λ is a dynamical variable that will be treated by both the Tinker and Colvars codes. The definition of the functions λ_e and λ_v depending on λ is done in the `def_lambdadyn` subroutine of the routine `mutate` in the Tinker code. The default values of b_e and b_v is $b_e = b_v = 0.5$, so that the van der Waals interactions are first entirely lit up, before switching the electrostatic interactions on. Of course, one can light up the electrostatic interactions before the van der Waals interactions are fully lit up: the user can choose the values of b_e and b_v , which are then written as inputs in the `.key` simulation file.

6.4.1.2 Extended potential and list of needed derivatives

Given the equations of motion (6.3), one will need to also implement the softcores' gradient and derivative with respect to λ . In the prospect of implementing the OSRW method, one also needs to obtain the force acting on the potential mean force $F_\lambda(\cdot) := \partial_\lambda V_{ext}(\cdot, \lambda)$ which itself is acting on λ . As a consequence, we shall also derive expressions for the Cartesian gradient $\nabla F_\lambda = \nabla \partial_\lambda V_{ext}(\cdot, \lambda)$ along with the derivative $\partial_\lambda F_\lambda = \partial_{\lambda\lambda}^2 V_{ext}(\cdot, \lambda)$. Given the new extended potential (6.8), we will decompose the second collective variable F_λ needed in the OSRW method into two terms. First notice that

$$\frac{\partial V_{ext}}{\partial \lambda} = \frac{\partial V_{SC-e}}{\partial \lambda} + \frac{\partial V_{SC-vdW}}{\partial \lambda} = \frac{\partial V_{SC-e}}{\partial \lambda_e} \frac{d\lambda_e}{d\lambda} + \frac{\partial V_{SC-vdW}}{\partial \lambda_v} \frac{d\lambda_v}{d\lambda}.$$

Hence the following decomposition:

$$F_\lambda := \frac{\partial V_{ext}}{\partial \lambda} = F_e + F_{vdw},$$

where $F_e := \frac{\partial V_{SC-e}}{\partial \lambda_e} \frac{d\lambda_e}{d\lambda}$ and $F_{vdw} = \frac{\partial V_{SC-vdW}}{\partial \lambda_v} \frac{d\lambda_v}{d\lambda}$.

▲ The following potentials will depend on the *distance* r between the positions \mathbf{r}_i and \mathbf{r}_j of two particles.

6.4.2 Softcores for van der Waals interactions

6.4.2.1 The Halgren potential

Definition – The r^{-6} term of the Lennard-Jones potential (2.1) is physically sensible, and theoretically justified. The r^{-12} term on the other hand, has no proper theoretical justification, and its primary utility is to facilitate computations. Many other models exist for the van der Waals interactions. Among them, one can distinguish the generic potential suggested by T. Halgren [36], which allows to treat a rare gas, and has the following expression:

$$V_{hal}(r) = \varepsilon \left(\frac{1 + \delta}{\frac{r}{r^*} + \delta} \right)^{n-m} \left(\frac{1 + \gamma}{\left(\frac{r}{r^*}\right)^m + \gamma} - 2 \right)$$

where r is the distance between two atoms and ε is the energy associated to the interaction between the pair of atom considered, at their *minimal separation distance* r^* . The *buffer* parameters $(n, m) \in \mathbb{N}^* \times \mathbb{N}^*$, $(\gamma, \delta) \in \mathbb{R}^+ \times \mathbb{R}^+$ are to be chosen by the user. We will from now on use the following form for the Halgren potential energy :

$$V_{hal}(\rho) = \varepsilon \left(\frac{1 + \delta}{\rho + \delta} \right)^{n-m} \left(\frac{1 + \gamma}{\rho^m + \gamma} - 2 \right), \quad (6.10)$$

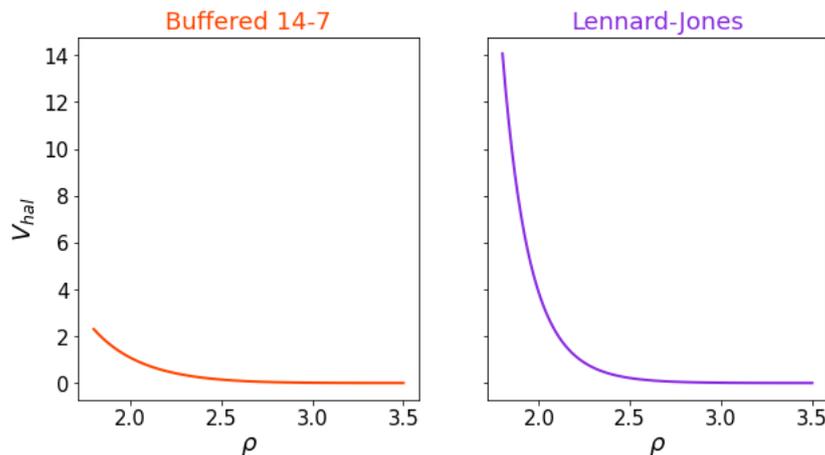
$$\text{where } \rho := \frac{r}{r^*} \text{ and } \frac{d\rho}{dr} = \frac{1}{r^*}.$$

Remark 28. The buffered 14-7 Halgren potential, which corresponds to the set of parameters $n = 14, m = 7, \delta = 0.07$ and $\gamma = 0.12$, has been implemented in Tinker-HP to treat van der Waals interactions in the implementation of the AMOEBA force field model [52, Section 4.2].

Note that the values $n = 12, m = 6$ and $\gamma = \delta = 0$ yield the classical Lennard-Jones potential:

$$V_{LJ}(\rho) = \varepsilon \frac{1}{\rho^6} \left(\frac{1}{\rho^6} - 2 \right) = \varepsilon \left(\frac{1}{\rho^{12}} - \frac{2}{\rho^6} \right).$$

Two classical Halgren potentials



Van der Waals parameter for H (nonpolar): $\sigma = 3.5\text{\AA}$, $\varepsilon = 0.005$

Figure 6.4 – Two classical choices of parameters for the Halgren potential: the buffer 14-7 potential (right) obtained with $n = 14, m = 7, \delta = 0.07$ and $\gamma = 0.12$ and the Lennard-Jones potential (left) obtained with $n = 12, m = 6$ and $\gamma = \delta = 0$. Values for ε and σ were taken for nonpolar hydrogen, and are given in [75, Table 2].

As we do not work with the AMOEBA force field model, but with the AMBER99 model, we will run simulations with values of the Halgren potential's parameters set to $n = 12, m = 6$ and $\gamma = \delta = 0$. In order to implement λ -dynamics, we will need the analytical expressions of the Halgren potential's space derivative, namely $dV_{hal}/d\rho$ and $d^2V_{hal}/d\rho^2$, the derivation of which being done in Section C.1 of Annex C.

6.4.2.2 Generic softcore

Definition – Now, for a given value of λ in $[0, 1]$, one can determine the value of the collective variable $\lambda_v = \lambda_v(\lambda)$, and consider the following, generic softcore potential, similar to that suggested by T. Steinbrecher and co-workers [83]:

$$V_{SC-hal}(r, \lambda_v) = \lambda_v^t V_{hal}(g(r, \lambda_v)),$$

where

$$g(r, \lambda_v) := (\alpha(1 - \lambda_v)^s \sigma^k + r^k)^{\frac{1}{k}}, \quad \text{avec } s, t, k \in \mathbb{N}^*, \alpha, \sigma \in \mathbb{R}^+.$$

With σ such that

$$\sigma 2^{\frac{1}{6}} = r^*.$$

If one considers $\rho = \frac{r}{r^*}$ as done in the Tinker code, the softcore potential then becomes:

$$V_{SC-hal}(\rho, \lambda_v) = \lambda_v^t V_{hal}(g(\rho, \lambda_v)), \quad (6.11)$$

and the softcore function to implement in Tinker then becomes:

$$g(\rho, \lambda_v) := r^* \left(\alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^s + \rho^k \right)^{\frac{1}{k}}, \quad \text{avec } s, t, k \in \mathbb{N}^*, \alpha, \sigma \in \mathbb{R}^+. \quad (6.12)$$

In Tinker, we choose to define the constant α_{LJ} as $\alpha_{LJ} := \alpha 2^{-\frac{k}{6}}$, the parameter α being already defined in the code for the buffered 14-7 Halgren potential. The parameters that will be chosen by the user are α , s , t and k . We shall keep in mind that $\frac{d\rho}{dr} = \frac{1}{r^*}$. The softcore potential (6.11) and the softcore function (6.12) will be implemented in the `elambdalj1c` subroutine of the `elj1` routine of the Tinker code.

Choice of the parameters – The AMBER force field uses the set of parameters $s = t = 1$ and $k = 6$ [83]. Let us first notice that fixing all the parameters equal to one is no good, as it will inevitably lead the system to explode (which we latter on confirmed runing test simulations), as shown in the following plot:

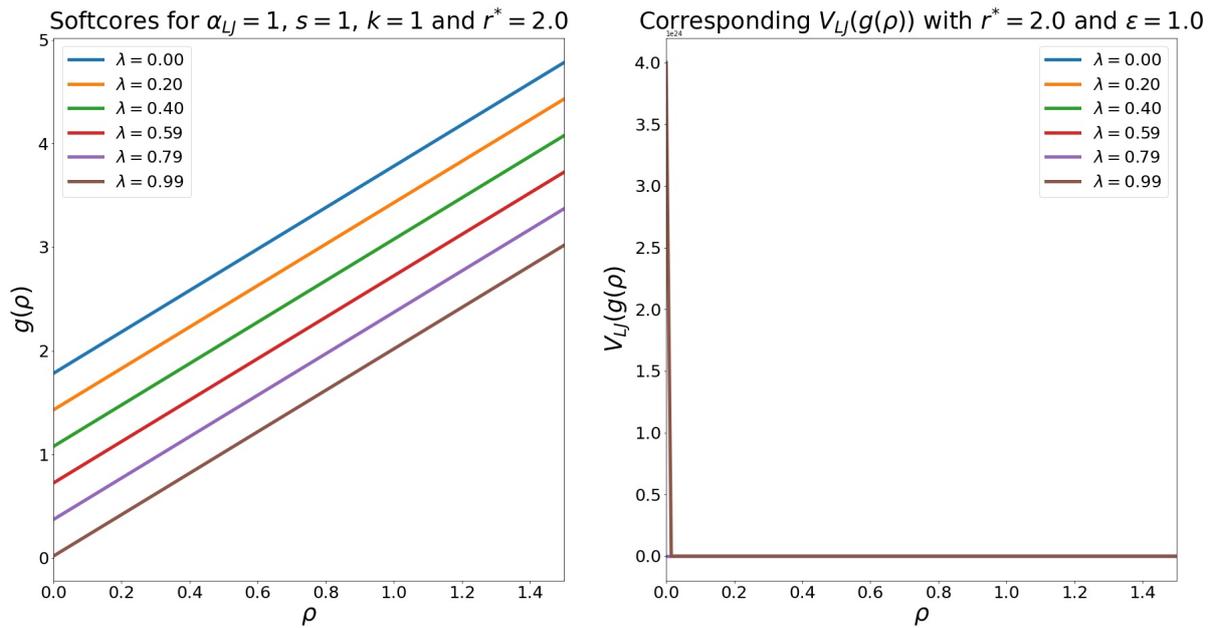


Figure 6.5 – Softcore function and associated modified Lennard-Jones potential with $s = t = k = 1$ and $\alpha = 1$, for λ between 0 and 0.99. Initial value of the softcore potential for $\lambda = 0.99$ is 4×10^{24} .

We refer to the discussion of T. Steinbrecher and his colleagues on the optimal values of α and k provided $s = t = 1$ [83] and point out that T. Beutler et al. suggest that other values of s and t can be chosen [12]. As such, we concluded that the set $\alpha = 0.5$, $s = 2$, $t = 1$ and $k = 6$ was a

good set of parameters to start with.

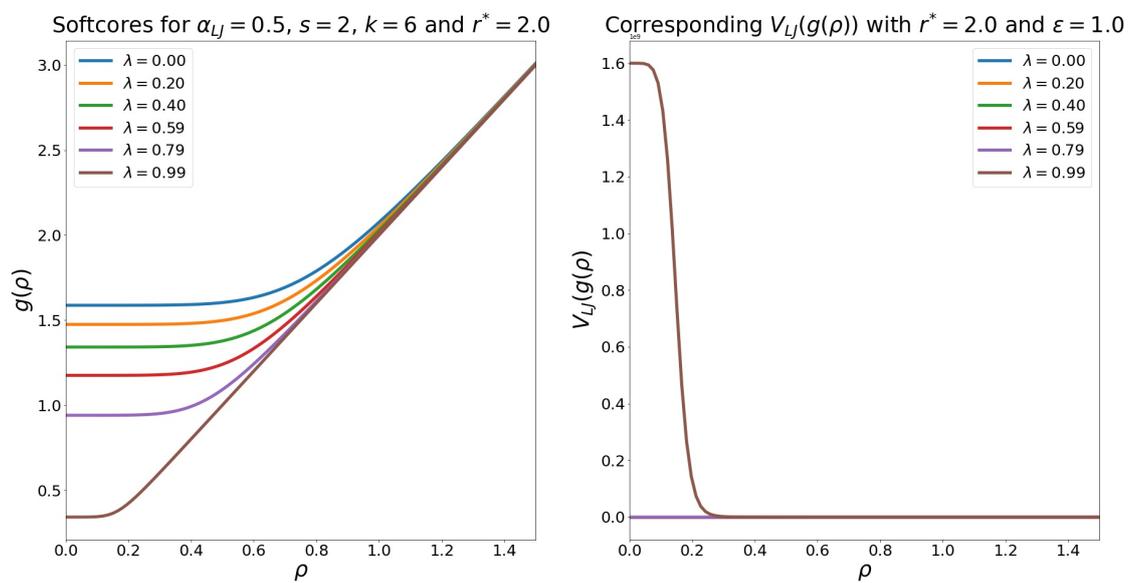


Figure 6.6 – Softcore function and associated modified Lennard-Jones potential with $s = 2$, $t = 1$, $k = 6$ and $\alpha = 1$, for λ between 0 and 0.99. Initial value of the softcore potential for $\lambda = 0.99$ is 1.6×10^9 .

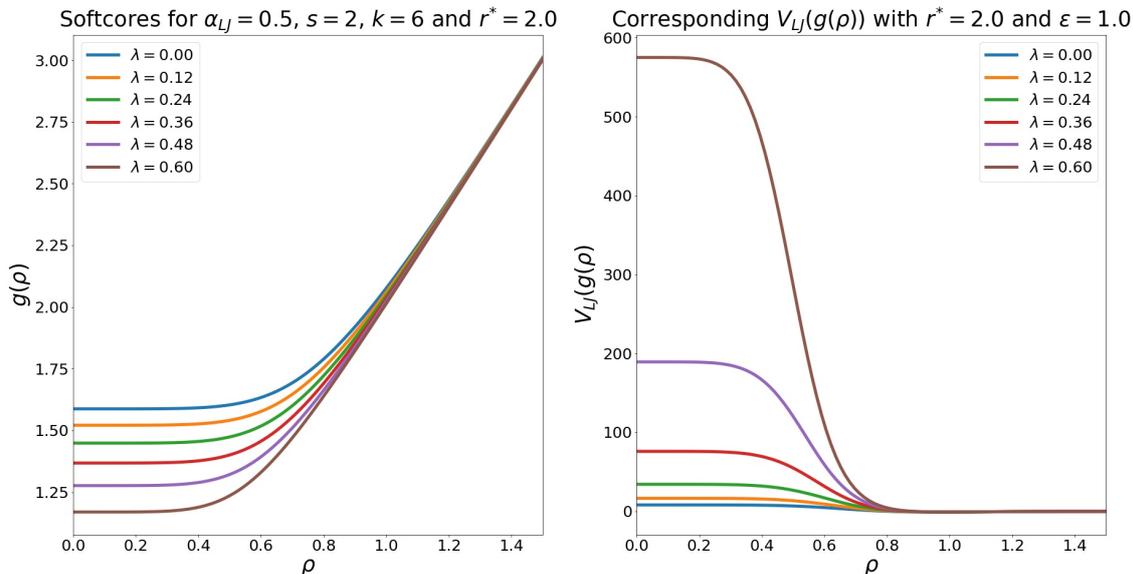


Figure 6.7 – Softcore function and associated modified Lennard-Jones potential with $s = 2$, $t = 1$, $k = 6$ and $\alpha = 1$, for λ between 0 and 0.6.

6.4.2.3 Useful derivatives

In the case where a softcore potential is used for van der Waals interactions, one would need the analytical expression of the space derivatives, the derivatives with respect to λ and the cross-derivatives of said potential. The derivation of these expressions may be found in Section C.2 of Annex C. All of the needed derivatives will be directly implemented in the `elambdaLJ1c` subroutine of the `elJ1` routine of the Tinker code.

6.4.3 Softcores for electrostatic interactions

6.4.3.1 Particle Mesh Ewald

The total electrostatic potential energy of a system of N particles is given by:

$$V_e(r, q) = \frac{1}{4\pi\epsilon_0} \sum_{i < j} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (6.13)$$

where ϵ_0 is the vacuum permittivity. If N is not too large and finite, the electrostatic potential energy (6.13) can be evaluated as such. However, in practice, the use of periodic boundary conditions as introduced in Section 2.1.3 implies that there is an infinity of periodic replicas for each particle: the potential (6.13) does not converge and one has to rely on numerical tricks to compute electrostatic interactions. The *Ewald summation* method [41] is used in this scope to

compute long-range interactions for periodic systems. It relies on the following equality:

$$\forall r \in \mathbb{R}_+^*, \quad \frac{1}{r} = \frac{\operatorname{erfc}\left(\frac{\sqrt{\eta}}{2}r\right)}{r} + \frac{\operatorname{erf}\left(\frac{\sqrt{\eta}}{2}r\right)}{r},$$

where the error function erf and its complement erfc are given by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \quad \text{and} \quad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2} dt, \quad \forall x \in \mathbb{R}^+.$$

Now, the first right-hand term is singular when the distance r goes to 0, but converges towards 0 as r goes to $+\infty$. On the other hand, the second right-hand term converges towards $\sqrt{\eta/\pi}$ as r goes to 0, but has a long tail when r goes to $+\infty$.

This decomposition of the inverse of the position leads to the division of the energy into two parts:

$$V_e(r, q) = \frac{1}{4\pi\epsilon_0} \left(\sum_{i<j} q_i q_j \frac{\operatorname{erfc}\left(\frac{\sqrt{\eta}|\mathbf{r}_i - \mathbf{r}_j|}{2}\right)}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i<j} q_i q_j \frac{\operatorname{erf}\left(\frac{\sqrt{\eta}|\mathbf{r}_i - \mathbf{r}_j|}{2}\right)}{|\mathbf{r}_i - \mathbf{r}_j|} + \right)$$

The first part is a short-range term calculated in the real space and can be evaluated *directly*, whereas the second is a long-range term calculated in the Fourier space. When compared to direct summation of pair interactions, the Ewald summation method yields a faster convergence of the energy. Several algorithms have been designed to scale with a smaller power of N than the regular Ewald summation method. The *Particle Mesh Ewald* (PME) method [27] is one of them, and its computational cost is relatively low compared to the Ewald method. Electrostatic interactions will be treated with the PME method in the Tinker code. More precisely, the PME method is implemented in the `echarge1c` subroutine of the `echarge1` routine. Let us quickly review its concept.

Let us denote by (q_1, \dots, q_N) the ponctual charges of the N particules, and denote by $(\mathbf{r}_1, \dots, \mathbf{r}_N)$ their positions. Denote by \mathcal{U} the unit cell of the periodic system (see Figure 2.1). In the unit cell, the charges satisfy the following condition:

$$q_1 + \dots + q_N = 0.$$

One has that $\mathcal{U} = \operatorname{Vect}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$, where the vectors $(\mathbf{a}_i)_{i \in \llbracket 1, 3 \rrbracket}$ are *a priori* not orthogonal. Now, consider $(\mathbf{a}_i^*)_{i \in \llbracket 1, 3 \rrbracket}$ the reciprocal conjugate vectors of the cell vectors $(\mathbf{a}_i)_{i \in \llbracket 1, 3 \rrbracket}$, defined as:

$$\mathbf{a}_1^* = 2\pi \frac{\mathbf{a}_2 \wedge \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \wedge \mathbf{a}_3)}, \quad \mathbf{a}_2^* = 2\pi \frac{\mathbf{a}_3 \wedge \mathbf{a}_1}{\mathbf{a}_2 \cdot (\mathbf{a}_3 \wedge \mathbf{a}_1)}, \quad \mathbf{a}_3^* = 2\pi \frac{\mathbf{a}_1 \wedge \mathbf{a}_2}{\mathbf{a}_3 \cdot (\mathbf{a}_1 \wedge \mathbf{a}_2)},$$

and let $\mathcal{N} = \{\mathbf{n} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3, | n_1, n_2, n_3 \in \mathbb{N}, \text{ all non-negative}\}$. The ponctual charge q_i interacts with the others charges in the unit cell $(q_j)_{j \neq i}$ along with their periodic images of positions $\{\mathbf{r}_j + \mathbf{n}\}_{\mathbf{n} \in \mathcal{N}}$, but it also interacts with its own periodic images of positions $\{\mathbf{r}_i + \mathbf{n}\}_{\mathbf{n} \in \mathcal{N}}$. Let $\mathcal{M} = \{\mathbf{m} = m_1 \mathbf{a}_1^* + m_2 \mathbf{a}_2^* + m_3 \mathbf{a}_3^*, | m_1, m_2, m_3 \in \mathbb{N}, \text{ all non-negative}\}$ be the ensemble of the reciprocal mesh vectors. Each ponctual charge q_i has fractional coordinates (s_1^i, s_2^i, s_3^i) which are defined as follows:

$$s_k^i = \mathbf{a}_k^* \cdot \mathbf{r}_i, \quad \forall k \in \llbracket 1, 3 \rrbracket.$$

Eventually, define \mathfrak{M} the set of the so-called "masked" pairs $(i, j) \in \llbracket 1, N \rrbracket^2$ for which the nonimaged, nonbond interactions are not taken into account. The total electrostatic energy of

the unit cell \mathcal{U} is then given by:

$$E_e^{\mathcal{U}} = \frac{1}{2} \sum_{\substack{\mathbf{n} \in \mathcal{N} \\ \mathbf{n} \neq \mathbf{0}}} \sum_{\substack{i,j \\ i \neq j}} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|}.$$

The infinite sum over \mathcal{N} is **conditionally convergent**: depending on the convergence of $\mathbf{n} \in \mathcal{N}$ towards $+\infty$ (i.e, on the convergence of each n_i towards $+\infty$), the limit may differ. The series is consequently not absolutely convergent. The Ewald summation method is then used to decompose the energy $E_e^{\mathcal{U}}$ into three absolutely convergent parts [27]:

- a *direct* term E_{dir} , also called *real* part, evaluated in the real, Cartesian space.
- a *reciprocal* term E_{rec} , evaluated in the Fourier space.
- and a correction term E_{corr} , evaluated in the real space.

Let us give their expressions. One first need to define the *structure factor*:

$$S(\mathbf{m}) = \sum_{j=1}^N q_j e^{2\pi i(\mathbf{m} \cdot \mathbf{r}_j)} = \sum_{j=1}^N q_j e^{2\pi i(m_1 s_1^j + m_2 s_2^j + m_3 s_3^j)}, \quad \forall \mathbf{m} \in \mathcal{M}.$$

We then have

$$E_e^{\mathcal{U}} = E_{dir} + E_{rec} + E_{corr}$$

with

$$\begin{aligned} E_{dir}(r, q) &= \frac{1}{2} \sum_{\substack{\mathbf{n} \in \mathcal{N} \\ \mathbf{n} \neq \mathbf{0}}} \sum_{\substack{i,j=1 \\ i \neq j \\ (i,j) \notin \mathfrak{M}}}^N q_i q_j \frac{\text{erfc}(\beta_e |\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|} \\ E_{rec}(r, q) &= \frac{1}{2\pi \text{vol}(\mathcal{U})} \sum_{\substack{\mathbf{m} \in \mathcal{M} \\ \mathbf{m} \neq \mathbf{0}}} \frac{S(\mathbf{m})S(-\mathbf{m})}{|\mathbf{m}|^2} e^{-\frac{\pi^2 |\mathbf{m}|^2}{\beta_e^2}}, \quad \text{vol}(\mathcal{U}) = \mathbf{a}_1 \cdot (\mathbf{a}_2 \wedge \mathbf{a}_3) \\ E_{corr}(r, q) &= -\frac{1}{2} \sum_{(i,j) \in \mathfrak{M}} q_i q_j \frac{\text{erf}(\beta_e |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{\beta} \sum_{i=1}^N q_i^2, \end{aligned}$$

the parameter β_e being fixed before the simulation. One can then define the following interaction forces

- the direct force $-\nabla E_{dir}(r, q)$,
- the reciprocal force $-\nabla E_{rec}(r, q)$,
- and the correction force $-\nabla E_{corr}(r, q)$,

where, for an energy E , ∇ designates its Cartesian gradient, namely,

$$\nabla E = (\partial_r E(r) \cdot \partial_x r, \partial_r E(r) \cdot \partial_y r, \partial_r E(r) \cdot \partial_z r)^\top.$$

6.4.3.2 First attempt: "hardcore" potentials

As a first step, one may modify the electrostatic potential in a very simple way. Let us consider what will be called a *hardcore* potential energy:

$$V_{HC-e}(r, q; \lambda_e) = V_{dir}(r; \lambda_e) + (1 - \lambda_e)V_{rec}^0(r, q) + \lambda_e V_{rec}^1(r, q) + V_{corr}(r, q; \lambda_e), \quad (6.14)$$

where $V_{rec}^0(r, q)$ (resp. $V_{rec}^1(r, q)$) is the reciprocal electrostatic energy at the initial state $\lambda_e = 0$ (resp. at the final state $\lambda_e = 1$). Consequently, one has:

$$\frac{\partial V_{HC-e}}{\partial \lambda_e} = \frac{\partial V_{dir}}{\partial \lambda_e} + V_{rec}^1(r, q) - V_{rec}^0(r, q) + \frac{\partial V_{corr}}{\partial \lambda_e}.$$

In the Tinker code, for a fixed value of the collective variable λ (and hence of λ_e), one updates the charges $(q_i)_{i \in [1, N]}$ by scaling them with λ_e . Hence the following expressions:

$$V_{dir}(r, q; \lambda_e) = \frac{1}{2} \sum_{\substack{\mathbf{n} \in \mathcal{N} \\ \mathbf{n} \neq \mathbf{0}}} \sum_{\substack{i, j=1 \\ i \neq j \\ (i, j) \notin \mathfrak{N}}}^N (\lambda_e)^2 q_i q_j \frac{\text{erfc}(\beta_e |\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|)}{|\mathbf{r}_j - \mathbf{r}_i + \mathbf{n}|} = (\lambda_e)^2 E_{dir}(r, q),$$

$$V_{corr}(r, q; \lambda) = -\frac{1}{2} \sum_{(i, j) \in \mathfrak{N}} (\lambda_e)^2 q_i q_j \frac{\text{erf}(\beta_e |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{1}{\beta_e} \sum_{i=1}^N (\lambda_e)^2 q_i^2 = (\lambda_e)^2 E_{corr}(r, q).$$

In the end:

$$\begin{aligned} V_{HC-e}(r, q; \lambda_e) &= V_{dir}(r; \lambda_e) + (1 - \lambda_e)V_{rec}^0(r, q) + \lambda_e V_{rec}^1(r, q) + V_{self}(r, q; \lambda_e) \\ &= (\lambda_e)^2 E_{dir}(r, q) + (1 - \lambda_e)E_{rec}^0(r, q) + \lambda_e E_{rec}^1(r, q) \\ &\quad + (\lambda_e)^2 E_{corr}(r, q) \end{aligned}$$

which yields

$$\frac{\partial V_{HC-e}}{\partial \lambda_e} = 2\lambda_e (E_{dir}(r, q) + E_{corr}(r, q)) + E_{rec}^1(r, q) - E_{rec}^0(r, q),$$

along with

$$\frac{\partial V_{HC-e}}{\partial \lambda} = \frac{\partial V_e}{\partial \lambda_e} \frac{d\lambda_e}{d\lambda} = 2\lambda_e \frac{d\lambda_e}{d\lambda} (E_{dir}(r, q) + E_{corr}(r, q)) + (E_{rec}^1(r, q) - E_{rec}^0(r, q)) \frac{d\lambda_e}{d\lambda}.$$

In order to implement the OSRW method, we will also need the following double derivative:

$$\frac{\partial V_{HC-e}}{\partial r \partial \lambda} = \frac{\partial V_e}{\partial \lambda \partial r} = \frac{d\lambda_e}{d\lambda} \left(2\lambda_e \left(\frac{\partial E_{dir}(r, q)}{\partial r} + \frac{\partial E_{corr}(r, q)}{\partial r} \right) + \left(\frac{\partial E_{rec}^1(r, q)}{\partial r} - \frac{\partial E_{rec}^0(r, q)}{\partial r} \right) \right),$$

so that one has, for all $i \in \{1, 2, 3\}$:

$$\frac{\partial}{\partial x_i} F_e(r, q; \lambda_e) := \frac{\partial}{\partial x_i} \frac{\partial V_{HC-e}}{\partial \lambda} = \frac{\partial V_e}{\partial r \partial \lambda} \frac{dr}{dx_i}.$$

Furthermore, we will need the following derivative

$$\frac{\partial}{\partial \lambda_e} F_e(r, q; \lambda_e) = 2(E_{dir}(r, q) + E_{corr}(r, q)).$$

The hardcore electrostatic potential was part of our first attempt at implementing the λ -dynamics. Combined with the softcore potential (6.11) for the van der Waals interaction, it showed already promising results, as we shall see further on. Hardcore potentials proved to be sufficient to obtain proper free energy differences in the context of a classical λ -dynamics. However, the introduction of the F_λ dynamical collective variable in the code needed for the OSRW method led to numerical instabilities. Many simulations were stopped before the end-time, as the energy of the system skyrocketed. Such instabilities could be explained by the fact that the extended variable ϕ , introduced later on in Section 6.5 whose evolution is coupled to that of F_λ cannot quickly follow the movements of λ , which is reflected when trespassing the boundaries 0 and 1. As a consequence, important forces have to be applied on F_λ , leading to a constant increase of the system's energy. An idea to tackle these instabilities would be to handle the electrostatic interactions with better care, especially for small distances. As a consequence, a second attempt has been made to incorporate softcore potentials for electrostatic interactions, in the manner of Section 6.4.2.

6.4.3.3 Second attempt: proper softcore potential

We will introduce a softcore for the electrostatic interactions, based on the suggestion of W. Yang, P. Ren, M. J. Schnieders and colleagues [80] along with the work of T. Steinbrecher et al. [83]. Only the real part and correction term of the PME potential will be modified. Let us consider two atoms, i and j . During the alchemical transitions, there can be three possibilities: either both atoms are being *mutated*, namely, they are being transformed, or only one of them is mutated, or none of them is mutated. Let us denote by $r = |\mathbf{r}_i - \mathbf{r}_j|$ the distance between the two atoms. We will consider the following modified electrostatic potential:

$$V_{SC-e}^{ij}(r, \lambda_e) = V_{dir}^{ij}(r, \lambda_e) + V_{corr}^{ij}(r, \lambda_e) + (1 - \lambda_e)E_{rec}^0(r) + \lambda_e E_{rec}^1(r) \quad (6.15)$$

where

$$V_{dir}^{ij}(r, \lambda_e) = \begin{cases} \lambda_e^{2t} E_{real}(f(r, \lambda_e)) & \text{if } i \text{ and } j \text{ are mutated} \\ \lambda_e^t E_{real}(f(r, \lambda_e)) & \text{if } i \text{ or } j \text{ is mutated, but not the other} \\ E_{real}(r) & \text{if none are mutated} \end{cases}$$

and

$$V_{corr}^{ij}(r, \lambda_e) = \begin{cases} \lambda_e^{2t} E_{corr}(f(r, \lambda_e)) & \text{if } i \text{ and } j \text{ are mutated} \\ \lambda_e^t E_{corr}(f(r, \lambda_e)) & \text{if } i \text{ or } j \text{ is mutated, but not the other} \\ E_{corr}(r) & \text{if none are mutated,} \end{cases}$$

where

$$f(r, \lambda_e) = (\alpha_e(1 - \lambda_e)^s + r^u)^{\frac{1}{u}}, \quad (6.16)$$

is the softcore function, with $\alpha_e \in \mathbb{R}_*^+$ and $s \in \mathbb{N}^*$ being tunable parameters.

As said beforehand, one is interested into implementing the derivative:

$$\begin{aligned} \frac{\partial V_{SC-e}^{ij}(r, \lambda)}{\partial \lambda} &= \frac{\partial V_{SC-e}^{ij}(r, \lambda_e)}{\partial \lambda_e} \frac{d\lambda_e}{d\lambda} \\ &= \left(\frac{\partial V_{dir}^{ij}(r, \lambda_e)}{\partial \lambda_e} + \frac{\partial V_{corr}^{ij}(r, \lambda_e)}{\partial \lambda_e} + E_{rec}^1(r) - E_{rec}^0(r) \right) \frac{d\lambda_e}{d\lambda} \end{aligned}$$

and the Cartesian gradient:

$$\nabla V_{SC-e}^{ij}(r, \lambda_e) = \nabla V_{dir}^{ij}(f(r, \lambda_e)) + \nabla V_{corr}^{ij}(f(r, \lambda_e)) + (1 - \lambda_e) \nabla E_{rec}^0(r) + \lambda_e \nabla E_{rec}^1(r). \quad (6.17)$$

We will also compute the following double derivatives, which will be needed for the OSRW implementation:

$$\frac{\partial^2 V_{SC-e}^{ij}(r, \lambda)}{\partial \lambda^2} = \frac{\partial^2 V_{SC-e}^{ij}(r, \lambda_e)}{\partial \lambda_e^2} \left(\frac{d\lambda_e}{d\lambda} \right)^2 \quad (6.18)$$

$$\begin{aligned} \nabla \left(\frac{\partial V_{SC-e}^{ij}(r, \lambda)}{\partial \lambda} \right) &= \frac{d\lambda_e}{d\lambda} \left(\nabla \frac{\partial V_{dir}^{ij}(r, \lambda_e)}{\partial \lambda_e} + \nabla \frac{\partial V_{corr}^{ij}(r, \lambda_e)}{\partial \lambda_e} \right. \\ &\quad \left. + \nabla E_{rec}^1(r) - \nabla E_{rec}^0(r) \right). \end{aligned} \quad (6.19)$$

In practice, in order to compute the Cartesian gradient, we will need to compute derivatives with respect to r . The derivation the softcore potential's derivatives may be found in Section C.3 of annex C. All of these expressions have been implemented in the Tinker-HP code in the `elambda-dareal1c` subroutine of the `echarge1` routine. The parameter α_e , s , t and u are declared in the `mutant` module, and their values need to be written by the user in the input simulation `.key` file.

Choice of the parameters Several works [12, 80] suggest the use of $\mathfrak{t} = 1$ and $\mathfrak{s} = \mathfrak{u} = 2$. As for the value of the constant α_e , it has been shown to be dependent of the parameter α_{LJ} of the van der Waals softcore (6.11): in [83], free energy calculations were done with values of α_{LJ} varying from 0.2 to 0.5 and α_e from 1.5 to 5, in order to determine the best parameter setting. From the above derivatives, one knows that the softcore $f(r, \lambda_e)$ is not allowed to be null for any value of λ_e . If $\lambda_e = 1$, one notices that

$$f(r, 1) = r,$$

so that a cutoff is necessary to avoid the singularity at $r = 0$. For values of λ in $[0, 1)$, the singularity is avoided. Provided the parameters \mathfrak{s} and \mathfrak{u} are set to 2, T. C. Beutler and colleagues [12] view α_e as the square root of a critical radius r_0 , namely $\alpha_e = r_0^2$ so that the interaction stop growing for distances lesser than $r_0(1 - \lambda)$. As a consequence, values of α_e between 0.6 and 1.6 would give values of r_0 between 0.77 and 1.26 Å. Standard parameter values are still to be determined: simulations will be run with several values of α_e , with \mathfrak{s} and \mathfrak{u} being usually set to 2.

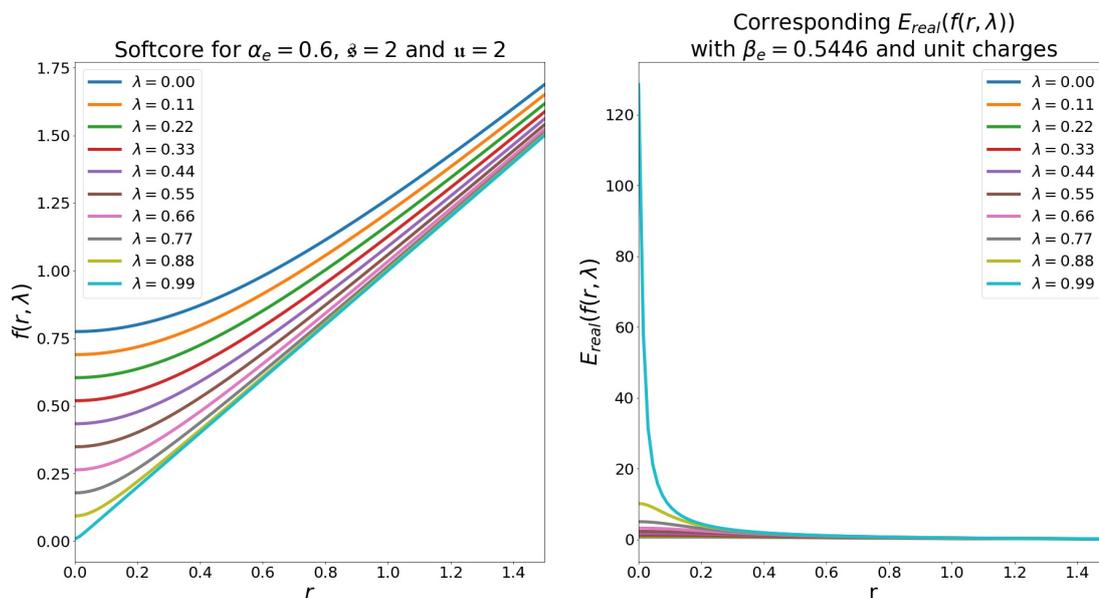


Figure 6.8 – Softcore function and associated modified direct potential energy with $s = u = 2$ and $\alpha_e = 0.6$, for λ between 0 and 0.99.

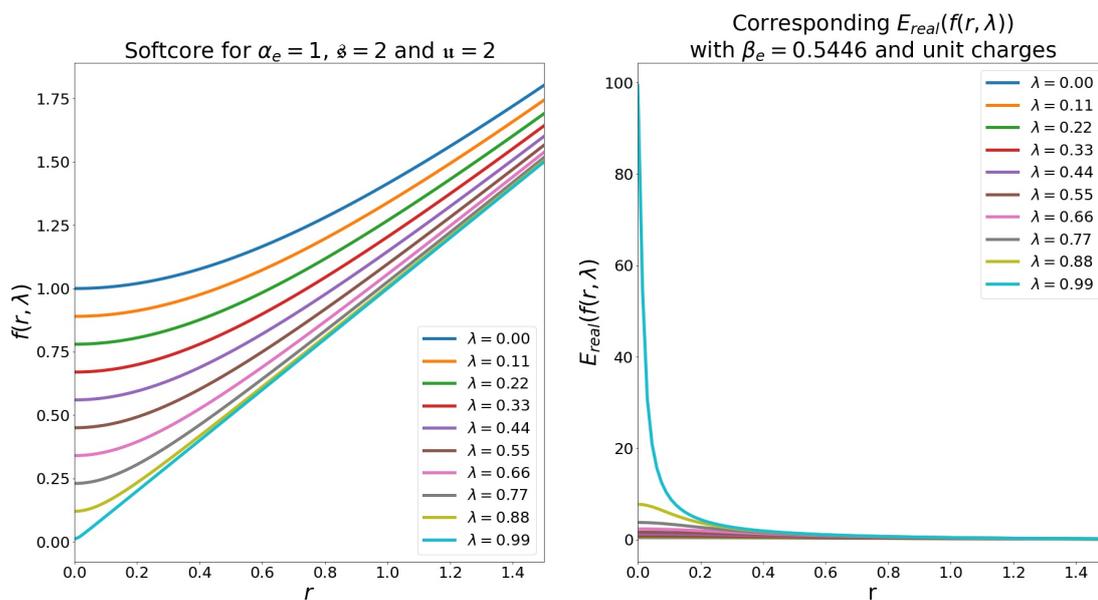


Figure 6.9 – Softcore function and associated modified direct potential energy with $s = u = 2$ and $\alpha_e = 1$, for λ between 0 and 0.99.

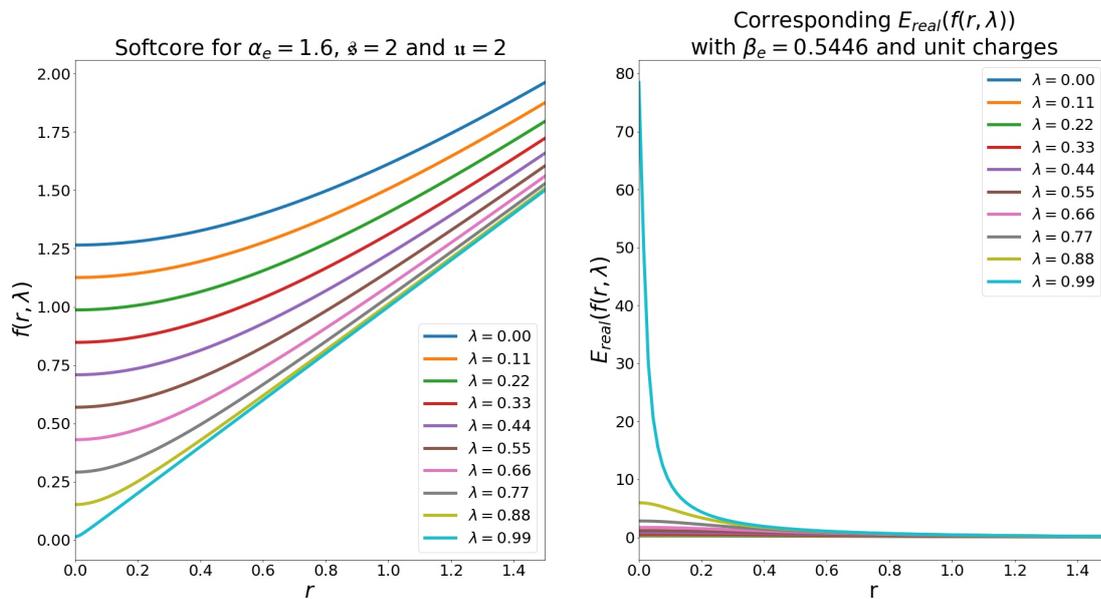


Figure 6.10 – Softcore function and associated modified direct potential energy with $s = u = 2$ and $\alpha_e = 1.6$, for λ between 0 and 0.99.

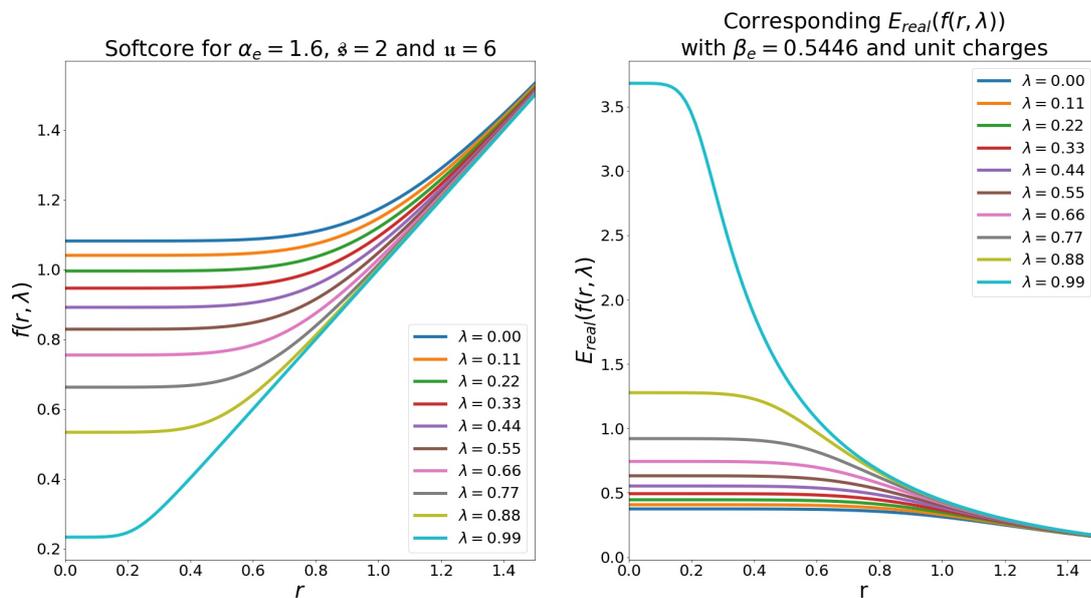


Figure 6.11 – Softcore function and associated modified direct potential energy with $s = 2$, $u = 6$ and $\alpha_e = 1.6$, for λ between 0 and 0.99.

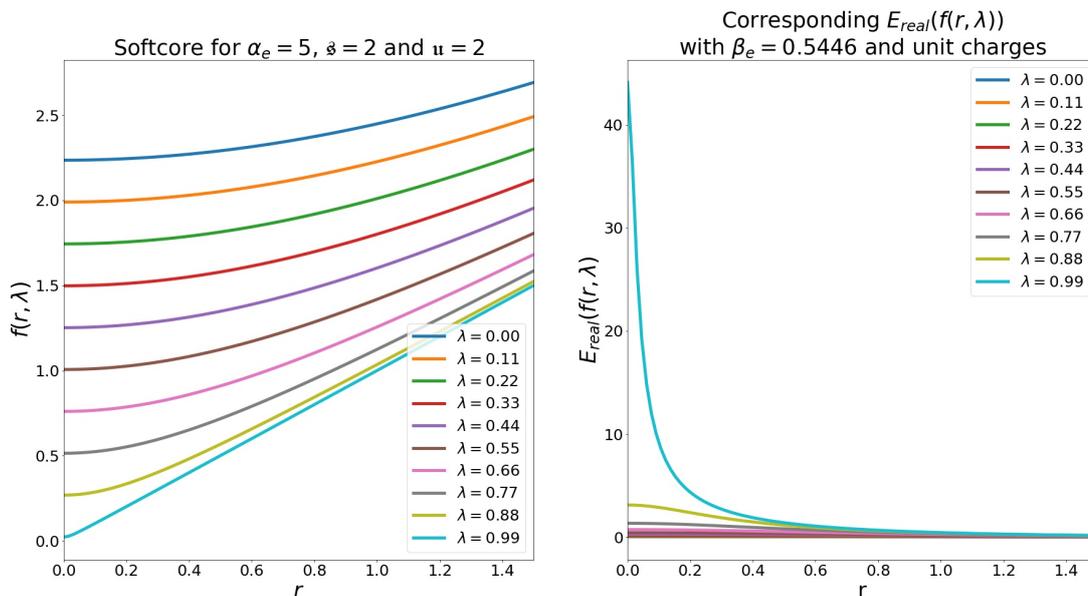


Figure 6.12 – Softcore function and associated modified direct potential energy with $s = u = 2$ and $\alpha_e = 5$, for λ between 0 and 0.99.

6.4.4 Tinker-HP – Colvars interface: first update

Let us quickly recap how a simulation using the λ -dynamics method is treated by the Tinker-HP–Colvars interface. At a given time t , Tinker-HP has a value for λ (`lambda`), and can define the corresponding $\lambda_e(\lambda)$ (`elambda`) and $\lambda_v(\lambda)$ (`vlambda`) in the subroutine `def_lambdadyn`. Tinker-HP then propagates the dynamics for one timestep dt , and calculates the force $\partial_\lambda V_{ext} = F_\lambda$ (`delambda`) applied on λ . At the end of the timestep, the interface routines `get_alch_lambda` and `get_delambda` send the value of λ and F_λ respectively to the Colvars module. The colvars module defines its own collective variable λ_c (`alch_lambda`) which here corresponds to λ , along with the force applied on it, F_{λ_c} (`alch_Flambda`) which here corresponds to F_λ . Colvars then propagates λ_c according to the Langevin dynamics (6.3) over one timestep, using leapfrog integration. If $\lambda_c(\lambda) \neq \lambda$, which is not our case here, Colvars then computes the corresponding $\lambda = \lambda_c^{-1}(\lambda_c(\lambda))$, along with corresponding $F_\lambda = F_{\lambda_c} \times \lambda'_c(\lambda)$. The interface routines `set_alch_lambda` and `apply_force_lambda` then send the new values of λ and of F_λ to Tinker-HP as initial values of λ and F_λ at time $t + dt$.

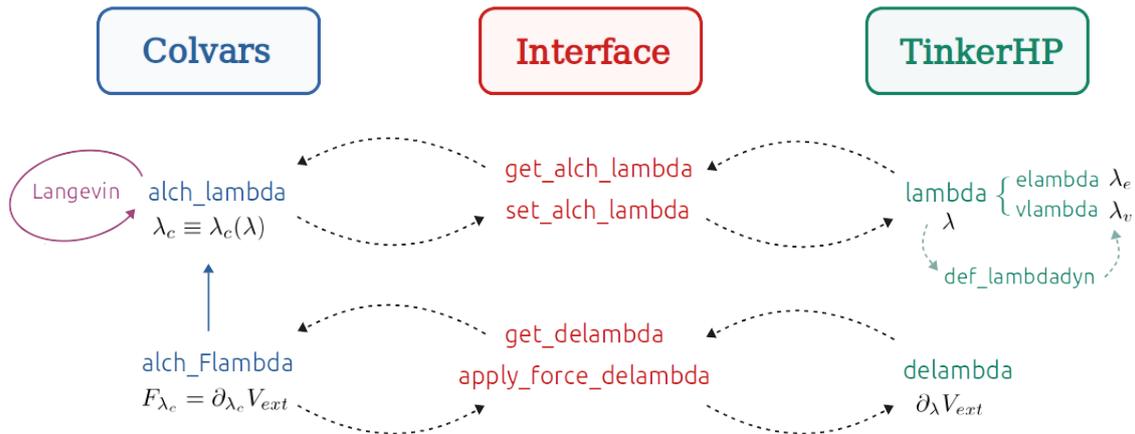


Figure 6.13 – Scheme of the communication between the Tinker-HP and Colvars module codes when running a simulation with the λ -dynamics method.

6.4.5 First numerical results

6.4.5.1 On the mass of the collective variable λ

First, one needs to treat the matter of the fictitious mass \mathbf{m}_λ of the collective variable λ .

Parametrising the mass in Colvars – In order for the Colvars module to launch a proper extended dynamics, we will need to tune the following parameters:

- ▷ **extendedLagrangian**, is a boolean with default value set to **off**. It allows the addition of a degree of freedom by considering the collective variable to be coupled to a fictitious particle via a harmonic spring.
- ▷ **extendedFluctuation**, σ , is a real number in the collective variable's unit. It defines the spring constant by setting the deviation between the collective variable and the fictitious particle induced by the thermal fluctuations of the system. This constant is computed internally and is given by $\frac{k_B T}{\sigma^2}$. The units of σ depends on the collective variable at hand.
- ▷ **extendedTimeConstant**, τ , is a real, non-negative number with default value set to 200 fs. It defines the oscillation period of the harmonic oscillator composed of the spring and the fictitious particle. Note that the period should be significantly higher than the timestep, so that the equations of motion for the particle are correctly integrated. The inertial mass of the particle coupled to the collective variable is then given by:

$$\mathbf{m}_\lambda = k_B T \left(\frac{\tau}{2\pi\sigma} \right)^2.$$

- ▷ **extendedTemp**, is a real, non-negative number with default value set to the thermostat's temperature, in Kelvins. It defines the temperature used to compute the coupling force constant given by $\frac{k_B T}{\sigma^2}$. This temperature can also be the target temperature for the extended Langevin dynamics.

- ▷ **extendedLangevinDamping**, γ , is a real, non-negative number with default value set to 1.0 ps⁻¹. It defines the damping factor γ of the extended Langevin dynamics. If the damping is not 0, the collective variable follows a Langevin dynamics only one collective variable λ , of momenta A , looking at the the Langevin dynamics (6.3) at temperature T fixed by **extendedTemp**. Using the damping can prove to be useful when using the ABF method, where the collective variable can heat up rapidly out of equilibrium. γ should be reasonably small, otherwise the friction term in the Langevin dynamics would slow down the phase space sampling.

Dimensional analysis – One may ask what is the dimension \mathfrak{m}_λ . The stiffness k of the spring which links the collective variable λ to the fictitious particle of mass \mathfrak{m}_λ has the dimension of an energy divided by the dimension of λ squared, as λ is a generalised coordinate (cf. Section B.2.3 of Annex B). Note that the fictitious particle has the same dimension as λ . If one denotes by U the collective variable unit compatible with SI units (which is not the case in the Collective variable module code in general), then one gets $[k] = ML^2U^{-2}T^{-2}$. Now, let us determine the dimension of the fluctuation parameter σ :

$$[k] = \frac{[k_B][T]}{[\sigma^2]} = \frac{ML^2T^{-2}\Theta^{-1} \times \Theta}{[\sigma^2]} = \frac{ML^2T^{-2}}{[\sigma^2]} = ML^2U^{-2}T^{-2},$$

which yields

$$[\sigma] = U.$$

In conclusion, the dimension of the parameter σ changes with the nature of λ : the collective variable can well be a distance, an angle, or even without dimension, as in our case. If there is only one collective variable λ , of momenta A , looking at the Langevin dynamics (6.3) one notes that:

$$[\lambda_t] = [\mathfrak{m}_\lambda^{-1}] \times [A_t] \times [dt], \quad \text{et} \quad [A_t] = [-\nabla V(\lambda_t)] \times [dt].$$

Since $-\nabla V_{ext}$ is a *generalized force* (cf. Section B.2.3 of Annex B), one has that $[-\nabla V(\lambda_t)] = ML^2U^{-1}T^{-2}$. Eventually:

$$[A_t] = ML^2U^{-1}T^{-2}T = ML^2U^{-1}T^{-1},$$

and

$$[\lambda] = [\mathfrak{m}_\lambda^{-1}] \times ML^2U^{-1}T^{-1}T,$$

so that

$$[\mathfrak{m}_\lambda] = ML^2U^{-2}. \quad (6.20)$$

A direct consequence of the expression (6.20) is that there is no easy way to infer the physical nature of the mass m_λ , except from the case where λ is a distance.

Characteristic mass – Given a collective variable λ , how does one chooses its mass before running a simulation with the λ -dynamics method? One might think of fixing as order of magnitude the mass of the heaviest atom in the system, the length of the simulation box along with the maximum value of λ . In the toy case of a box of water of length $L = 18.643$, the heaviest

atom is the oxygen atom, and its molar mass is of $m_O = 15.999$ amu. The characteristic mass M_{car} of the collective variable would then be:

$$\begin{aligned} M_{car} &= (15.995 \text{ amu}) \times (18.643)^2 = (15.995 \text{ g.mol}^{-1}) \times (18.643 \times 10^{-10} \text{ m})^2 \\ &= \left(15.995 \times \frac{10^{-3} \text{ kg}}{6.022 \times 10^{23}} \right) \times (18.643 \times 10^{-10} \text{ m})^2 \\ &= (2.656 \times 10^{-26} \text{ kg}) \times 3.476 \times 10^2 \times 10^{-20} \text{ m}^2 \\ &= 9.232 \times 10^{-44} \text{ kg} \cdot \text{m}^2 \end{aligned}$$

and the associated frequency would be

$$\begin{aligned} \nu_{car} &= \frac{1}{2\pi} \sqrt{\frac{k}{M_{car}}} = \frac{1}{2\pi} \sqrt{\frac{k_B T}{M_{car}}} \\ &= \frac{1}{2\pi} \sqrt{\frac{1.38 \times 10^{-23} \times 300}{9.232 \times 10^{-44}}} = \frac{1}{2\pi} \sqrt{4.484 \times 10 \times 10^{21}} \\ &= \frac{1}{2\pi} \sqrt{4.484} \times 10^{11} \\ &= 3.370 \times 10^{10} \text{ s}^{-1}, \end{aligned}$$

where we approximated $k_b \simeq 1.38 \times 10^{-23} \text{ m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{K}^{-1}$. Since

$$\frac{k_B T}{(2\pi)^2} \simeq 1.049 \times 10^{-22},$$

if one sets $\sigma = 1\text{U}$, then

$$\mathbf{m}_\lambda = 1.049 \times 10^{-22} \times \tau^2.$$

And, in order to get $\mathbf{m}_\lambda = M_{car}$ one needs to take

$$\begin{aligned} \tau &\simeq \sqrt{\frac{M_{car}}{1.049 \times 10^{-22}}} \simeq \sqrt{\frac{9.232}{1.049} \times 10^{-44} \times 10^{22}} \\ &\simeq \sqrt{8.801 \times 10^{-22}} \\ &\simeq 2.967 \times 10^{-11} \text{ s} \\ &\simeq 2.967 \times 10^4 \text{ fs}. \end{aligned}$$

Of course, one can toy with other parameters rather than just τ in order to set the mass \mathbf{m}_λ . In practice, we noticed using parameters that gave a mass of the order of M_{car} led to a systematic explosion of the system. As a consequence, our simulations were run with an extended time constant set to $\tau = 300\text{fs}$, a fluctuation constant of $\sigma = 1$ (so that $\mathbf{m}_\lambda \simeq 9.438 \times 10^{-48} \text{ kg} \cdot \text{m}^2$ is significantly smaller than M_{car}) and a Langevin damping $\gamma = 200 \text{ ps}^{-1}$.

6.4.5.2 On the boundary conditions

In our first runs of λ -dynamics simulations, we observed that the collective variable often went outside the interval $[0, 1]$. One explanation is that λ may cross one of the boundaries $\lambda^* \in \{0, 1\}$ with a speed so large the reflecting boundary conditions implemented in the Colvars code are not sufficient to bring λ back in $[0, 1]$, as shown in Figure 6.14. Indeed, if λ is at a distance d of the boundary λ^* then the Colvars code will consider the new collective variable that is located

at $\lambda^* \pm 2d$: as soon as $d > 1$, the reflecting boundary conditions fails. In order to bypass this issue, we added an artificial safeguard in the subroutine `def_lambdadyn`: if λ is below $\lambda^* = 0$ (resp. above $\lambda^* = 1$), $\lambda_e = \lambda_v = 0$ (resp. $\lambda_e = \lambda_v = 0$) and the force acting on λ is set to 0. Such safeguard works well combined with reflecting boundary conditions in the case of our classical λ -dynamics (with both softcores for van der Waals and electrostatic interactions and for softcores for van der Waals interactions coupled with hardcore electrostatic interactions). However, as we will see further on, as soon as one wishes to implement the OSRW method, one will need a new kind of boundary conditions.

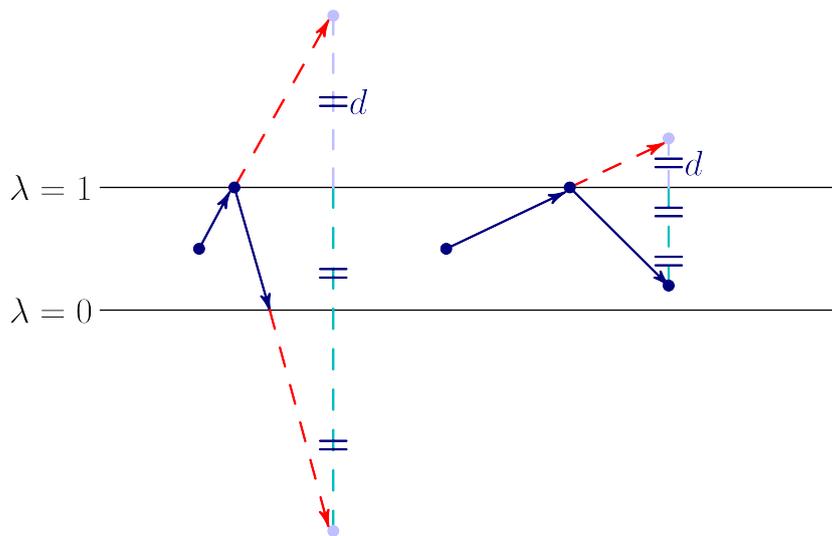


Figure 6.14 – Exemple where $\lambda^* = 1$. The first particle (left) is not correctly reflected and goes outside of the boundaries, whereas the second particle (right) is correctly reflected.

6.4.5.3 Free energy profiles

All simulations were run on the Laboratoire Jacques-Louis Lions' hpc2 calculator. The hpc2 calculator is a Altix UV 2000 computer which contains 32 CPUs Intel Xeon 64 bits EvyBridge E4650 of 10 cores each.

- **Softcore potential for van der Waals interactions, hardcore potential for electrostatic interactions ("softhard" case)**– We first tested the λ -dynamics with a softcore potential for the van der Waals interactions as in equation (6.11), along with a hardcore potential for electrostatic interactions as in (6.14). The softcore parameters of expression (6.12) were set to `sca`= $\alpha = 0.5$, `sca`= $s = 2$, `sct`= $t = 1$ and `sck`= $k = 6$. Two toy models were considered: the hydration of a water molecule and of a sodium cation in a water box of 18.643 Å length, with the Amber99 force field parameters. First, dynamics were run in the NVE ensemble in order to check energy conservation: *hysteresis*, namely the violation of the energy conservation, would mean our implementation was somewhat faulty. This ensured the derivatives expression to be properly implemented. We then ran simulations in order to obtain the hydration free energies of both the water molecule and sodium cation. Simulations were run in the NVT ensemble at a

temperature of 300K, over 2ns, with a timestep of 1fs.

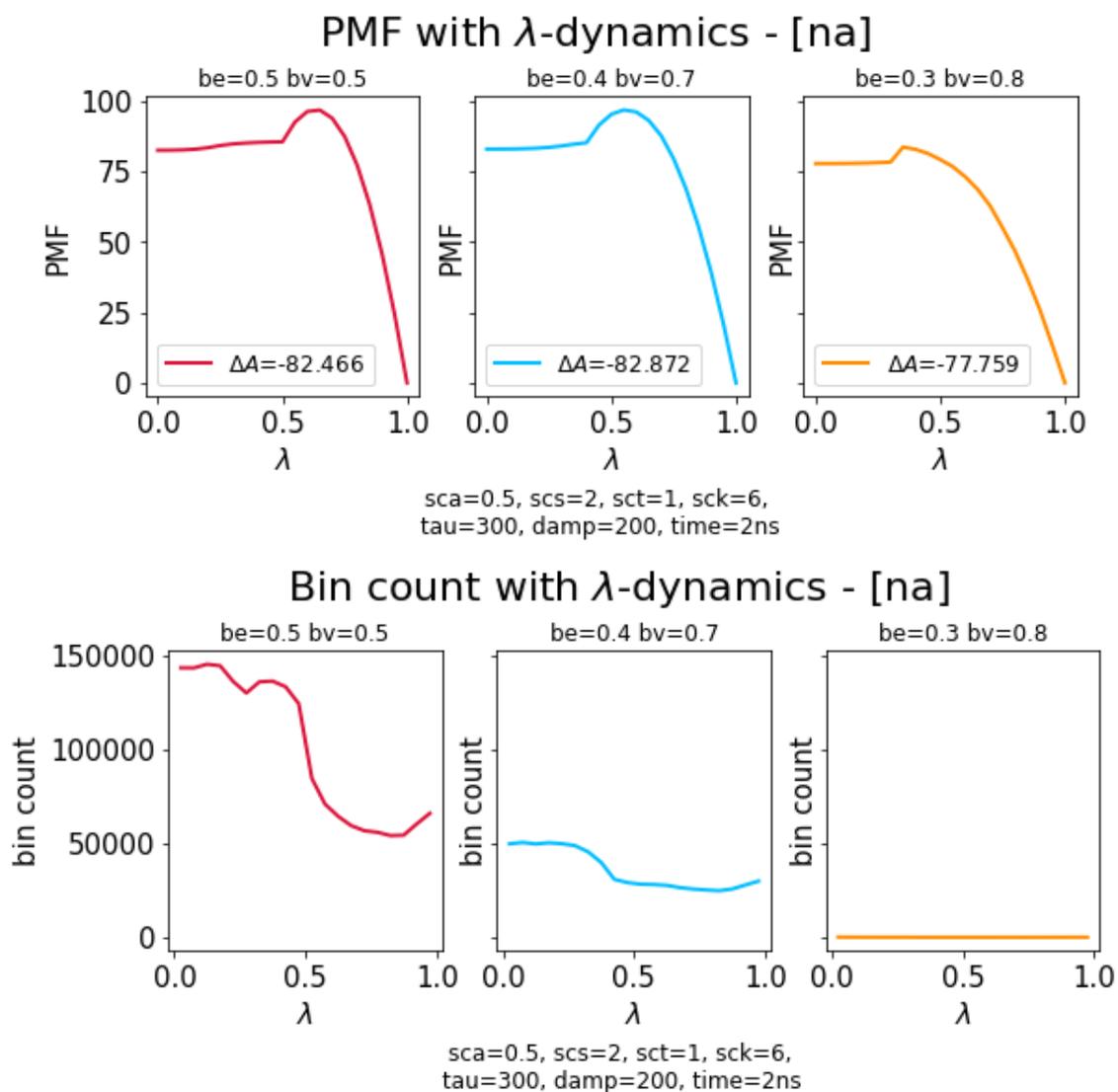


Figure 6.15 – Hydration of a sodium cation. Potential Mean Force profile (up) and bin count (down) obtained for several values of the bound b_e and b_v . The free energy difference $\Delta_{0 \rightarrow 1}A$ (in kcal.mol⁻¹) between final and initial state is printed on the PMF plots.

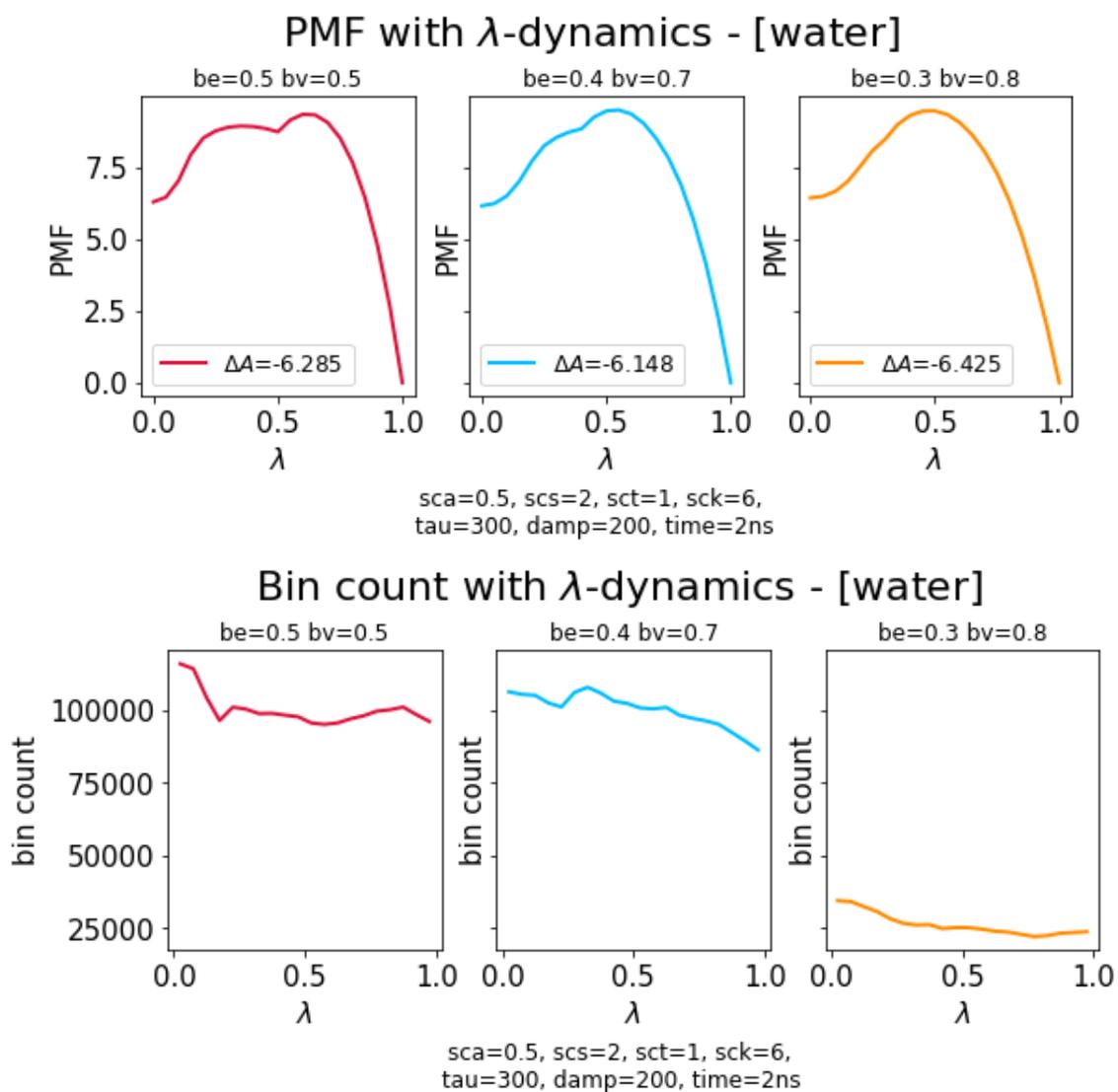


Figure 6.16 – Hydration of a water molecule. Potential Mean Force profile (up) and bin count (down) obtained for several values of the bound b_e and b_v . The free energy difference $\Delta_{0 \rightarrow 1}A$ (in kcal.mol⁻¹) between final and initial state is printed on the PMF plots.

In order to compare the Potential Mean Force obtained for a given system, we relied on the BAR method introduced in Section 3.4.2: we discretised $[0, 1]$ with a step $\Delta\lambda = 0.05$, in order to obtain 21 values $(\lambda_i)_{i \in \llbracket 0, 20 \rrbracket}$ for the collective variable λ . To each value λ_i corresponds a value of λ_v^i (resp. λ_e^i) and of λ_v (resp. λ_e). Depending on the chosen bounds b_e and b_v , the discretisation steps $\Delta\lambda_v$ and $\Delta\lambda_e$ of the λ_v and λ_e intervals may differ, as shown in Figure 6.17.

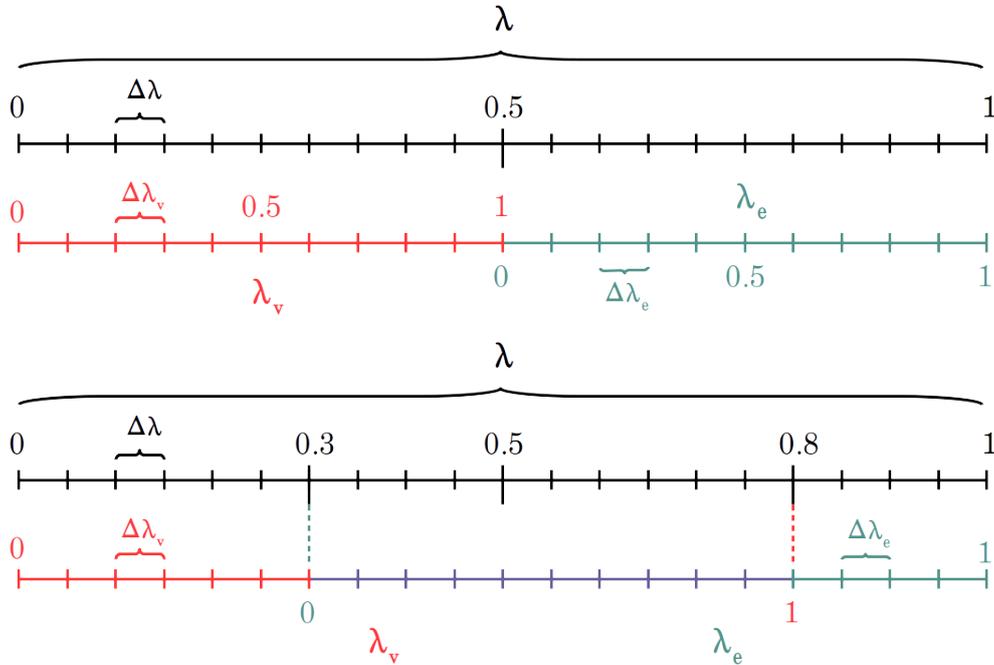


Figure 6.17 – Given a discretisation of the interval $[0, 1]$, and the values of b_e and b_v , the discretisation of the λ_v and λ_e may differ. When $b_e = b_v = 0.5$ (up) one has 10 discrete values of λ_v and 10 discrete values of λ_e . When $b_e = 0.3$ and $b_v = 0.8$ (down), one has 17 discrete values of λ_v and 15 discrete values of λ_e .

For each value λ_i (and consequently for each corresponding pair $(\lambda_v^i, \lambda_e^i)$), a 2 ns simulation was run in the NVT ensemble with the Tinker-HP program: note that since here λ does not evolve dynamically, there is no communication needed with the Colvars module. The BAR method implemented in the Tinker code was then used to compute the free energy A_i of the i -th simulation. Afterwards, the free energy difference $\Delta A_i = A_{i+1} - A_i$ of the window $[\lambda_i, \lambda_{i+1}]$ was evaluated for every $i \in \llbracket 0, 20 \rrbracket$. The Potential Mean Force is then recovered by linear interpolation on each window with a Python script. Obtaining the free energy profile of the whole transition from $\lambda = 0$ to $\lambda = 1$, and hence the free energy difference $\Delta_{0 \rightarrow 1} A = A(\lambda = 1) - A(\lambda = 0)$ then takes 2×21 ns. We ran simulations for both the hydration of a water molecule and a sodium cation, and obtained the following results:

Test	$\Delta_{0 \rightarrow 1} A$ (kcal.mol ⁻¹) λ -dynamics	$\Delta_{0 \rightarrow 1} A$ (kcal.mol ⁻¹) BAR method	Absolute error (kcal.mol ⁻¹)	Relative error
Na ⁺	-82.466	-83.4669	~ 1.0009	$\sim -1.2\%$
H ₂ O	-6.28523	-6.3218	0.03657	-0.6%

Table 6.1 – Hydration free energies for the cation Na⁺ and a water molecule, obtained from λ -dynamics and BAR method. The softcore potential (6.11) was used for van der Waals interactions and the hardcore potential (6.14) was used for electrostatic interactions.

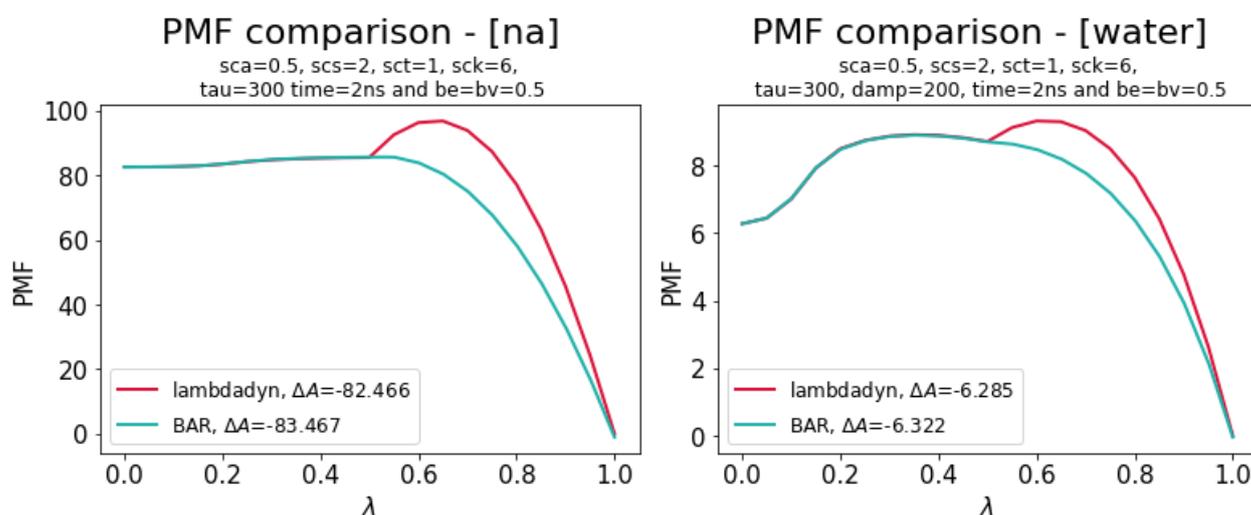


Figure 6.18 – Potential Mean Force profiles obtained for the hydration of a sodium cation (up), and of a water molecule (down). The red profiles are obtained with 2 ns λ -dynamics simulations, whereas the blue profiles are obtained with 42 ns BAR simulations. Free energy differences $\Delta_{0 \rightarrow 1} A$ (in kcal.mol⁻¹) between final and initial state are printed. The softcore potential (6.11) was used for van der Waals interactions and the hardcore potential (6.14) was used for electrostatic interactions.

Observations:

- (O1) We first notice that the free energy difference computed by the implemented λ -dynamics in 2 ns is close to the free energy difference computed with the BAR method in 42 ns for both toy models: the gain in computation time when using the λ -dynamics with a softcore potential for the van der Waals interactions and hardcore potential for electrostatic interactions is significant.

- (O2) The way the λ -dynamics is implemented does not allow the simultaneous lighting up of van der Waals and electrostatic interactions: the nearest to 0 the bound b_e and the nearest to 1 the bound b_v is, the poorest the sampling. Note that the hydration of the sodium cation is more sensible to the change in bounds than the hydration of a water molecule, as electrostatic interactions between the cation and the surrounding water molecules are non-negligible. The choice of $b_e = b_v = 0.5$ is to this day the safest to obtain satisfying free energy profiles.
- (O3) Overall, histograms show that the the dynamics does not visit the interval $[0, 1]$ equally, the van der Waals interactions being sampled more than the electrostatic interactions, hence the need to use softcore potentials for the former: this motivates the use of softcore potentials for electrostatic interactions.

Softcore potentials for both van der Waals and electrostatic interactions ("softsoft" case)– As a consequence, we consequently ran simulations with softcore potentials for both kind of interactions. The setup remained the same. Preliminary simulations were run in the NVE ensemble to check if the softcore derivatives were correct. Simulations were then run in the NVT ensemble at a temperature of 300 K, over 2 ns, with a timestep of 1 fs. We focused on the hydration of the sodium cation. The parameters of the van der Waals softcore (6.12) were always set to `sca`= 0.5, `sca`= 2, `sct`= 1 and `sck`= 6. A first test was to set the parameters of the softcore (6.16) to `fsca`= $\alpha_e = 0$, `fsca`= `s` = 2 and `fsck`= `u` = 2, so that one would get $f(r, \lambda_e) = r$ and the classical, hardcore potential (6.14) for electrostatic interactions. We compared the PMF obtained with this set of parameters with the PMF obtained formerly with only a softcore potential for the van der Waals interactions: a different PMF would have meant our implementation was somehow faulty.

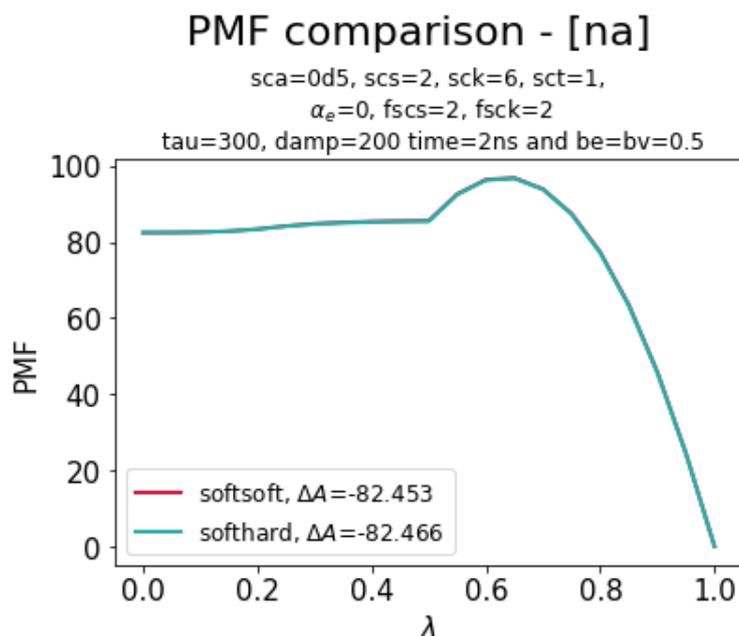


Figure 6.19 – Coinciding Potential Mean Force profiles for two simulations, one the "softsoft" case (where the softcore potential (6.11) was used for van der Waals interactions and the hardcore potential (6.14) was used for electrostatic interactions) the other in the "softhard" case (where the softcore potential (6.11) was used for van der Waals interactions and the softcore potential (6.15) was used for electrostatic interactions). The free energy difference $\Delta_{0 \rightarrow 1} A$ (in kcal.mol⁻¹) between final and initial state is printed.

Choosing the values of the parameters `fsca`, `fscs` and `fsck` proved to be difficult, and we relied on values suggested in [12] for first simulations. We ran simulations for several values of `fsca` and `fsck` with `fscs=5` being set to 2, and with bounds $b_e = b_v = 0.5$ in order to compare the PMF profiles along with the associated bin counts:

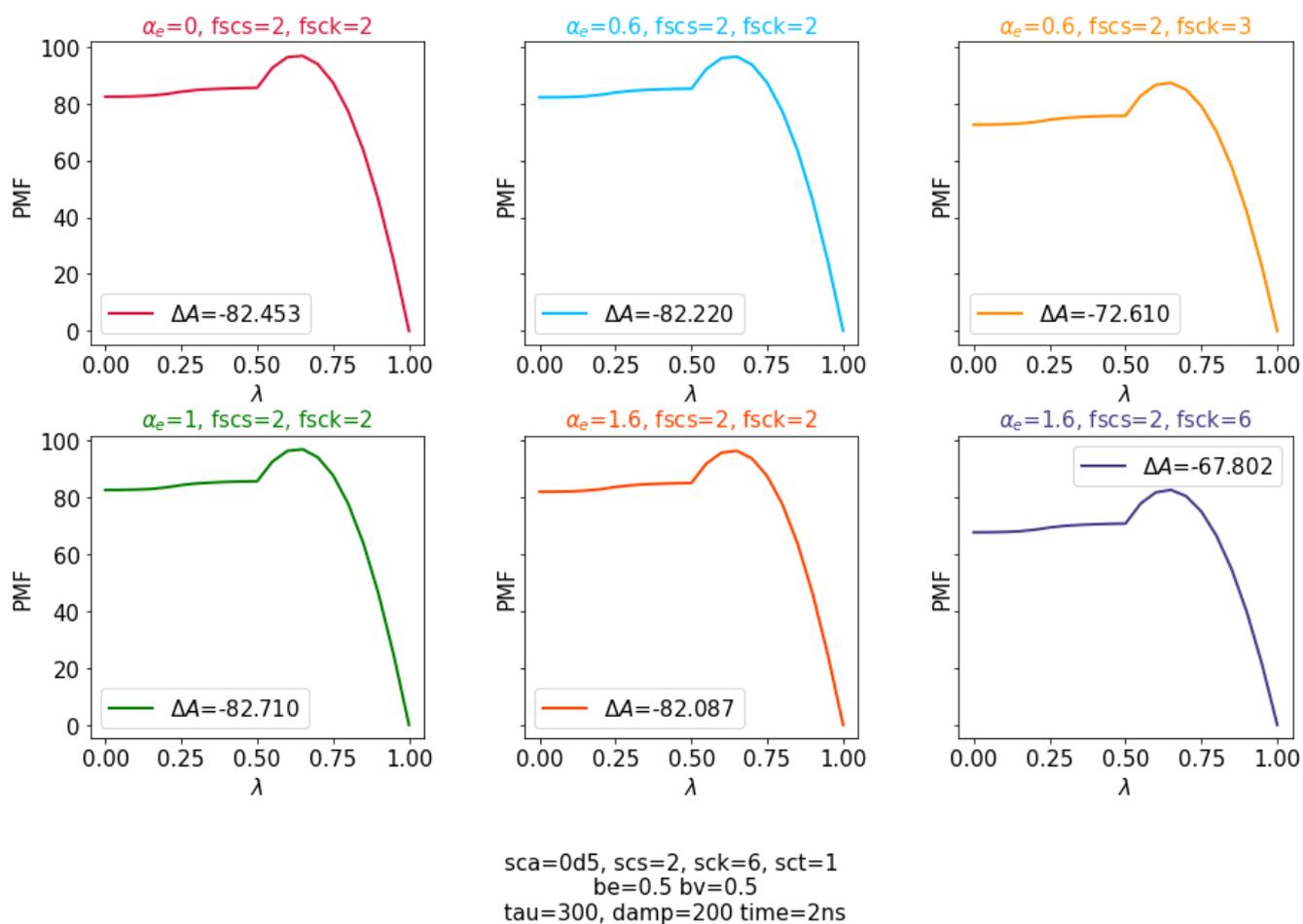
PMF with λ -dynamics - [na]

Figure 6.20 – Potential Mean Force profiles for the hydration of a sodium cation, with fixed bounds $b_e = b_v = 0.5$ and parameter $fscs = s$ set to 2. The free energy difference $\Delta_{0 \rightarrow 1} A$ (in kcal.mol⁻¹) between final and initial state is printed. The softcore potential (6.11) was used for van der Waals interactions and the softcore potential (6.15) was used for electrostatic interactions.

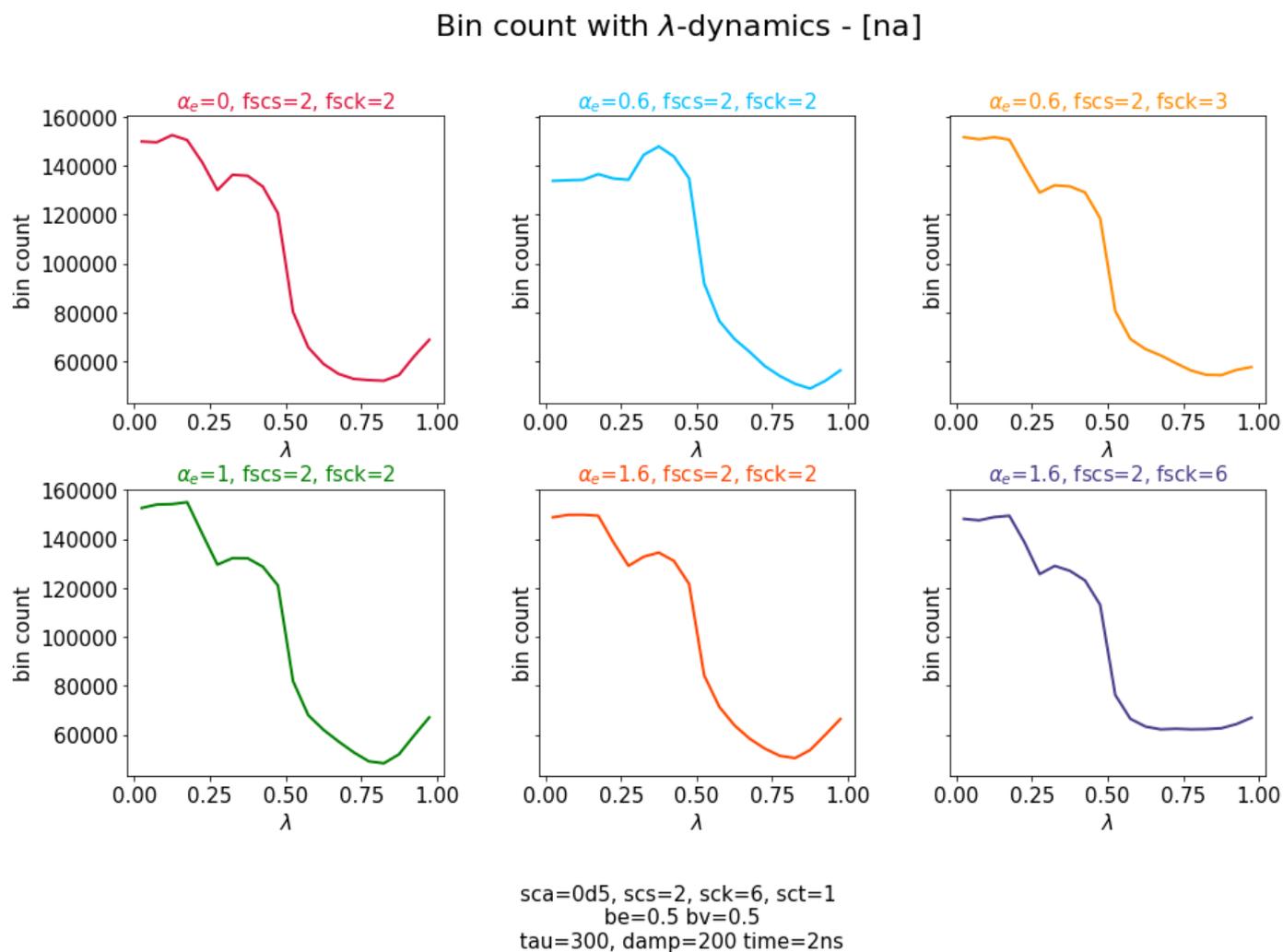


Figure 6.21 – Bin counts for the hydration of a sodium cation, with fixed bounds $b_e = b_v = 0.5$ and parameter $fscs = \mathfrak{s}$ set to 2.

We then arbitrarily choose the softcore parameters $fsca = \alpha_e = 0$, $fscs = \mathfrak{s} = 2$ and $fsck = \mathfrak{u} = 2$, in order to compare the obtained free energy difference with a reference one computed with the BAR method, in the same manner as above. Bounds were set to $b_e = b_v = 0.5$.

Test	$\Delta_{0 \rightarrow 1} A$ (kcal.mol ⁻¹) λ -dynamics	$\Delta_{0 \rightarrow 1} A$ (kcal.mol ⁻¹) BAR method	Absolute error (kcal.mol ⁻¹)	Relative error
Na ⁺	-82.7096	-83.3879	~ 0.6783	$\sim -0.008\%$

Table 6.2 – Solvation free energies for the cation Na⁺ obtained from λ -dynamics and BAR method. Softcore potentials were used for both van der Waals and electrostatic interactions.

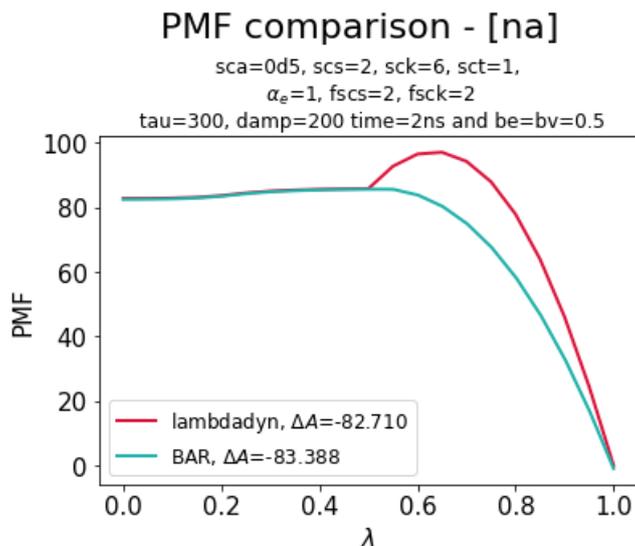


Figure 6.22 – Potential Mean Force profiles obtained for the hydration of a sodium cation. The red profiles are obtained for 2ns λ -dynamics simulation, whereas the blue profiles are obtained for 42 ns BAR simulations. Free energy differences $\Delta_{0 \rightarrow 1} A$ (in kcal.mol⁻¹) between final and initial state are printed. Softcore potentials were used for both van der Waals and electrostatic interactions.

Observations:

- (O4) One first notices that adding a softcore potential for the electrostatic interactions does seem to meliorate the computation of the free energy difference $\Delta_{0 \rightarrow 1} A$. One should of course run simulations for other systems, like the hydration of a water molecule, as done previously.
- (O5) There nonetheless does not seem to be improvement regarding the sampling of the electrostatic region: for $b_e = b_v = 0.5$, the parameter λ tends to spend more time in $[0, 0.5]$ as in the "softhard" case. The expected effect of "smoothing" the PMF profile at $\lambda = 0.5$ with the addition of a softcore potential for electrostatic interaction is not observed.

Observation (O5) led us to run a simulation with different values for the bounds b_e and b_v , in order to compare the obtained PMF and bin counts to the ones obtained with $b_e = b_v = 0.5$.

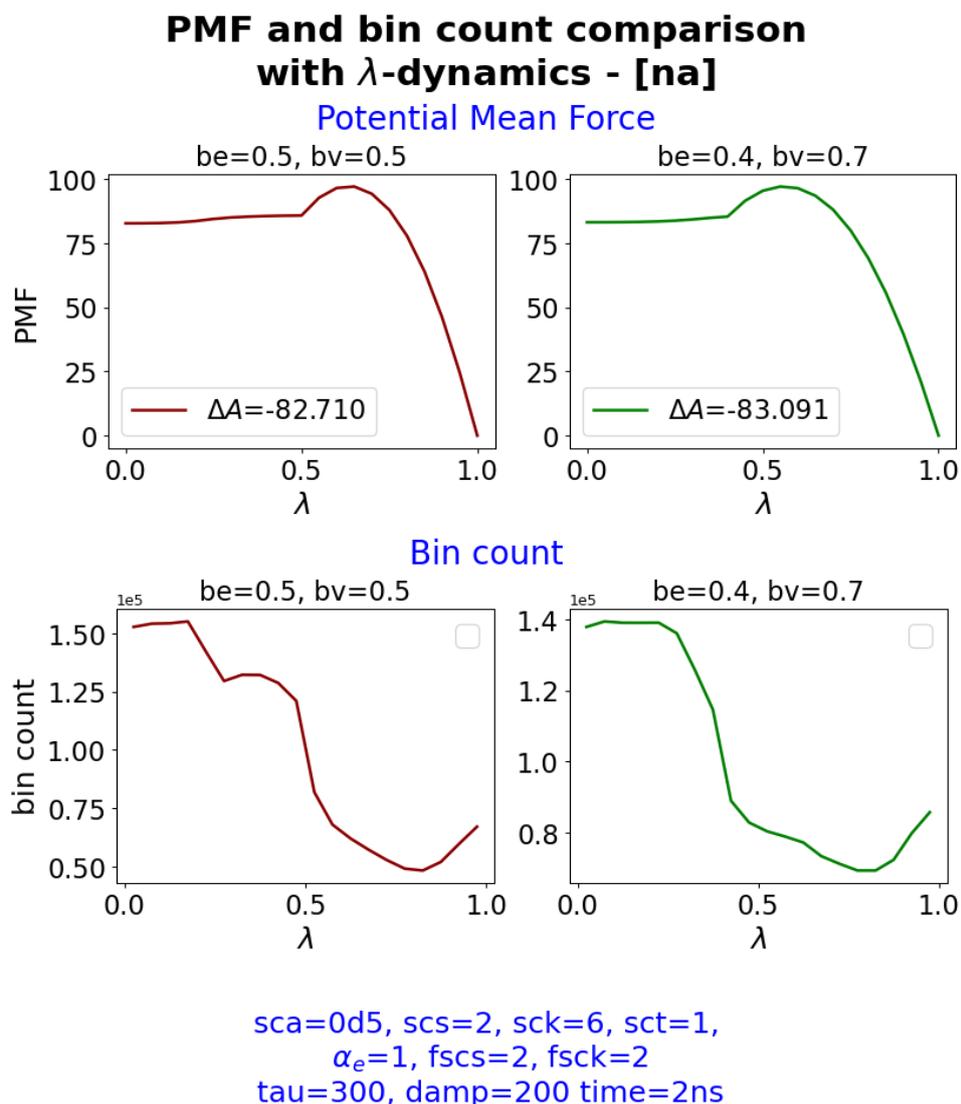


Figure 6.23 – Potential Mean Force profiles and bin counts for the hydration of a sodium cation, with parameter $fscs = s$ set to 2, and either $b_e = b_v = 0.5$ (left) or $b_e = 0.4$, $b_v = 0.7$ (right). The free energy difference $\Delta_{0 \rightarrow 1} A$ (in kcal.mol^{-1}) between final and initial state is printed.

Observations:

- (O6) Changing the values of the bounds b_e and b_v does seem to have an effect around $\lambda = 0.5$. However, it does not seem to significantly impact the sampling of the electrostatic region.

Conclusions The implementation of softcore potentials for van der Waals and electrostatic interactions allow for a fast computation of free energy differences, at least for the hydration

of a sodium cation and of a water molecule. The use of combined softcore potentials for both interactions instead of a softcore potential for only van der Waals interactions does not seem to have much impact on the computation efficiency. However, several questions still need to be answered before drawing any conclusions:

- (Q0) Is the mass m_λ we have chosen for our simulations adequate? Would using a mass of the order of M_{car} –provided the system does not explode– more appropriate in order to properly sample the electrostatic region?
- (Q1) Were the parameters of the electrostatic softcore potential (6.15) the optimal choice of parameters?
- (Q2) The use of softcore potentials combined to the λ -dynamics is promising to compute precise free energy differences, compared to the BAR method: can one estimate the gain in computation time?

until then, we can still conclude that one may opt for any of the two options, and still obtain a proper PMF profile in a sensibly short time.

6.5 Implementation of the OSRW method

6.5.1 Equations of motion

Now that the λ -dynamics has been implemented in the Tinker-HP code, let us sketch how one would design an adaptive biasing method in the manner of the OSRW method. Let us recall that we are interested in the two variables (λ, F_λ) , where $F_\lambda = -\partial_\lambda V(\mathbf{r}, \lambda)$ is the force acting on the variable λ . Thanks to the communication between the Tinker-HP and Colvars module codes, one now can make λ evolve dynamically. As shown in Figure 6.13, in a simulation run with the λ -dynamics, Colvars uses F_λ to propagate λ dynamically. However, F_λ is now a *dynamical variable* that one will have to propagate. To do so, one has to evaluate the force F^* acting on F_λ . This requires to compute $F^*(\mathbf{r}, \lambda) = \frac{\partial}{\partial F_\lambda} H_{ext}(\mathbf{r}, \mathbf{p}; \lambda)$, where H_{ext} is the extended Hamiltonian of the system, to be precised further down. Problem is, this computation is too difficult. One idea to bypass this issue is to apply an extended Adaptive Biasing Force method for the time-evolution of F_λ . In which case, we couple the variable F_λ to a classical extended coordinate ϕ using a spring of constant k . The extended Hamiltonian then becomes:

$$H_{ext}(\mathbf{r}, \mathbf{p}; \lambda, \phi) = E^{kin}(\mathbf{p}; \lambda) + U_{ext}(\mathbf{r}, \lambda, \phi)$$

where the extended potential is now

$$U_{ext}(\mathbf{r}, \lambda, \phi) = V_{ext}(\mathbf{r}; \lambda) + \frac{1}{2}k_\phi (F_\lambda(\mathbf{r}) - \phi)^2,$$

with V_{ext} given by (6.8). The force F_ϕ acting on ϕ is easily computable:

$$\begin{aligned} F_\phi &:= -\nabla \left(\frac{1}{2}k_\phi (F_\lambda(\mathbf{r}) - \phi)^2 \right) = -k_\phi (F_\lambda(\mathbf{r}) - \phi) \nabla (F_\lambda(\mathbf{r}) - \phi) \\ &= -k_\phi (F_\lambda(\mathbf{r}) - \phi) \nabla F_\lambda(\mathbf{r}). \end{aligned}$$

Note that the initial extended Hamiltonian is $E^{kin}(\mathbf{p}, \lambda) + V_{ext}(\mathbf{r}, \lambda)$, where λ is an **extended coordinate**, which is linked to a **fictitious particle** $\tilde{\lambda}$ of mass \mathbf{m}_λ , with the constraint that both are constantly equal. This is not the same idea we use while dealing with the variable F_λ , which is not an extended Cartesian coordinate, but a **classical collective variable**, linked to an extended variable ϕ of mass \mathbf{m}_ϕ via a spring of constant k_ϕ . It is the variable ϕ that is represented by a fictitious particle $\tilde{\phi}$ in the same way as λ . not F_λ . As a consequence, the extended Hamiltonian $H_{ext}(q, p; \lambda, \phi)$ will lead to **6N equations of motion treated by the Tinker-HP code**, and **2 additional equations of motion treated by Colvars**. One has:

$$\begin{cases} \ddot{\mathbf{r}} = -M^{-1}\nabla U_{ext} \\ \ddot{\lambda} = -\mathbf{m}_\lambda^{-1}\partial_\lambda U_{ext} \\ \ddot{\phi} = -\mathbf{m}_\phi^{-1}\partial_\phi U_{ext} \end{cases} .$$

Where, since: $U_{ext}(\mathbf{r}, \lambda, \phi) = V_{ext}(\mathbf{r}, \lambda) + \frac{1}{2}k_\phi (F_\lambda(r) - \phi)^2$:

$$\begin{aligned} \partial_\lambda U_{ext}(\mathbf{r}, \lambda, \phi) &= \partial_\lambda V_{ext}(\mathbf{r}, \lambda) + k (F_\lambda(\mathbf{r}) - \phi) \partial_\lambda F_\lambda(\mathbf{r}) \\ \nabla U_{ext}(\mathbf{r}, \lambda, \phi) &= \nabla V_{ext}(\mathbf{r}, \lambda) + k (F_\lambda(\mathbf{r}) - \phi) \nabla_{\mathbf{r}} F_\lambda(\mathbf{r}), \end{aligned}$$

hence the need to know both values of $\partial_\lambda F_\lambda = \frac{\partial^2 V}{\partial \lambda^2}$ and $\nabla F_\lambda = \nabla \partial_\lambda V$.

In practice, for a given time t , we first set the Cartesian derivatives of the force F_λ to zero. We then proceed to compute the forces and the Cartesian derivatives of the force F_λ . Let us recall the expression of ∇F_λ :

$$\begin{aligned} \nabla F_\lambda &= \nabla (-\partial_\lambda V(\mathbf{r}, \lambda)) = -\nabla (\partial_\lambda V_e(\mathbf{r}, \lambda) + \partial_\lambda V_v(\mathbf{r}, \lambda)) \\ &= -\nabla \partial_\lambda V_e(\mathbf{r}, \lambda) - \nabla \partial_\lambda V_v(\mathbf{r}, \lambda). \end{aligned}$$

Now, given the value $F_{t-\Delta t}^*$ of the force acting on F_λ at the previous timestep, one **locally** updates the total force acting on the system as follows:

$$\nabla H_{ext}^t(\mathbf{r}, \mathbf{p}; \lambda, \phi) = \nabla H_{ext}^{t-dt}(\mathbf{r}, \mathbf{p}; \lambda, \phi) + F_{t-dt}^* \times \nabla F_\lambda^{t-dt}(\mathbf{r}, \lambda).$$

6.5.2 Tinker-HP – Colvars interface: second update

Treating the variable F_λ as a dynamical variable led to several additions to the Tinker-HP–Colvars interface: let us quickly summarise them.

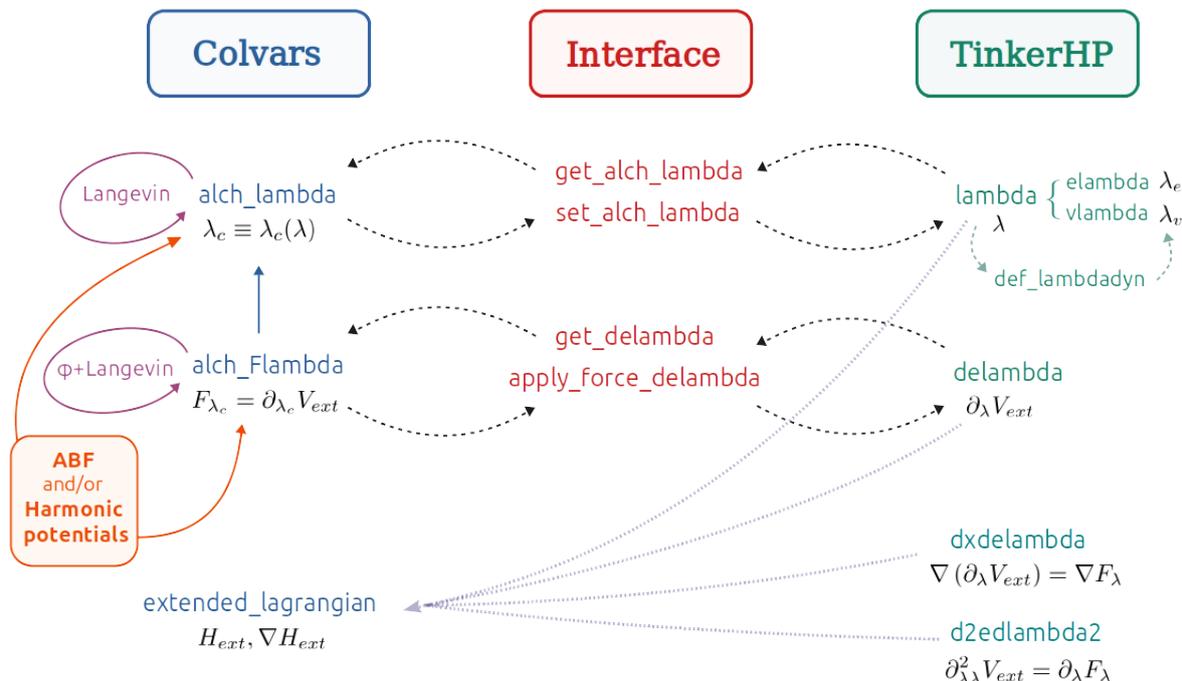


Figure 6.24 – Scheme of the communication between the Tinker-HP and Colvars module codes when running a simulation with the λ -dynamics method where both λ and F_λ evolve dynamically.

At a given time t , the Tinker-HP code does not only have a value of the variables λ and F_λ , but also have one of the derivative $\partial_\lambda F_\lambda$ (`delambda`) and the gradient ∇F_λ (`dxdelambda`). These values will be sent to Colvars in order to obtain that of the extended Hamiltonian at time t : similarly to the routines `get_alch_lambda` and `get_alch_Flambda`, interface routines have been introduced to do so. After having defined its internal variable $\lambda_c(\lambda)$ with the received alchemical λ (and the corresponding force F_{λ_c}), Colvars will propagate λ_c and the variable ϕ coupled to F_{λ_c} according to a Langevin dynamics, using leapfrog integration. The new values of λ and F_λ will be sent to the Tinker-HP program, which will then proceed to run the dynamics at time $t + dt$. Note that since λ and F_λ are now both dynamical, the Colvars module can apply harmonic potentials to both variables, along with using an Adaptive Biasing Force method.

6.5.3 Numerical results: what is done and what remains to be done

Implementing the OSRW method in the Tinker-HP code has proved to be trickier than expected. We ran simulations within the same framework as in Section 6.4.5: temperature was set to 300 K, simulations were run over 2 ns with a timestep of 1 fs, and the simulation box's length was of 18.643 Å. We will here sketch the approach that has been adopted, the issues we faced and how we intend to solve them. Let us recall that by "OSRW", one here means that the reaction coordinate considered is given by (λ, F_λ) , which both evolve dynamically.

6.5.3.1 OSRW with "softhard" λ -dynamics: inherent instabilities

We first ran simulations with a softcore potential for the van der Waal interactions only. Quick simulations were run in the NVE ensemble to check if double derivatives related to the dynamical evolution of F_λ were properly implemented: adding a harmonic potential on the variable λ , with strong spring force and centered in 0.7 was needed in order to observe energy conservation. Indeed, in practice, as the λ -dynamics start with $\lambda = 1$, values of the collective variable λ tends to be above the upper value 1, leading to explosion. Such a value for the harmonic potential's center was motivated by [87, Figure B, p.3] and [88, Figure 3 (a)].

We then came back to the NVT ensemble. However, first attempts to run simulations with dynamical (λ, F_λ) variables where met with quasi-instantaneous explosion of the system. One idea would be to change the boundary conditions required by the Colvars module for the propagation of both variables. We consequently switched from reflecting boundary conditions as introduced in Section 6.4.5 to the following ones:

$$\begin{cases} \frac{1}{2}k_{up}(\lambda - \lambda_{max})^2, & \lambda > \lambda_{max} \\ \frac{1}{2}k_{low}(\lambda - \lambda_{min})^2, & \lambda < \lambda_{min} \\ 0 & \text{else.} \end{cases} .$$

Namely, if λ is above a maximum value λ_{max} (resp. below a minimum value λ_{min}) set by the user, an harmonic potential of center λ_{max} (resp. λ_{min}) with constant k_{up} (resp. k_{low}) will be applied to λ . In practice, λ_{max} and λ_{min} are set up with the keywords `UpperWall` and `LowerWall`, and the constants k_{up} and k_{low} with the keywords `UpperWallConstant` and `LowerWallConstant`.

However, this change of boundary conditions did not have the expected impact on the stability of our simulations. Since these preliminary tests were run with our first implemented λ -dynamics, which used a softcore potential only for the van der Waals interactions, we decided to implement a softcore potential for electrostatic interactions, in the hope that this will prevent instabilities around $\lambda = 1$. Yet, an additional explanation could be that when λ goes outside of the interval $[0, 1]$, the force F_λ applied on λ is immediately set to 0 by the Tinker-HP code. This will lead the extended coordinate ϕ coupled to F_λ by Colvars to behave badly, quickly leading the system to explode. The question is then: how does one "mollify" the fall of F_λ to 0?

One idea would be to make everything go to 0 when λ approaches the boundary values 0 and 1. Another is to consider two different variables to treat λ : one "alchemical", which is the λ treated by the Tinker code, and another "dynamical" that is treated by the Colvars code. Right now, those are the same: at a given time step, Tinker gives the value of λ (which is then "alchemical") to Colvars, Colvars makes λ evolves dynamically (so that it is then "dynamical") and sends it back to Tinker for the next timestep. But one could change the expression of the dynamical variable, that we denote by λ_c , motivating the notation already used in Section 6.4.4 and 6.5.2. The dynamical variable λ_c treated by Colvars is now a function of the alchemical variable λ sent by Tinker, and defined on the whole \mathbb{R} space. We opt for the following dynamical variable:

$$\lambda_c(\lambda) := \frac{1}{\pi} \arccos(1 - 2\lambda),$$

so that:

$$\lambda = \frac{1 - \cos(\pi\lambda_c)}{2} \text{ if } \lambda_c \in [0, 1], \text{ and } 0 \text{ or } 1 \text{ otherwise.}$$

We will color any quantity computed by Tinker in purple, and any quantity computed in Colvars in teal. Colvars will then consider the force that is applied not to λ , but to the dynamical λ_c , defined as:

$$F_{\lambda_c} := \frac{\partial V_{ext}(r, \lambda)}{\partial \lambda} \cdot \frac{d\lambda}{d\lambda_c},$$

with

$$\frac{d\lambda}{d\lambda_c} = \frac{\pi}{2} \sin(\pi\lambda_c) \text{ if } \lambda_c \in [0, 1], \text{ and } 0 \text{ otherwise.}$$

In order to make λ_c and F_{λ_c} evolve dynamically, the Colvars code also needs the force applied on F_{λ_c} . We will consequently need:

$$\begin{aligned} \nabla F_{\lambda_c} &= \nabla \left(\frac{\partial V_{ext}(r, \lambda)}{\partial \lambda} \cdot \frac{d\lambda}{d\lambda_c} \right) = \nabla F_{\lambda} \frac{d\lambda}{d\lambda_c} \\ \frac{\partial F_{\lambda_c}}{\partial \lambda_c} &= \frac{\partial}{\partial \lambda_c} \left(\frac{\partial V_{ext}(r, \lambda)}{\partial \lambda} \cdot \frac{d\lambda}{d\lambda_c} \right) = \frac{\partial^2 V_{ext}(r, \lambda)}{\partial \lambda^2} \cdot \left(\frac{d\lambda}{d\lambda_c} \right)^2. \end{aligned}$$

The PMF generated by the Colvars code will give the free energy derivative with respect to λ_c :

$$\frac{dA}{d\lambda_c} = \left\langle \frac{\partial H_{ext}}{\partial \lambda_c} \right\rangle, \quad (6.21)$$

One may want to use thermodynamic integration to obtain the more conventional PMF \tilde{A} with respect to the alchemical λ of Tinker:

$$\frac{d\tilde{A}}{d\lambda} = \frac{dA}{d\lambda_c} \cdot \frac{d\lambda_c}{d\lambda},$$

with

$$\frac{d\lambda_c}{d\lambda} = \frac{1}{\pi} \frac{1}{\sqrt{\lambda(1-\lambda)}}. \quad (6.22)$$

Note that as λ goes towards 0 or 1, (6.22) goes towards $+\infty$. Therefore one will be quickly facing the issue of multiplying $+\infty$ by 0. As such, it would be better to:

(A) Run a classical λ -dynamics, obtain the PMF (6.21) with respect to the dynamical λ_c .

(B) Run a BAR test with several fixed values of the alchemical λ in order to build a (intermediary) PMF with respect to the alchemical λ , that we will denote by $\frac{d\tilde{A}}{d\lambda}$.

(C) Bring together the PMF of the λ -dynamics simulation of step **(A)** with respect to the alchemical λ :

$$\frac{dA^{\text{MBAR}}}{d\lambda_c} = \frac{d\tilde{A}}{d\lambda} \frac{d\lambda}{d\lambda_c}$$

which is much easier and avoids singularities.

This would allow for a direct comparison of $A(\lambda_c)$ and $A^{\text{MBAR}}(\lambda_c)$, which are two estimators

of the same quantity. Note however that this would imply for the user to post-treat several simulations. Proper python scripts should be made accessible in order to ease these steps.

Notes on the biasing of F_λ – All of the above is necessary for the extended dynamics on F_λ if we need a 2d free energy surface on λ_c and F_λ . This can be avoided by applying a different type of bias on F_λ , either metadynamics, or a constant bias $V^b(F_\lambda)$. This may be sufficient to enhance the sampling of the orthogonal space. Then to reconstruct the unbiased derivative $\frac{dA^{\text{MBAR}}}{d\lambda_c}$, we need the 2d-histogram $\rho(\lambda_c, F_\lambda)$ of the (Step A) simulation (with or without a bias along λ_c).

Then the unbiased free energy derivative is the reweighted conditional average:

$$\frac{dA}{d\lambda_c} = - \frac{\int F_\lambda e^{+\beta V^b(F_\lambda)} \rho(\lambda_c, F_\lambda) dF_\lambda}{\int e^{+\beta V^b(F_\lambda)} \rho(\lambda_c, F_\lambda) dF_\lambda}$$

Note that biasing along λ_c does not bias conditional averages on λ_c (this property is used in ABF).

The definition of λ_c as a function of the alchemical λ in the Colvars module has not been implemented yet, as one hopes the use of softcore potentials for both van der Waals and electrostatic interactions would resolve the issue of system explosion. For now, $\lambda_c(\lambda) \equiv \lambda$ in the Colvars code, and as such, the following question is yet to be answered:

(Q3) Is the implementation of λ_c as a function of λ in the Colvars module necessary to treat the system's inherent instability due to the strong oscillations of F_λ ?

6.5.3.2 OSRW with "softsoft" λ -dynamics

Before attempting to implement λ_c in the Colvars module, we started to confront the use of a softcore potential for electrostatic interactions in the OSRW method. We first ran simulations in the NVE ensemble to check for energy conservation. The energy is conserved provided an harmonic potential is applied on λ , with strong spring force and center in 0.7: without it, the system explodes. In order to apply a harmonic spring to the collective variable F_λ , one first needs to identify its behaviour: we generated a 2d histogram of (λ, F_λ) by running a simulation over 0.1 ns with timestep of 1 fs in the NVE ensemble, for the solvation of a sodium cation.

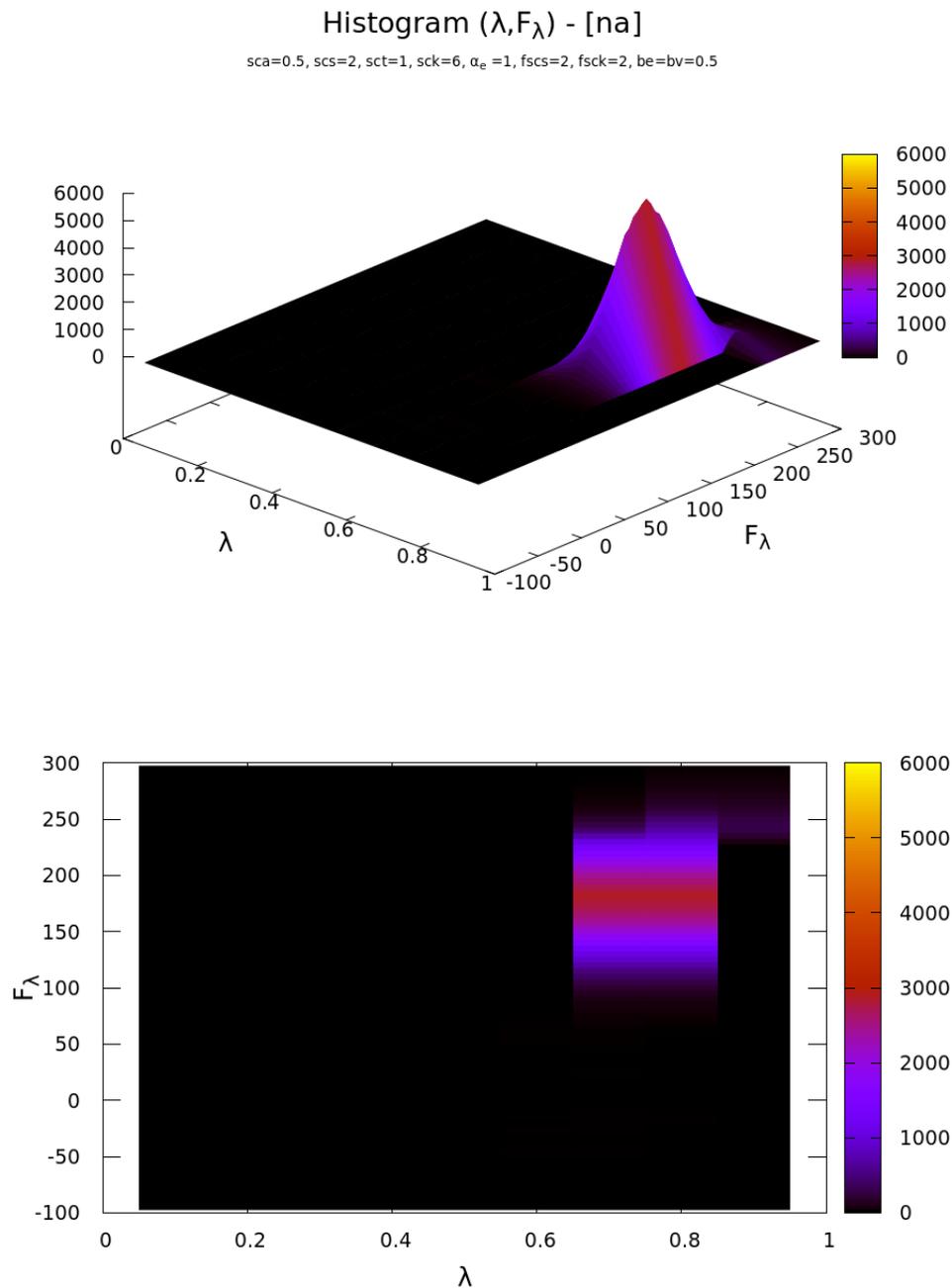


Figure 6.25 – Histogram of (λ, F_λ) with λ -dynamics run with softcore potentials for van der Waals and electrostatic interactions, and harmonic potential on λ centered in 0.7.

One can make out that a good harmonic potential to apply on F_λ if λ is restricted to 0.7 would have a center of around 180. However, until now, doing so does not ensure the energy conservation. One should consequently:

- (Q4) Ensure the softcore derivatives needed for the dynamical evolution of the collective variable F_λ are correctly implemented in the Tinker-HP code.
- (Q5) If question (Q1) is answered, and if the energy is still not conserved in a NVE simulation, one should try to apply an adequate harmonic potential on F_λ .
- (Q6) If solution to (Q2) still fails to ensure energy conservation, then one may proceed to the implementation of λ_c as a function of λ in the Colvars module, given by (6.5.3.1).

Provided tackling (Q1)–(Q3) solves the issue of energy conservation for an OSRW simulation, one may then attempt to tackle our final open problem:

- (Q7) Can one apply the Adaptive Biasing Force method on both coordinates (λ, F_λ) ? Is our algorithm as efficient as the original OSRW algorithm suggested by W. Yang and colleagues?
- (Q8) Is it also more efficient than the BAR method? If so, can one estimate the gain in computation time compared to the BAR method, and compared to our implemented λ -dynamics method?

Work is in progress to answer questions (Q0)–(Q8), and one may hope that the implementation of the λ -dynamics, along with the use of the ABF method with the reaction coordinate (λ, F_λ) , will enable the computation of alchemical free energies in a fast, efficient and robust manner. If this proves to be the case, the resulting method would allow the user to obtain a 2d free energy profile, providing new insights on the system at hand. One may then wish to test the implemented OSRW method on greater systems, with the long-term goal of estimating a ligand's affinity with a given receptor.

Appendix **A**

Notes on stochastic processes

A.1 Markov processes

A *Markov process* is a family $(X_t)_{t \geq 0}$ of random variables on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a set E equipped with a σ -field \mathcal{E} . We assume the measurable space (E, \mathcal{E}) to be satisfying the *measure decomposition theorem*: given a measure μ on $(E \times E, \mathcal{E} \otimes \mathcal{E})$, if μ^x denotes its projection on the first coordinate x , then one has the following decomposition:

$$\mu(dx, dy) = \kappa(x, dy)\mu^x(dx),$$

for some kernel κ [5]. The measure decomposition theorem will be crucial in our case. Indeed, as said in Section 1.4.1, one can decompose the canonical measure as the product of two measures, as done in equation (1.8). Such decomposition will prove to be the key to ease the study of many algorithms used in the scope of molecular dynamics. The process $(X_t)_{t \geq 0}$ starting from the initial point $x \in E$ is said to be Markovian if it satisfies the *Markov property*:

[M1] If one denotes by $\{\mathcal{F}_t\}_{t \geq 0} := \{\sigma(X_s | s \leq t)\}_{t \geq 0}$ the natural filtration of $(X_t)_{t \geq 0}$, then, for all $A \in \mathcal{E}$, for all $s < t$:

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s).$$

A.2 Markov semigroups

There exists a natural semigroup associated to the Markov process $(X_t)_{t \geq 0}$. Before defining it, let us first consider a family of operators $(P_t)_{t \geq 0}$ defined on real-valued measurable functions f satisfying the following conditions:

[SG1] For every $t \geq 0$, P_t is a linear operator which maps the set \mathcal{B} of bounded measurable functions defined on (E, \mathcal{E}) , onto itself.

[SG2] $P_0 = Id$

[SG3] $P_t(\mathbf{1}) = \mathbf{1}$ where $\mathbf{1}(x) = 1, \quad \forall x \in E$.

[SG4] If $f \geq 0$ then $P_t f \geq 0$.

[SG5] For all $t, s \geq 0, P_{t+s} = P_t \circ P_s$.

[SG6] There exists an *invariant measure*, namely a positive, σ -finite measure μ on (E, \mathcal{E}) such that for all measurable function $f : E \rightarrow \mathbb{R}$ which is bounded and positive, for every $t \geq 0$, one has:

$$\int_E P_t f \, d\mu = \int_E f \, d\mu.$$

Furthermore, $(P_t)_{t \geq 0}$ is continuous at $t = 0$, in other words, for every $f \in L^2(\mu)$,

$$\lim_{t \rightarrow 0} P_t f = f.$$

[SG7] For every $1 \leq p < +\infty$, the family of operators $(P_t)_{t \geq 0}$ can be extended as bounded (i.e contraction) operators on the $L^p(\mu)$ space.

A semigroup $(P_t)_{t \geq 0}$ satisfying properties **[SG1–SG7]** is called a *Markov semigroup*. The natural, Markov semigroup associated to the Markov process $(X_t)_{t \geq 0}$ is defined as:

$$\forall t \geq 0, \forall x \in E, \quad P_t f(x) := \mathbb{E}_x [f(X_t)] = \mathbb{E} [f(X_t) | X_0 = x],$$

where \mathbb{E} is taken with respect to the Wiener measure associated to $(W_t)_{t \geq 0}$, f is a test function usually taken to be either bounded or measurable and positive.

Any Markov semigroup can be represented by the *probability kernels* of the Markov process $(X_t)_{t \geq 0}$. The family $(p_t)_{t \geq 0}$ are probability kernels if:

- (i) For all $x \in E$, for all $t \geq 0$, $p_t(x, \cdot)$ is a probability measure.
- (ii) For all set $A \in \mathcal{E}$, $t \mapsto p_t(x, A)$ is measurable.

The probability $p_t(x, \cdot)$ is consequently the distribution at time $t \geq 0$ of the process $(X_t)_{t \geq 0}$ starting in point $x \in E$. Under the assumption that the space (E, \mathcal{E}) is nice enough, the associated Markov semigroup satisfies, for any bounded measurable function f on E :

$$\forall t \geq 0, \forall x \in E, \quad P_t f(x) = \int_E f(y) p_t(x, dy).$$

Provided the density kernels have densities with respect to the Lebesgue measure (or any other measure) λ , which we will also denote by $(p_t)_{t \geq 0}$, one has for any bounded measurable function f on E :

$$\forall t \geq 0, \forall x \in E, \quad P_t f(x) = \int_E f(y) p_t(x, y) d\lambda(y).$$

The invariant measure μ is said to be *reversible with respect to the semigroup* $(P_t)_{t \geq 0}$ if, for all functions $f, g \in L^2(\mu)$, and all $t \geq 0$,

$$\int_E f P_t g \, d\mu = \int_E g P_t f \, d\mu. \quad (\text{A.1})$$

A direct consequence of (A.1) is that the densities of the probability kernels $(p_t)_{t \geq 0}$ are symmetric on $E \times E$. The semigroup $(P_t)_{t \geq 0}$ is thus said to be *symmetric with respect to* μ .

A.3 Infinitesimal generator

Since the semigroup $(P_t)_{t \geq 0}$ is by construction a contraction semigroup, one can apply the Hille-Yosida theorem [5, Annex A.1]: there exists a dense linear subset \mathcal{D} of $L^2(\mu)$, where μ is the invariant measure given in [SG6], called the *domain* of the semigroup, on which the derivative

$$\lim_{t \rightarrow 0} \frac{1}{t} (P_t f - f) := \left. \frac{dP_t}{dt} \right|_{t=0},$$

exists for all $f \in \mathcal{D}$ and is in $L^2(\mu)$.

The *infinitesimal generator* \mathcal{L} of domain $\mathcal{D}(\mathcal{L}) = \mathcal{D}$ of the associated Markov process $(X_t)_{t \geq 0}$ is then defined as:

$$\begin{aligned} \mathcal{L} : \mathcal{D} &\longrightarrow L^2(\mu) \\ f &\longmapsto \left. \frac{dP_t}{dt} \right|_{t=0} \end{aligned}$$

The infinitesimal generator \mathcal{L} is *symmetric with respect to* μ if, for all $f, g \in \mathcal{D}(\mathcal{L})$,

$$\int_E f \mathcal{L} g \, d\mu = \int_E g \mathcal{L} f \, d\mu.$$

One can show that the semigroup $(P_t)_{t \geq 0}$ satisfies the *heat equation* with respect to \mathcal{L} : for every $t \geq 0$,

$$\partial_t P_t = \mathcal{L} P_t = P_t \mathcal{L}. \quad (\text{A.2})$$

A.4 Fokker-Planck-Kolmogorov equation

Since the semigroup $(P_t)_{t \geq 0}$ can be represented by probability kernels, we rewrite the two formulations of equation (A.2) using the density of the kernels. The equation

$$\partial_t P_t = \mathcal{L} P_t$$

yields that for all $x \in E$, for all $t \geq 0$,

$$\begin{cases} \partial_t p_t(x, y) = \mathcal{L}_x p_t(x, y) \\ p_0(x, y) d\lambda(y) = \delta_x \end{cases},$$

where \mathcal{L}_x represents the action of \mathcal{L} on the variable x . We usually prefer to use the dual formulation of this equation, namely $\partial_t P_t = P_t \mathcal{L}$: for all $x \in E$, for all $t > 0$,

$$\partial_t p_t(x, y) = \mathcal{L}^* p_t(x, y), \quad (\text{A.3})$$

with $p_0(x, \cdot)$ as initial density, and where \mathcal{L}^* is the adjoint of the operator \mathcal{L} with respect to the Lebesgue measure λ (or any reference measure):

$$\int_E f \mathcal{L} g \, d\lambda = \int_E g \mathcal{L}^* f \, d\lambda.$$

The equation (A.3) is called the *Fokker-Planck equation*.

Remark 29. Let us denote by $(P_t^*)_{t \geq 0}$ the adjoint of the semigroup $(P_t)_{t \geq 0}$. In both cases, the actions of the operators $(P_t)_{t \geq 0}$ and $(P_t^*)_{t \geq 0}$ yield solutions of an initial value problem. This motivates the use of the notations $e^{t\mathcal{L}} = P_t$ and $e^{t\mathcal{L}^*} = P_t^*$.

A.5 Hypocoellipticity

Provided the operator $\partial_t - \mathcal{L}^*$ is hypoelliptic, the solution to the Fokker-Planck equation (A.3) π_t is smooth and bounded, irrespective of the initial condition [5, Section 1.12]. An operator \mathcal{A} is said to be hypoelliptic if for any solution π of the equation $\mathcal{A}\pi = f$ for a given f , then

$$f \in H_s^{\text{loc}} \Rightarrow \pi \in H_{s+\varepsilon}^{\text{loc}}, \quad \forall \varepsilon > 0,$$

where H_s^{loc} denotes the local Sobolev space of order $s \in \mathbb{N}$. Hypocoellipticity is given by the following *Hörmander theorem*: [5, Proposition 1.12.1]

Theorem 4. Consider the following differential operator

$$\mathcal{L} = a_0 \cdot \nabla + \sum_{i=1}^N (a_i \cdot \nabla)^\top (a_i \cdot \nabla), \quad (\text{A.4})$$

where the $(a_i)_{i \in \llbracket 1, N \rrbracket}$ are smooth vector field in \mathbb{R}^d , and \top denotes the adjoint with respect to the $L^2(\lambda)$ -scalar product. Consider the following family of vector fields:

$$\begin{aligned} V_0 &:= \text{span} \{a_i \mid i \in \llbracket 0, N \rrbracket\}, \\ V_{n+1} &:= \text{span} \{V_n \cup \{[v, a_i] \mid v \in V_n, i \in \llbracket 1, N \rrbracket\}\}, \quad \forall n \in \llbracket 0, N \rrbracket, \end{aligned}$$

where for two smooth vector fields X, Y in \mathbb{R}^d one has:

$$[X, Y] = \nabla Y \cdot X - \nabla X \cdot Y.$$

If the following parabolic Hörmander condition

$$\exists k^* \in \llbracket 0, N \rrbracket, \quad V_{k^*} = \mathbb{R}^d, \quad (\text{A.5})$$

is satisfied, then the operator \mathcal{L} is hypoelliptic.

Remark 30. When considering differential operators of the form $\partial_t - \mathcal{L}$ where \mathcal{L} is of the form (A.4), one needs to consider $a_0 = -a_0 + \partial_t$ and $a_i = -a_i$ for all $i \in \llbracket 1, N \rrbracket$. The parabolic Hörmander condition is verified this time with the set $V_0 := \text{span} \{a_i \mid i \in \llbracket 1, N \rrbracket\}$, and the definition of the sets $(V_k)_{k \in \llbracket 1, N+1 \rrbracket}$ do not change. As a consequence, the term a_0 no longer intervene in the parabolic Hörmander condition.

A.6 Diffusion processes

A (time-homogeneous) diffusion process on \mathbb{R}^d starting from $x \in \mathbb{R}^d$ is a process $(X_t)_{t \geq 0}$ satisfying the following stochastic differential equation:

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t) dW_t \\ X_0 = x, \end{cases} \quad (\text{A.6})$$

where $(W_t)_{t \geq 0}$ is standard d -dimensional Brownian motion, b is a vector, σ a $d \times d$ matrix. We assume b and σ to be smooth functions with bounded derivatives.

The associated infinitesimal generator is :

$$\mathcal{L} = \sum_{i=1}^d b_i \partial_i + \frac{1}{2} \sum_{i,j=1}^d a_{i,j} \partial_{ij}^2,$$

where $a = \sigma \sigma^\top$.

Assume (A.6) admits a strong solution, and that there exists an invariant measure μ whose density π satisfying $0 = \mathcal{L}^* \pi$ is positive and smooth. Provided the infinitesimal generator \mathcal{L} satisfies the Hörmander parabolic condition (A.5), then:

- (i) The $(X_t)_{t \geq 0}$ is *irreducible*, namely, for any Borel set A such that $\lambda(A) > 0$, for λ -almost all $x_0 \in \mathbb{R}^d$, for all $t > 0$,

$$P_t \mathbf{1}_A x_0 > 0.$$

In other words, the process can reach any subset of \mathbb{R}^d with positive probability, whatever the time and initial condition considered.

- (ii) The process is *ergodic*: for any observable $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^d)$, its trajectorial average converge in the long-time limit towards its statistical mean:

$$\lim_{T \rightarrow +\infty} \int_0^T \varphi(X_t) dt = \mathbb{E}_\mu[\varphi].$$

- (iii) The invariant measure μ is unique.

[5] .

Appendix **B**

Some notions in analytical mechanics

Classical (or *Newtonian*) mechanics rely on a vectorial description of physical systems. Its foundations are based on the concept of mass points, where we assume one can describe a given physical system's constituents with a mass and a point in an Euclidean space. This description implies working with vector quantities of motion such as the velocity, the acceleration, forces, as well as momentum. But one could be interested in working with scalar quantities of motions such as the system's kinetic energy, a particle's moment of inertia and so on. Such is the purpose of analytical mechanics, which by construction allows a far more general description of complicated physical systems.

B.1 Some reminders of classical mechanics

From now on, we will denote by $\dot{\varphi}$ the time derivative of any quantity φ , and quantities written in bold should be seen as vectorial.

Let us consider a system of N mass points $(X_i, m_i)_{i \in \llbracket 1, N \rrbracket}$ where X_i is a point in the Euclidean space \mathbb{R}^3 equipped with the orthogonal basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$, and m_i is its mass. For all i in $\llbracket 1, N \rrbracket$, the point X_i will be described by its coordinates (x_i, y_i, z_i) in the inertial Cartesian axis system $Oxyz$. We will denote by \mathbf{r}_i the vector $\mathcal{O}X_i$, and by $\mathbf{v}_i = \dot{\mathbf{r}}_i$ the velocity of the point i . By $\nabla_{\mathbf{r}_i}$ we denote the vector $(\partial_{x_i}, \partial_{y_i}, \partial_{z_i})^\top$.

B.1.1 Newton's second law

Classical mechanics rely on the *d'Alembert principle*, also called *Newton's second law*:

$$\mathbf{F}_i = \frac{d}{dt} (m_i \mathbf{v}_i), \quad (\text{B.1})$$

which states that the net force on the point i is equal to the time derivative of the point's momentum $m_i \mathbf{v}_i$. To facilitate the reading of the following sections, and avoid any misunderstanding,

we will for now prevent ourselves from using the notation \mathbf{p} for the momentum.

For a system of N particles, one would need to solve $3N$ equations, as one needs to project (B.1) against the three Cartesian axis.

B.1.2 Work and kinetic energy

Let us consider a mass point (X, m) in space, of mass m , position \mathbf{r} and velocity $\mathbf{v} = \dot{\mathbf{r}}$. Let us consider two positions \mathbf{r} and $\tilde{\mathbf{r}}$. The energy produced along a displacement (or deformation) of the mass point when a force \mathbf{F} is applied is called the work $W_{\mathbf{r} \rightarrow \tilde{\mathbf{r}}}$ of the force \mathbf{F} . One has:

$$W_{\mathbf{r} \rightarrow \tilde{\mathbf{r}}} = \int_{\mathbf{r}}^{\tilde{\mathbf{r}}} \mathbf{F} \cdot d\mathbf{r} = E_c(\tilde{\mathbf{r}}) - E_c(\mathbf{r}),$$

where $E_c(\mathbf{r}) = \frac{1}{2}m\mathbf{v}$ is the kinetic energy of the mass point at position \mathbf{r} .

Now let us consider our system of N mass points $(X_i, m_i)_{i \in \llbracket 1, N \rrbracket}$. The net force \mathbf{F}_i applied to the i -th mass point is the sum of external forces applied to the point and interaction forces with the other mass points. Hence:

$$\mathbf{F}_i = \mathbf{F}_i^{ext} + \sum_{j \neq i} \mathbf{F}_{ji},$$

where $\mathbf{F}_{ji} = -\mathbf{F}_{ij}$ is the interaction force that the mass point j applies to the mass point i .

If the point i moves by an elementary displacement $d\mathbf{r}_i$ then the elementary work of the force \mathbf{F}_i is given by

$$dw_i = \mathbf{F}_i^{ext} \cdot d\mathbf{r}_i + \sum_{j \neq i} \mathbf{F}_{ji} \cdot d\mathbf{r}_i = d\left(\frac{1}{2}m_i v_i^2\right),$$

where v_i stands for the norm $\|\mathbf{v}_i\|$. Hence the total elementary work of the system is given by:

$$dW = \sum_{i=1}^N dw_i = \sum_{i=1}^n d\left(\frac{1}{2}m_i v_i^2\right) = d\sum_{i=1}^n \left(\frac{1}{2}m_i v_i^2\right)$$

and one can conclude that the kinetic energy E_c of the whole system is equal to the sum of the kinetic energies $(E_c^i)_{i \in \llbracket 1, N \rrbracket}$ of the N mass points.

B.1.3 Conservative forces

A **conservative system** is a physical system in which the work W done by a force \mathbf{F} between two positions \mathbf{r} and $\tilde{\mathbf{r}}$ is independant of the chosen path, is reversible, and is equal to the difference between the initial and final value of a given energy. Let us work within such a system.

A direct consequence is that the force \mathbf{F}_i applied on the mass point i can be written as the derivative of a potential energy V_i :

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} V_i.$$

Just like the kinetic energy, if V_i denotes the potential energy from which the force \mathbf{F}_i derives from, then the potential energy of the total system V can be written as the following sum:

$$V = \sum_{i=1}^N V_i,$$

and $\mathbf{F}_i = -\nabla V$.

Note that the elementary work done by the force \mathbf{F}_i along an elementary displacement $d\mathbf{r}_i$ is equal to

$$dw_i = \mathbf{F}_i \cdot d\mathbf{r}_i = -\nabla_{\mathbf{r}_i} V_i \cdot d\mathbf{r}_i = -dV_i,$$

hence the total work of the system between two positions \mathbf{r} and $\tilde{\mathbf{r}}$ is given by:

$$W_{\mathbf{r} \rightarrow \tilde{\mathbf{r}}} = \sum_{i=1}^N \int_{\mathbf{r}}^{\tilde{\mathbf{r}}} dw_i = - \sum_{i=1}^N \int_{\mathbf{r}}^{\tilde{\mathbf{r}}} dV_i = \sum_{i=1}^N (V_i(\tilde{\mathbf{r}}) - V_i(\mathbf{r})) = V(\tilde{\mathbf{r}}) - V(\mathbf{r}).$$

B.2 Foundations of analytical mechanics

B.2.1 Relationship between potential and kinetic energies

We have, starting from vectorial quantities in classical mechanics, constructed scalar quantities: the system's total potential and kinetic energies. Let us establish a nice relationship between those. By Newton's second law,

$$\mathbf{F}_i = \frac{d}{dt} (m_i \mathbf{v}_i) = -\nabla_{\mathbf{r}_i} V_i$$

and since

$$\begin{aligned} \nabla_{\mathbf{v}_i} E_c &= \sum_{j=1}^N \nabla_{\mathbf{v}_i} E_c^j = \sum_{j=1}^N \nabla_{\mathbf{v}_i} \left(\frac{1}{2} m_j v_j^2 \right) \\ \nabla_{\mathbf{v}_i} E_c &= \sum_{j=1}^N m_j v_j, \end{aligned}$$

one gets:

$$-\nabla_{\mathbf{r}_i} V = \frac{d}{dt} \nabla_{\mathbf{v}_i} E_c. \quad (\text{B.2})$$

Note that the quantities involved in equality (B.2) involves only scalar quantities: it is a reformulation of Newton's second law and lays the foundation of analytical mechanics.

B.2.2 Lagrangian

We will now define a function of both position \mathbf{r} and speed $\mathbf{v} = \dot{\mathbf{r}}$, called the Lagrangian, using both potential and kinetic energies:

$$\mathcal{L}(\mathbf{r}, \mathbf{v}) = E_c(\mathbf{v}) - V(\mathbf{r}).$$

Given equation (B.2), one can write the generalised Lagrangian equation:

$$\nabla_{\mathbf{r}}\mathcal{L} = \frac{d}{dt}\nabla_{\mathbf{v}}\mathcal{L}, \quad (\text{B.3})$$

in other words, for a given mass point $i \in \llbracket 1, N \rrbracket$:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}_i}, \quad \frac{\partial \mathcal{L}}{\partial y_i} = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{y}_i}, \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial z_i} = \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{z}_i}.$$

B.2.3 Constraints, degrees of freedom and generalised coordinates

Now, for a given system of N particle, if all of the mass point's Cartesian coordinates $\{(x_i, y_i, z_i)\}_{i \in \llbracket 1, N \rrbracket}$ are independent, in regard of equation (B.3), one has to solve $3N$ equations. Most of the time however, some coordinates are not independent and are linked by equations called **mecanical constraints**. There are two kind of constraints:

- (i) **holonomic constraints**, which are algebraic equations of the form

$$f(\mathbf{r}_1, \dots, \mathbf{r}_N; t) = 0.$$

If the equation is time-dependent then the constraint is **rheonomic**, whereas if it is not, it is called **scleronomic**.

- (ii) **non-holonomic constraints**, which stands for every constraint that is not holonomic. Among them are the **semi-holonomic** constraints, which depend on both positions and velocities and are of the form

$$g(\mathbf{r}_1, \dots, \mathbf{r}_N; \dot{\mathbf{r}}_1, \dots, \dot{\mathbf{r}}_N; t).$$

In the case of semi-holonomic constraints, the method of Lagrange multipliers can be applied to incorporate the constraints to the equations of motion. Assume one has K semi-holonomics constraints $(g_k)_{k \in \llbracket 1, K \rrbracket}$. Then it can be shown, with a slight abuse of notations, that the Lagrangian of the extended system can be written as:

$$\mathcal{L}(\mathbf{r}, \mathbf{v}; \lambda, t) = \mathcal{L}(\mathbf{r}, \mathbf{v}; t) + \sum_{k=1}^K \lambda_k(t) g_k(\mathbf{r}_1, \dots, \mathbf{r}_N; t),$$

where λ_k is the Lagrange multiplier associated to the k -th semi-holonomic constraint.

We will from now on work only with holonomic constraints, and define the number of **degrees of freedom** g , which is the number of independent coordinates. Provided there exists a finite number l of constraints, one would have $g = 3N - l$.

Let us assume our system has $g = 3N - l$ degrees of freedom. We will reorganise the coordinates and renumerate them so that the g first coordinates are independent, the l remaining coordinates being linked by constraints. We ditch the classical set of Cartesian coordinates

$$\{(x_1, y_1, z_1), \dots, (x_N, y_N, z_N)\}$$

for the following set of coordinates:

$$\{q_1, \dots, q_g; q_{g+1}, \dots, q_{g+l}\},$$

where

$$\left\{ \begin{array}{l} q_1 = q_1(q_1, 0, \dots, 0; t) \\ \vdots \\ q_g = q_g(0, \dots, 0, q_g; t) \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} q_{g+1} = q_{g+1}(q_1, \dots, q_g; t) \\ \vdots \\ q_{g+l} = q_{g+l}(q_1, \dots, q_g; t) \end{array} \right.$$

By construction, one can get rid of the l coordinates $(q_{g+1}, \dots, q_{g+l})$ and work only with the so-called **generalised coordinates** (q_1, \dots, q_g) . These new, independant coordinate constitute a basis in which one can rewrite the positions $(\mathbf{r}_i)_{i \in \llbracket 1, N \rrbracket}$ of the N mass points as $\mathbf{r}_i = \mathbf{r}_i(q_1, \dots, q_g; t)$ for all $i \in \llbracket 1, N \rrbracket$. Note that generalised coordinates are not bound to be Cartesian coordinates. One can work with angles, lengths, and so on.

Remark 31. *As a consequence, one can rewrite, for the i -th particle:*

(i) *The velocity \mathbf{v}_i as:*

$$\mathbf{v}_i = \frac{d}{dt} \mathbf{r}_i = \sum_{k=1}^g \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t}.$$

(ii) *The total differential of the position as:*

$$d\mathbf{r}_i = \sum_{k=1}^g \frac{\partial \mathbf{r}_i}{\partial q_k} dq_k + \frac{\partial \mathbf{r}_i}{\partial t} dt.$$

(iii) *The virtual displacement of the particle i provided there is no virtual variation of the time variable as:*

$$\delta \mathbf{r}_i = \sum_{k=1}^g \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k.$$

Remark 32. *Having non-holonomic constraints does not ensure a reduction of the number of degrees of freedom! Indeed, each given holonomic constraint allows a rewriting of the dependent coordinates as a function of the generalised coordinate. It is a priori not the case for non-holonomic constraint, where positions and velocities can be linked.*

Example 1. *A classical example is the pendulum of length L in \mathbb{R}^2 . The position of the pendulum is given by $\mathbf{r} = (x, y)^\top$, and the angle between the pendulum and the Oy axis is denoted by θ . One has $\|\mathbf{r}\|^2 = x^2 + y^2 = L^2$, and the number of degrees of freedom is $g = 2 - 1 = 1$. The generalised coordinate to consider is consequently θ , and one can rewrite $x = L \sin(\theta)$, $y = L \cos(\theta)$.*

From now on, we will assume the system has g degrees of freedom and we will denote by (q_1, \dots, q_g) the associated generalised coordinate. The vector space defined by the generalised coordinates is called the **configuration space**. The **state space**, or **phase space**, is the space of all points $(q, p) \in \mathbb{R}^{2g}$ which characterise all of the system's possible states.

B.2.4 Conjugate momentum

When working with Cartesian coordinates, one had $m\mathbf{v} = \nabla_{\mathbf{r}}E_c$. A logical thing to do is to define a similar object for generalised coordinates. The **generalised momentum**, or **conjugate momentum** p_k of the generalised coordinate q_k , $k \in \llbracket 1, g \rrbracket$ is given by:

$$p_k = \frac{\partial E_c}{\partial q_k}, \quad \forall k \in \llbracket 1, g \rrbracket. \quad (\text{B.4})$$

B.2.5 Generalised forces

Provided the system is at static equilibrium, the virtual work δW of the applied force \mathbf{F} is null for all virtual displacements $\delta \mathbf{r}$. Let us write this statement in the terms of classical mechanics:

$$\delta W = \sum_{i=1}^N \mathbf{F}_i \cdot \delta \mathbf{r}_i = 0.$$

Now, for all $i \in \llbracket 1, N \rrbracket$ one has $\mathbf{r}_i = \mathbf{r}_i(q_1, \dots, q_g)$ and :

$$\delta \mathbf{r}_i = \frac{\partial \mathbf{r}_i}{\partial q_1} \delta q_1 + \dots + \frac{\partial \mathbf{r}_i}{\partial q_g} \delta q_g,$$

so that:

$$\delta W = \sum_{i=1}^N \mathbf{F}_i \cdot \left(\sum_{k=1}^g \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \right) = \sum_{k=1}^g \left(\sum_{i=1}^N \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \delta q_k.$$

We consequently define \mathbb{F}_k the force applied on the generalised coordinate q_k , $k \in \llbracket 1, g \rrbracket$ as:

$$\mathbb{F}_k = \sum_{i=1}^N \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k}.$$

In case the system is conservative, one has $\mathbf{F}_i = -\nabla_{\mathbf{r}_i} V$, for all $i \in \llbracket 1, N \rrbracket$, where V is the system's total potential energy. This yields, for $k \in \llbracket 1, g \rrbracket$:

$$\begin{aligned} \mathbb{F}_k &= \sum_{i=1}^N \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = - \sum_{i=1}^N \nabla_{\mathbf{r}_i} V \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = - \sum_{i=1}^N \nabla_{\mathbf{r}_i} V_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \\ &= - \sum_{i=1}^N \left(\frac{\partial V_i}{\partial x_i} \frac{\partial x_i}{\partial q_k} \mathbf{e}_1 + \frac{\partial V_i}{\partial y_i} \frac{\partial y_i}{\partial q_k} \mathbf{e}_2 + \frac{\partial V_i}{\partial z_i} \frac{\partial z_i}{\partial q_k} \mathbf{e}_3 \right) \\ &= - \sum_{i=1}^N \frac{\partial V_i}{\partial q_k} = - \frac{\partial}{\partial q_k} \left(\sum_{i=1}^N V_i \right) \end{aligned}$$

hence

$$\mathbb{F}_k = - \frac{\partial V}{\partial q_k}.$$

B.2.6 Hamiltonian

Let us rewrite the Lagrangian \mathcal{L} as $\mathcal{L}(\mathbf{q}, \mathbf{p}; t)$ where $\mathbf{q} = (q_1, \dots, q_g)$ and $\dot{\mathbf{q}} = (\dot{q}_1, \dots, \dot{q}_g)$:

$$\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}; t) = E_c(\dot{\mathbf{q}}) - V(\mathbf{q})$$

From the generalised Lagrangian equation (B.3), one can express the conjugate momentum of the k -th generalised coordinate as follows:

$$\dot{p}_k = \frac{\partial \mathcal{L}}{\partial \dot{q}_k}. \quad (\text{B.5})$$

Equation (B.5) is a direct rewriting of the d'Alembert principle when p_k has the dimension of a momentum, whereas it is a rewriting of the angular momentum theorem when p_k has the dimension of an angular momentum.

Now, provided the Lagrangian \mathcal{L} is time-independent, using (B.5), one can write its total differential as follows:

$$\begin{aligned} d\mathcal{L} &= \frac{\partial \mathcal{L}}{\partial q_k} dq_k + \frac{\partial \mathcal{L}}{\partial \dot{q}_k} d\dot{q}_k \\ &= \sum_{k=1}^g \dot{p}_k dq_k + \frac{\partial \mathcal{L}}{\partial \dot{q}_k} d\dot{q}_k. \end{aligned}$$

Recall that equation (B.4) can be rewritten as $p_k = \frac{\partial \mathcal{L}}{\partial \dot{q}_k}$, which yields

$$d\mathcal{L} = \sum_{k=1}^g \dot{p}_k dq_k + p_k d\dot{q}_k.$$

Now, since $d(p_k \dot{q}_k) = dp_k \dot{q}_k + p_k d\dot{q}_k$:

$$d\mathcal{L} = \sum_{k=1}^g \dot{p}_k dq_k + \sum_{k=1}^g d(p_k \dot{q}_k) - \sum_{k=1}^g dp_k \dot{q}_k$$

i.e

$$d \left(\mathcal{L} - \sum_{k=1}^g p_k \dot{q}_k \right) = \sum_{k=1}^g \dot{p}_k dq_k - \sum_{k=1}^g dp_k \dot{q}_k.$$

This implies that $\frac{d}{dt} \left(\mathcal{L} - \sum_{k=1}^g p_k \dot{q}_k \right) = 0$, and the quantity $\left(\mathcal{L} - \sum_{k=1}^g p_k \dot{q}_k \right)$ is consequently an invariant of the system. This leads us to define the **Hamiltonian of the system**:

$$H(\mathbf{q}, \dot{\mathbf{q}}; t) = \sum_{k=1}^g p_k \dot{q}_k - \mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}; t). \quad (\text{B.6})$$

Let us rewrite the equations of motion by considering the following differential:

$$dH = \sum_{k=1}^g \left(\frac{\partial H}{\partial q_k} dq_k + \frac{\partial H}{\partial p_k} dp_k \right) + \frac{\partial H}{\partial t} dt.$$

H being defined by (B.6), one also has:

$$\begin{aligned} dH &= \sum_{k=1}^g (dp_k \dot{q}_k + p_k d\dot{q}_k) - d\mathcal{L}(\mathbf{q}, \dot{\mathbf{q}}; t) \\ &= \sum_{k=1}^g (dp_k \dot{q}_k + p_k d\dot{q}_k) - \sum_{k=1}^g \left(\frac{\partial \mathcal{L}}{\partial q_k} dq_k + \frac{\partial \mathcal{L}}{\partial \dot{q}_k} d\dot{q}_k \right) - \frac{\partial \mathcal{L}}{\partial t} dt \\ &= \sum_{k=1}^g (dp_k \dot{q}_k + p_k d\dot{q}_k - \dot{p}_k dq_k - p_k dq_k) - \frac{\partial \mathcal{L}}{\partial t} dt, \end{aligned}$$

where we used equations (B.4) and (B.5). Hence:

$$dH = \sum_{k=1}^g (-\dot{p}_k dq_k + p_k d\dot{q}_k) - \frac{\partial \mathcal{L}}{\partial t} dt,$$

which yields the so-called **Hamiltonian equations of motion**:

$$\frac{\partial H}{\partial q_k} = -\dot{p}_k \quad \text{and} \quad \frac{\partial H}{\partial p_k} = \dot{q}_k.$$

Appendix C

Derivatives for the implementation of λ -dynamics and the OSRW method

C.1 Halgren potential

In the following section, the Lennard-Jones potential will be used as a reference to check our computations. We will rely on

$$\frac{dV_{LJ}(\rho)}{d\rho} = \varepsilon \left(-\frac{12}{\rho^{13}} + \frac{12}{\rho^7} \right) = 12\varepsilon \left(\frac{1}{\rho^7} - \frac{1}{\rho^{13}} \right),$$

and

$$\frac{d^2V_{LJ}(\rho)}{d\rho^2} = 12\varepsilon \left(\frac{13}{\rho^{14}} - \frac{7}{\rho^8} \right).$$

Note that we have $\frac{d^2V_{LJ}(r)}{dr^2} = \frac{1}{(r^*)^2} \frac{d^2V_{vdW}(\rho)}{d\rho^2}$.

Derivatives of the Halgren potential – We will need to implement the space derivative of the Halgren potential (6.10). One has

$$V_{hal}(\rho) = \varepsilon(1 + \delta)^{n-m} (\rho + \delta)^{m-n} (1 - \gamma - 2\rho^m) (\rho^m + \gamma)^{-1}$$

i.e

$$V_{hal}(\rho) = \varepsilon(1 + \delta)^{n-m} f(\rho)g(\rho)$$

with

- $f(\rho) = (\rho + \delta)^{m-n}$ and $\frac{df(\rho)}{d\rho} = (m - n) (\rho + \delta)^{m-n-1}$
- $g(\rho) = \frac{(1 - \gamma - 2\rho^m)}{(\rho^m + \gamma)}$ and $\frac{dg}{d\rho} = -m\rho^{m-1} \frac{(\gamma + 1)}{(\rho^m + \gamma)^2}$.

This yields

$$\begin{aligned} \frac{dV_{hal}(\rho)}{d\rho} &= \varepsilon(1+\delta)^{n-m} \left[\frac{df(\rho)}{d\rho} g(\rho) + f(\rho) \frac{dg(\rho)}{d\rho} \right] \\ &= \varepsilon(1+\delta)^{n-m} \left[(m-n)(\rho+\delta)^{m-n-1} \frac{(1-\gamma-2\rho^m)}{(\rho^m+\gamma)} - m\rho^{m-1} \frac{(\gamma+1)}{(\rho^m+\gamma)^2} (\rho+\delta)^{m-n} \right] \end{aligned}$$

i.e

$$\frac{dV_{hal}(\rho)}{d\rho} = \varepsilon(1+\delta)^{n-m} \frac{(\rho+\delta)^{m-n}}{(\rho^m+\gamma)} \left[(m-n) \frac{(1-\gamma-2\rho^m)}{(\rho+\delta)} - m(\gamma+1) \frac{\rho^{m-1}}{(\rho^m+\gamma)} \right].$$

We will also need its second space derivative. Let us consider

$$\begin{cases} h_0(\rho) := \frac{(\rho+\delta)^{m-n}}{(\rho^m+\gamma)} \\ h_1(\rho) := (m-n) \frac{(1-\gamma-2\rho^m)}{(\rho+\delta)} \\ h_2(\rho) := m\rho^{m-1} \frac{(\gamma+1)}{(\rho^m+\gamma)} \end{cases}$$

so that:

$$\frac{dV_{hal}(\rho)}{d\rho} = \varepsilon(1+\delta)^{n-m} h_0(\rho) [h_1(\rho) - h_2(\rho)],$$

and

$$\frac{d^2V_{hal}(\rho)}{d\rho^2} = \varepsilon(1+\delta)^{n-m} \left(\frac{dh_0(\rho)}{d\rho} [h_1(\rho) - h_2(\rho)] + h_0(\rho) \left[\frac{dh_1(\rho)}{d\rho} - \frac{dh_2(\rho)}{d\rho} \right] \right), \quad (C.1)$$

the derivatives of h_0 , h_1 and h_2 being given by:

$$\begin{aligned} \frac{dh_0(\rho)}{d\rho} &= \frac{(\rho+\delta)^{m-n}}{(\rho^m+\gamma)^2} \left[(m-n) \frac{(\rho^m+\delta)}{(\rho+\delta)} - m\rho^{m-1} \right] \\ \frac{dh_1(\rho)}{d\rho} &= (m-n) \frac{(2(1-m)\rho^m - 2m\delta\rho^{m-1} + (\gamma-1))}{(\rho+\delta)^2} \end{aligned}$$

and

$$\frac{dh_2(\rho)}{d\rho} = m(\gamma+1) \frac{(\gamma(m-1)\rho^{m-2} - \rho^{2m-2})}{(\rho^m+\gamma)^2}.$$

Verification – These analytical expressions have been verified with the computer algebra program Mathematica. We can additionally check if our derivatives are sound by comparing it

with the Lennard-Jones potential. One has:

$$\begin{aligned}\frac{dV_{LJ}(\rho)}{d\rho} &= \varepsilon \frac{\rho^{-6}}{\rho^6} \left(-6 \frac{(1-2\rho^6)}{\rho} - 6 \frac{\rho^5}{\rho^6} \right) \\ &= \varepsilon \frac{1}{\rho^{12}} \left(12\rho^5 - \frac{12}{\rho} \right) \\ &= 12\varepsilon \left(\frac{1}{\rho^7} - \frac{1}{\rho^{13}} \right).\end{aligned}$$

Furthermore,

$$\begin{cases} h_0(\rho) = \frac{1}{\rho^{12}}, & \frac{dh_0(\rho)}{d\rho} = -\frac{12}{\rho^{13}} \\ h_1(\rho) = -\frac{6}{\rho} + 12\rho^5, & \frac{dh_1(\rho)}{d\rho} = \frac{6}{\rho^2} + 12 \times 5\rho^4 \\ h_2(\rho) = \frac{6}{\rho}, & \frac{dh_2(\rho)}{d\rho} = -\frac{6}{\rho^2} \end{cases}$$

and one then has:

$$\begin{aligned}\frac{d^2V_{LJ}(\rho)}{d\rho^2} &= \varepsilon \left(\frac{-12}{\rho^{13}} \right) \left[-\frac{6}{\rho} + 12\rho^5 - \frac{6}{\rho} \right] + \varepsilon \frac{1}{\rho^{12}} \left[\frac{6}{\rho^2} + 12 \times 5\rho^4 + \frac{6}{\rho^2} \right] \\ &= \varepsilon \left[\frac{12 \times 12}{\rho^{14}} - \frac{12 \times 12}{\rho^8} + \frac{12}{\rho^{14}} + \frac{12 \times 5}{\rho^8} \right] \\ &= 12\varepsilon \left(\frac{13}{\rho^{14}} - \frac{7}{\rho^8} \right).\end{aligned}$$

C.2 Softcore potential for van der Waals interactions

Let us now derive the expression of the generic softcore potential (6.11) derivatives, that will be needed for the implementation of the λ -dynamics.

Spatial derivatives – One has

$$\frac{\partial}{\partial \rho} g(\rho, \lambda_v) = \frac{1}{k} k \rho^{k-1} r^* \left(\alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^s + \rho^k \right)^{\frac{1}{k}-1}$$

i.e

$$\frac{\partial}{\partial \rho} g(\rho, \lambda_v) = \rho^{k-1} (r^*)^k g(\rho, \lambda_v)^{1-k}.$$

This yields that:

$$\begin{aligned}\frac{\partial V_{SC-hal}(\rho, \lambda_v)}{\partial \rho} &= \lambda_v^t \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) \\ \frac{\partial V_{SC-hal}(r, \lambda_v)}{\partial r} &= \frac{d\rho}{dr} \frac{\partial V_{SC-hal}(\rho, \lambda_v)}{\partial \rho} = \frac{1}{r^*} \lambda_v^t \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)).\end{aligned}$$

Derivatives with respect to λ_v – One has

$$\frac{\partial}{\partial \lambda_v} g(\rho, \lambda_v) = -\frac{s}{k} \alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^{s-1} (r^*)^k g(\rho, \lambda_v)^{1-k}.$$

and

$$\begin{aligned} \frac{\partial^2}{\partial \lambda_v^2} g(\rho, \lambda_v) &= \frac{s(s-1)}{k} \alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^{s-2} (r^*)^k g(\rho, \lambda_v)^{1-k} \\ &\quad + (1-k) \left(\frac{s}{k} \alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^{s-1} (r^*)^k \right)^2 g(\rho, \lambda_v)^{1-2k}. \end{aligned}$$

One consequently has:

$$\begin{aligned} F_v(q, \lambda) &:= \frac{\partial V_{SC-hal}(\rho, \lambda_v)}{\partial \lambda_v} = t \lambda_v^{t-1} V_{hal}(g(\rho, \lambda_v)) + \lambda_v^t \frac{\partial V_{hal}(g(\rho, \lambda_v))}{\partial \lambda_v} \\ &= t \lambda_v^{t-1} V_{hal}(g(\rho, \lambda_v)) + \lambda_v^t \frac{\partial g(\rho, \lambda_v)}{\partial \lambda_v} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial \lambda_v} F_v(q, \lambda) &= \frac{\partial^2 V_{SC-hal}(\rho, \lambda_v)}{\partial \lambda_v^2} \\ &= t(t-1) \lambda_v^{t-2} V_{hal}(g(\rho, \lambda_v)) + 2t \lambda_v^{t-1} \frac{\partial g(\rho, \lambda_v)}{\partial \lambda_v} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) \\ &\quad + \lambda_v^t \left[\frac{\partial^2 g(\rho, \lambda_v)}{\partial \lambda_v^2} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) + \left(\frac{\partial g(\rho, \lambda_v)}{\partial \lambda_v} \right)^2 \frac{d^2 V_{hal}}{d\rho^2}(g(\rho, \lambda_v)) \right]. \end{aligned}$$

Let us recall that $\frac{d^2 V_{hal}}{d\rho^2}$ is given by (C.1).

Cross derivatives – One has:

$$\frac{\partial^2}{\partial \rho \partial \lambda_v} g(\rho, \lambda_v) = \frac{\partial^2}{\partial \lambda_v \partial \rho} g(\rho, \lambda_v) = -\frac{s(1-k)}{k} \alpha 2^{-\frac{k}{6}} (1 - \lambda_v)^{s-1} \rho^{k-1} (r^*)^{2k} g(\rho, \lambda_v)^{1-2k}.$$

So that:

$$\begin{aligned}
\frac{\partial^2 V_{SC-hal}}{\partial \rho \partial \lambda} &= \frac{\partial^2 V_{SC-hal}}{\partial \lambda \partial \rho} = \frac{\partial}{\partial \lambda} \left(\lambda_v^t \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) \right) \\
&= \frac{d\lambda_e}{d\lambda} \frac{\partial}{\partial \lambda_e} \left(\lambda_v^t \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) \right) \\
&= \frac{d\lambda_e}{d\lambda} \left(t \lambda_v^{t-1} \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) \right. \\
&\quad \left. + \lambda_v^t \left(\frac{\partial^2 g(\rho, \lambda_v)}{\partial \rho \partial \lambda_v} \frac{dV_{hal}}{d\rho}(g(\rho, \lambda_v)) + \frac{\partial g(\rho, \lambda_v)}{\partial \rho} \cdot \frac{\partial g(\rho, \lambda_v)}{\partial \lambda_v} \frac{d^2 V_{hal}}{d\rho^2}(g(\rho, \lambda_v)) \right) \right).
\end{aligned}$$

All of these expressions have been implemented in the Tinker-HP code in the `elambdaalj1c` subroutine of the `elj1` routine. The parameter α , s , t and k are declared in the `mutant` module, and their values need to be written by the user in the input simulation `.key` file.

Potential derivatives – Eventually, one gets:

$$\frac{\partial V_{SC-hal}(r, \lambda_v)}{\partial x_i} = \frac{dr}{dx_i} \frac{1}{r^*} \frac{\partial V_{SC-hal}(\rho, \lambda_v)}{\partial \rho}, \quad \forall i \in \llbracket 1, 3 \rrbracket.$$

$$\frac{\partial V_{SC-hal}(r, \lambda_v)}{\partial \lambda} = \frac{d\lambda_v}{d\lambda} \frac{\partial V_{SC-hal}(r, \lambda_v)}{\partial \lambda_v}.$$

and

$$\frac{\partial F_v(r, \lambda_v)}{\partial x_i} = \frac{dr}{dx_i} \frac{1}{r^*} \frac{\partial^2 V_{SC-hal}}{\partial \rho \partial \lambda}$$

C.3 Softcore potential for electrostatic interactions

Softcore derivatives – One has that:

$$\begin{aligned}
\frac{\partial f(r, \lambda_e)}{\partial r} &= r (\alpha_e (1 - \lambda_e)^s r^u)^{\frac{1}{u}-1} = r f(r, \lambda_e)^{1-u} \\
\frac{\partial f(r, \lambda_e)}{\partial \lambda_e} &= -\frac{s}{u} \alpha_e (1 - \lambda_e)^{s-1} f(r, \lambda_e)^{1-u}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f(r, \lambda_e)}{\partial r \partial \lambda_e} &= -\frac{s}{u} \alpha_e (1 - \lambda_e)^{s-1} (1 - u) \frac{\partial f(r, \lambda_e)}{\partial r} f(r, \lambda_e)^{-u} \\
&= \frac{s(u-1)}{u} \alpha_e (1 - \lambda_e)^{s-1} r f(r, \lambda_e)^{1-2u},
\end{aligned}$$

and

$$\frac{\partial^2 f(r, \lambda_e)}{\partial \lambda_e^2} = \frac{\mathfrak{s}(\mathfrak{s} - 1)}{\mathfrak{u}} \alpha_e (1 - \lambda_e)^{\mathfrak{s}-2} f(r, \lambda_e)^{1-\mathfrak{u}} - \frac{\mathfrak{s}}{\mathfrak{u}} \alpha_e (1 - \lambda_e)^{\mathfrak{s}-1} (1 - \mathfrak{u}) \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} f(r, \lambda_e)^{-\mathfrak{u}}$$

i.e

$$\frac{\partial^2 f(r, \lambda_e)}{\partial \lambda_e^2} = \frac{\mathfrak{s}(\mathfrak{s} - 1)}{\mathfrak{u}} \alpha_e (1 - \lambda_e)^{\mathfrak{s}-2} f(r, \lambda_e)^{1-\mathfrak{u}} + (1 - \mathfrak{u}) \left(\frac{\mathfrak{s}}{\mathfrak{u}} \alpha_e (1 - \lambda_e)^{\mathfrak{s}-1} \right)^2 f(r, \lambda_e)^{1-2\mathfrak{u}}.$$

Potential derivatives – In practice, the energy term $V_{corr}(r, \lambda_e)$ is not directly implemented, and is treated by the Tinker-HP code simultaneously with the direct energy term $V_{dir}(r, \lambda_e)$, so that one only needs to properly modify the real part of the electrostatic energy to take the softcore into account. In order to implement (6.17), (6.19) and (6.18), one needs to compute the derivatives of $V_{dir}(r, \lambda_e)$. Given two atoms labelled i and j , we distinguish between the three cases:

(1) **The two atoms i and k are mutated** – In which case

$$V_{dir}^{ij}(f(r, \lambda_e)) = \lambda_e^{2\mathfrak{t}} E_{real}(f(r, \lambda_e))$$

and :

$$\frac{\partial V_{dir}^{ij}(f(r, \lambda_e))}{\partial r} = \lambda_e^{2\mathfrak{t}} \frac{\partial f(r, \lambda_e)}{\partial r} \frac{dE_{real}(f(r, \lambda_e))}{dr}$$

$$\frac{\partial V_{dir}^{ij}(f(r, \lambda_e))}{\partial \lambda_e} = 2\mathfrak{t} \lambda_e^{2\mathfrak{t}-1} E_{real}(f(r, \lambda_e)) + \lambda_e^{2\mathfrak{t}} \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr}$$

$$\begin{aligned} \frac{\partial^2 V_{dir}^{ij}(f(r, \lambda_e))}{\partial r \partial \lambda_e} &= 2\mathfrak{t} \lambda_e^{2\mathfrak{t}-1} \frac{\partial f(r, \lambda_e)}{\partial r} \frac{dE_{real}(f(r, \lambda_e))}{dr} \\ &+ \lambda_e^{2\mathfrak{t}} \left(\frac{\partial^2 f(r, \lambda_e)}{\partial r \partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr} + \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{\partial f(r, \lambda_e)}{\partial r} \frac{d^2 E_{real}(f(r, \lambda_e))}{dr^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 V_{dir}^{ij}(f(r, \lambda_e))}{\partial \lambda_e^2} &= 2\mathfrak{t}(2\mathfrak{t} - 1) \lambda_e^{2\mathfrak{t}-2} E_{real}(f(r, \lambda_e)) + 4\mathfrak{t} \lambda_e^{2\mathfrak{t}-1} \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr} \\ &+ \lambda_e^{2\mathfrak{t}} \left(\frac{\partial^2 f(r, \lambda_e)}{\partial \lambda_e^2} \frac{dE_{real}(f(r, \lambda_e))}{dr} + \left(\frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \right)^2 \frac{d^2 E_{real}(f(r, \lambda_e))}{dr^2} \right). \end{aligned}$$

(2) Only one of the two atoms i and k is mutated – In which case

$$V_{dir}^{ij}(f(r, \lambda_e)) = \lambda_e^t E_{real}(f(r, \lambda_e))$$

and :

$$\frac{\partial V_{dir}^{ij}(f(r, \lambda_e))}{\partial r} = \lambda_e^t \frac{\partial f(r, \lambda_e)}{\partial r} \frac{dE_{real}(f(r, \lambda_e))}{dr}$$

$$\frac{\partial V_{dir}^{ij}(f(r, \lambda_e))}{\partial \lambda_e} = t \lambda_e^{t-1} E_{real}(f(r, \lambda_e)) + \lambda_e^t \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr}$$

$$\begin{aligned} \frac{\partial^2 V_{dir}^{ij}(f(r, \lambda_e))}{\partial r \partial \lambda_e} &= t \lambda_e^{t-1} \frac{\partial f(r, \lambda_e)}{\partial r} \frac{dE_{real}(f(r, \lambda_e))}{dr} \\ &+ \lambda_e^t \left(\frac{\partial^2 f(r, \lambda_e)}{\partial r \partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr} + \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{\partial f(r, \lambda_e)}{\partial r} \frac{d^2 E_{real}(f(r, \lambda_e))}{dr^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 V_{dir}^{ij}(f(r, \lambda_e))}{\partial \lambda_e^2} &= t(t-1) \lambda_e^{t-2} E_{real}(f(r, \lambda_e)) + 2t \lambda_e^{t-1} \frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \frac{dE_{real}(f(r, \lambda_e))}{dr} \\ &+ \lambda_e^t \left(\frac{\partial^2 f(r, \lambda_e)}{\partial \lambda_e^2} \frac{dE_{real}(f(r, \lambda_e))}{dr} + \left(\frac{\partial f(r, \lambda_e)}{\partial \lambda_e} \right)^2 \frac{d^2 E_{real}(f(r, \lambda_e))}{dr^2} \right). \end{aligned}$$

(3) None of the atoms i and k are mutated – In which case

$$V_{dir}^{ij}(r, \lambda_e) = V_{dir}(r) = E_{real}(r)$$

and :

$$\frac{\partial V_{dir}^{ij}(r, \lambda_e)}{\partial r} = \frac{dE_{real}(f(r, \lambda_e))}{dr}$$

$$\frac{\partial V_{dir}^{ij}(r, \lambda_e)}{\partial \lambda_e} = \frac{\partial^2 V_{dir}(r, \lambda_e)}{\partial r \partial \lambda_e} = \frac{\partial^2 V_{dir}(r, \lambda_e)}{\partial \lambda_e^2} = 0.$$

References

- [1] J. R. Abella, S. Y. Cheng, Q. Wang, W. Yang, and P. Ren. “Hydration Free Energy from Orthogonal Space Random Walk and Polarizable Force Field”. In: *Journal of Chemical Theory and Computation* 10.7 (2014), pp. 2792–2801. DOI: 10.1021/ct500202q.
- [2] H. Alrachid and T. Lelièvre. “Long-time convergence of an adaptive biasing force method: variance reduction by Helmholtz projection”. en. In: *Journal of computational mathematics* 1 (2015), pp. 55–82. DOI: 10.5802/smai-jcm.4.
- [3] L. Ambrosio, A. Carlotto, and A. Massaccesi. *Lectures on Elliptic Partial Differential Equations*. Jan. 2018. DOI: 10.1007/978-88-7642-651-3.
- [4] P. Atkins. *Physical Chemistry: Sixth Edition*. W.H. Freeman and Company, 2000.
- [5] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and Geometry of Markov Diffusion operators*. Grundlehren der mathematischen Wissenschaften, Vol. 348. Springer, Jan. 2014.
- [6] A. Barducci, G. Bussi, and M. Parrinello. “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: *Phys. Rev. Lett.* 100 (2 Jan. 2008), p. 020603. DOI: 10.1103/PhysRevLett.100.020603.
- [7] M. Benaïm and C.-É. Bréhier. “Convergence of adaptive biasing potential methods for diffusions”. In: *Comptes Rendus Mathématique* 354.8 (2016), pp. 842–846. DOI: 10.1016/j.crma.2016.05.011.
- [8] M. Benaïm and C.-E. Bréhier. “Convergence analysis of Adaptive Biasing Potential methods for diffusion processes”. In: *Communications in Mathematical Sciences* (2019).
- [9] M. Benaïm, C.-E. Bréhier, and P. Monmarché. “Analysis of an Adaptive Biasing Force method based on self-interacting dynamics”. In: *Electronic Journal of Probability* 25.none (2020), pp. 1–28. DOI: 10.1214/20-EJP490.
- [10] C. H. Bennett. “Efficient estimation of free energy differences from Monte Carlo data”. In: *Journal of Computational Physics* 22.2 (1976), pp. 245–268. DOI: 10.1016/0021-9991(76)90078-4.
- [11] A. Bernardin et al. *Collective variables module: Reference manual for NAMD*. 2020.
- [12] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. “Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations”. In: *Chemical Physics Letters* 222.6 (1994), pp. 529–539. DOI: 10.1016/0009-2614(94)00397-1.
- [13] R. Bitetti-Putzer, W. Yang, and M. Karplus. “Generalized ensembles serve to improve the convergence of free energy simulations”. In: *Chemical Physics Letters* 377.5 (2003), pp. 633–641. DOI: 10.1016/S0009-2614(03)01057-1.

- [14] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov. *Fokker–Planck–Kolmogorov equations*. American Mathematical Society, 2015.
- [15] J.-P. Bouchaud, G. Montambaux, A. Georges, and M. Mézard. *Physique statistique 1*. Département de Physique de l’Ecole Polytechnique. 2017.
- [16] F. Bouchet and J. Reygner. “Generalisation of the Eyring–Kramers Transition Rate Formula to Irreversible Diffusion Processes”. In: *Annales Henri Poincaré* 17.12 (2016), pp. 3499–3532. DOI: 10.1007/s00023-016-0507-4.
- [17] H. Brézis. *Analyse fonctionnelle : théorie et applications*. Masson, 1987.
- [18] G. Bussi and D. Branduardi. “Free-Energy Calculations with Metadynamics: Theory and Practice”. In: *Reviews in Computational Chemistry Volume 28*. John Wiley & Sons, Ltd, 2015. Chap. 1, pp. 1–49. DOI: 10.1002/9781118889886.ch1.
- [19] C. Camilloni et al. “Towards a structural biology of the hydrophobic effect in protein folding”. In: *Scientific Reports* 6.1 (July 2016), p. 28285. DOI: 10.1038/srep28285.
- [20] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. Schütt, and K.-R. Müller. “Machine Learning of Accurate Energy-Conserving Molecular Force Fields”. In: *Science Advances* 3.5 (May 2017), e1603015. DOI: 10.1126/sciadv.1603015.
- [21] H. Claudine. *Physique statistique : et illustrations en physique du solide*. Éd. de l’École polytechnique, 2003.
- [22] D. Comez. “Modern Ergodic Theory: From a Physics Hypothesis to a Mathematical Theory with Transformative Interdisciplinary Impact”. In: 2019.
- [23] D. S. Corti and G. Soto-Campos. “Deriving the isothermal–isobaric ensemble: The requirement of a “shell” molecule and applicability to small systems”. In: *The Journal of Chemical Physics* 108.19 (1998), pp. 7959–7966. DOI: 10.1063/1.476236.
- [24] E. Darve and A. Pohorille. “Calculating free energies using average force”. In: *The Journal of Chemical Physics* 115.20 (2001), pp. 9169–9183. DOI: 10.1063/1.1410978.
- [25] E. B. Dynkin. *Markov Processes Volume I*. Springer Verlag, 1965.
- [26] E. B. Dynkin. *Markov Processes Volume II*. Springer Verlag, 1965.
- [27] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. “A smooth particle mesh Ewald method”. In: *The Journal of Chemical Physics* 103.19 (1995), pp. 8577–8593. DOI: 10.1063/1.470117.
- [28] L. C. Evans. *Partial Differential Equations*. American Mathematical Society, 2010.
- [29] A. L. Ferguson. “Machine learning and data science in soft materials engineering”. In: *Journal of Physics: Condensed Matter* 30.4 (Dec. 2017), p. 043002. DOI: 10.1088/1361-648x/aa98bd.
- [30] G. Fiorin, M. L. Klein, and J. Hémin. “Using collective variables to drive molecular dynamics simulations”. In: *Mol. Phys.* 111.22-23 (2013), pp. 3345–3362. DOI: 10.1080/00268976.2013.813594.
- [31] H. Fu, X. Shao, C. Chipot, and W. Cai. “Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations”. In: *J. Chem. Theory Comput.* 12(8) (2016), pp. 3506–3513. DOI: 10.1021/acs.jctc.6b00447.
- [32] P. Gkeka et al. “Machine learning force fields and coarse-grained variables in molecular dynamics: application to materials and biological systems”. In: *arXiv e-prints*, arXiv:2004.06950 (Apr. 2020), arXiv:2004.06950. arXiv: 2004.06950 [physics.comp-ph].

- [33] *GROMACS User Manual version 5.0.4*, www.gromacs.org. (2014).
- [34] Z. Guo, C. L. Brooks, and X. Kong. “Efficient and Flexible Algorithm for Free Energy Calculations Using the λ -Dynamics Approach”. In: *The Journal of Physical Chemistry B* 102.11 (1998), pp. 2032–2036. DOI: 10.1021/jp972699+.
- [35] I. Gutman and O. E. Polansky. *Mathematical Concepts in Organic Chemistry*. Springer-Verlag Berlin Heidelberg, 1986.
- [36] T. A. Halgren. “The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters”. In: *Journal of the American Chemical Society* 114.20 (1992), pp. 7827–7843. DOI: 10.1021/ja00046a032.
- [37] P. Hänggi and F. Marchesoni. “Artificial Brownian motors: Controlling transport on the nanoscale”. In: *Rev. Mod. Phys.* 81 (1 Mar. 2009), pp. 387–442. DOI: 10.1103/RevModPhys.81.387.
- [38] R. L. Hayes, K. A. Armacost, J. Z. Vilseck, and C. L. Brooks. “Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ -Dynamics”. In: *The Journal of Physical Chemistry B* 121.15 (2017), pp. 3626–3635. DOI: 10.1021/acs.jpcc.6b09656.
- [39] J. Hénin and C. Chipot. “Overcoming free energy barriers using unconstrained molecular dynamics simulations”. In: *The Journal of Chemical Physics* 121.7 (2004), pp. 2904–2914. DOI: 10.1063/1.1773132.
- [40] T. Hocquet. “Physique statistique”. Master. Lecture. France, Apr. 2018.
- [41] N. Holzwarth. *LNotes on the Ewald summation*. <http://users.wfu.edu/natalie/f15phy752/lecturenote/>. Accessed: 2021-08-31. 2015.
- [42] L. Hörmander. “Hypoelliptic second order differential equations”. In: *Acta Mathematica* 119.none (1967), pp. 147–171. DOI: 10.1007/BF02392081.
- [43] K. Huang. *Introduction to Statistical Physics*. Graduate Texts in Contemporary Physics. CRC Press LLC, 2001.
- [44] W. Jia et al. *Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning*. 2020. arXiv: 2005.00223.
- [45] B. Jourdain, T. Lelièvre, and R. Roux. “Existence, uniqueness and convergence of a particle approximation for the Adaptive Biasing Force process”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 44.05 (Sept. 2010), pp. 831–865. DOI: 10.1051/m2an/2010044.
- [46] J. Kästner. “Umbrella sampling”. In: *WIREs Computational Molecular Science* 1.6 (2011), pp. 932–942. DOI: 10.1002/wcms.66.
- [47] J. L. Knight and C. L. Brooks. “Lambda-dynamics free energy simulation methods.” In: *Journal of computational chemistry* 30 (2009), pp. 1692–1700. DOI: 10.1002/jcc.21295.
- [48] J. M. Kolos, A. M. Voll, M. Bauder, and F. Hausch. “FKBP Ligands - Where We Are and Where to Go?” In: *Frontiers in Pharmacology* 9 (2018), p. 1425. DOI: 10.3389/fphar.2018.01425.
- [49] X. Kong and C. L. Brooks. “Lambda-dynamics: A new approach to free energy calculations”. In: *The Journal of Chemical Physics* 105.6 (1996), pp. 2414–2423. DOI: 10.1063/1.472109.
- [50] N. Krylov. “On diffusion processes with drift in L_d ”. In: *Probab. Theory Relat. Fields* (2020). DOI: 10.1007/s00440-020-01007-3.

- [51] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. “THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method”. In: *Journal of Computational Chemistry* 13.8 (1992), pp. 1011–1021. DOI: 10.1002/jcc.540130812.
- [52] L. Lagardère et al. “Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields”. In: *Chemical science* 9.4 (Nov. 2017). DOI: 10.1039/c7sc04531j.
- [53] A. Laio and M. Parrinello. “Escaping free-energy minima”. In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12562–12566. DOI: 10.1073/pnas.202427399. eprint: <https://www.pnas.org/content/99/20/12562.full.pdf>.
- [54] B. Leimkuhler and C. Matthews. *Molecular Dynamics. With Deterministic and Stochastic Numerical Methods*. Vol. 39. Interdisciplinary Applied Mathematics. Springer International Publishing Switzerland, 2015. DOI: 10.1007/978-3-319-16375-8.
- [55] T. Lelièvre. “Two mathematical tools to analyze metastable stochastic processes”. 2013.
- [56] T. Lelièvre, M. Rousset, and G. Stoltz. *Free Energy Computations*. Imperial College Press, 2010.
- [57] T. Lelièvre, M. Rousset, and G. Stoltz. “Long-time convergence of an Adaptive Biasing Force method”. In: *Nonlinearity* 21 (2008). DOI: 10.1088/0951-7715/21/6/001.
- [58] T. Lelièvre and G. Stoltz. “Partial differential equations and stochastic methods in molecular dynamics”. In: *Acta Numerica* 25 (2016).
- [59] Q. Liao. “Chapter Four - Enhanced sampling and free energy calculations for protein simulations”. In: *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*. Ed. by B. Strodel and B. Barz. Vol. 170. Progress in Molecular Biology and Translational Science. Academic Press, 2020, pp. 177–213. DOI: 10.1016/bs.pmbts.2020.01.006.
- [60] C. Lu, X. Li, D. Wu, L. Zheng, and W. Yang. “Predictive Sampling of Rare Conformational Events in Aqueous Solution: Designing a Generalized Orthogonal Space Tempering Method”. In: *Journal of Chemical Theory and Computation* 12.1 (2016), pp. 41–52. DOI: 10.1021/acs.jctc.5b00953.
- [61] E. J. Maginn and J. R. Elliott. “Historical Perspective and Current Outlook for Molecular Dynamics As a Chemical Engineering Tool”. In: *Industrial & Engineering Chemistry Research* 49.7 (2010), pp. 3059–3078. DOI: 10.1021/ie901898k.
- [62] S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino. “Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations”. In: *The Journal of Physical Chemistry B* 110.29 (2006), pp. 14011–14013. DOI: 10.1021/jp062755j.
- [63] G. Menz and A. Schlichting. “Poincaré and logarithmic Sobolev inequalities by decomposition of the energy landscape”. In: *The Annals of Probability* 42.5 (Sept. 2014), pp. 1809–1884. DOI: 10.1214/14-aop908.
- [64] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: 10.1063/1.1699114.
- [65] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer Verlag, London, 1993.

- [66] D. Min, L. Zheng, W. Harris, M. Chen, C. Lv, and W. Yang. “Practically Efficient QM/MM Alchemical Free Energy Simulations: The Orthogonal Space Random Walk Strategy”. In: *Journal of Chemical Theory and Computation* 6.8 (2010). PMID: 26613484, pp. 2253–2266. DOI: 10.1021/ct100033s.
- [67] P. Monmarché. “Generalized Γ calculus and application to interacting particles on a graph”. In: *Potential Analysis* 50.3 (2019), pp. 439–466. DOI: 10.1007/s11118-018-9689-3.
- [68] C. C. Moore. “Ergodic theorem, ergodic theory, and statistical mechanics”. In: *Proceedings of the National Academy of Sciences* 112.7 (2015), pp. 1907–1911. DOI: 10.1073/pnas.1421798112.
- [69] A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe. “Time-reversible Born-Oppenheimer molecular dynamics”. In: (Apr. 2006). DOI: 10.1103/PhysRevLett.97.123001. arXiv: cond-mat/0604643 [cond-mat.mtrl-sci].
- [70] F. Otto and C. Villani. “Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality”. In: *Journal of Functional Analysis* 173.2 (2000), pp. 361–400. DOI: 10.1006/jfan.1999.3557.
- [71] G. A. Pavliotis. *Stochastic Processes and Applications Book Subtitle Diffusion Processes, the Fokker-Planck and Langevin Equations*. Texts in Applied Mathematics. Springer, New York, 2014. DOI: 10.1007/978-1-4939-1323-7.
- [72] D. A. Pearlman and P. A. Kollman. “A new method for carrying out free energy perturbation calculations: Dynamically modified windows”. In: *The Journal of Chemical Physics* 90.4 (1989), pp. 2460–2470. DOI: 10.1063/1.455988.
- [73] J. C. Phillips et al. “Scalable molecular dynamics with NAMD”. In: *Journal of computational chemistry* 26 (Dec. 2005), pp. 1781–1802. DOI: 10.1002/jcc.20289.
- [74] P. Pulay and G. Fogarasi. “Fock matrix dynamics”. In: *Chemical Physics Letters* 386.4 (2004), pp. 272–278. DOI: 10.1016/j.cpllett.2004.01.069.
- [75] R. Qi, Q. Wang, and P. Ren. “General van der Waals potential for common organic molecules”. In: *Bioorganic and Medicinal Chemistry* 24.20 (2016). Advances in Computational and Medicinal Chemistry, pp. 4911–4919. DOI: 10.1016/j.bmc.2016.07.062.
- [76] J. A. Rackers et al. “Tinker 8: Software Tools for Molecular Design”. In: *Journal of Chemical Theory and Computation* 14.10 (2018), pp. 5273–5289. DOI: 10.1021/acs.jctc.8b00529.
- [77] S. E. Radford. “Protein folding: progress made and promises ahead”. In: *Trends in Biochemical Sciences* 25.12 (2000), pp. 611–618. DOI: 10.1016/S0968-0004(00)01707-2.
- [78] J. R. Ray, H. Graben, and J. Haile. “Statistical mechanics of the isoenthalpic-isobaric ensemble”. In: *Il Nuovo Cimento B (1971-1996)* 64 (1981), pp. 191–206.
- [79] J. R. Ray and H. W. Graben. “Fundamental treatment of the isoenthalpic-isobaric ensemble”. In: *Phys. Rev. A* 34 (3 Sept. 1986), pp. 2517–2519. DOI: 10.1103/PhysRevA.34.2517.
- [80] M. J. Schnieders et al. “The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field”. In: *Journal of Chemical Theory and Computation* 8.5 (2012), pp. 1721–1736. DOI: 10.1021/ct300035u.
- [81] H. Sidky, W. Chen, and A. L. Ferguson. “Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation”. In: *Molecular Physics* 118.5 (2020), e1737742. DOI: 10.1080/00268976.2020.1737742.
- [82] Smythe Mark L., Huston Shawn E., and Marshall Garland R. “Free energy profile of a 310- to α -helical transition of an oligopeptide in various solvents”. In: *Journal of the American Chemical Society* 115.24 (1993), pp. 11594–11595. DOI: 10.1021/ja00077a067.

- [83] T. Steinbrecher, I. Joung, and D. A. Case. “Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations”. In: *Journal of computational chemistry* 32.15 (2011), pp. 3253–3263. DOI: 10.1002/jcc.21909.
- [84] “The calculation of the potential of mean force using computer simulations”. In: *Computer Physics Communications* 91.1 (1995), pp. 275–282. DOI: 10.1016/0010-4655(95)00053-I.
- [85] G. Tiberio. “Molecular dynamics simulations of liquid crystals and photoresponsive systems”. PhD thesis. Università degli Studi di Bologna, 2021.
- [86] G. Torrie and J. Valleau. “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: *Journal of Computational Physics* 23.2 (1977), pp. 187–199. DOI: 10.1016/0021-9991(77)90121-8.
- [87] L. Zheng, M. Chen, and W. Yang. “Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems”. In: *Proceedings of the National Academy of Sciences* 105.51 (2008), pp. 20227–20232. DOI: 10.1073/pnas.0810631106. eprint: <https://www.pnas.org/content/105/51/20227.full.pdf>.
- [88] L. Zheng, M. Chen, and W. Yang. “Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling”. In: *The Journal of chemical physics* 130.23 (2009), 06B618. DOI: 10.1063/1.3153841.
- [89] L. Zheng and W. Yang. “Practically efficient and robust free energy calculations: Double-integration orthogonal space tempering”. In: *Journal of chemical theory and computation* 8.3 (2012), pp. 810–823. DOI: 10.1021/ct200726v.
- [90] R. W. Zwanzig. “High-Temperature Equation of State by a Perturbation Method. I. Non-polar Gases”. In: *The Journal of Chemical Physics* (1954).

Abstract

This thesis is dedicated to the study of adaptive biasing algorithms for molecular dynamics simulations, from both theoretical and numerical perspectives. The goal of molecular dynamics is to obtain macroscopic informations about a system of particles, given its microscopic description. Adaptive biasing algorithms are powerful tools in molecular dynamics, especially when one needs to compute a system's free energy. We will mainly focus on the *Adaptive Biasing Force* algorithm, whose key idea is to bias the interaction force between the particles in order to enhance the sampling of the system's configuration space. In particular, we will study its robustness in the case of non-conservative interaction forces. We will then proceed to design an enhanced sampling algorithm in the scope of alchemical transitions, where the system's evolution from an initial state to a final state is indexed by a parameter λ in $[0, 1]$. Such transitions are often used in pharmacology, as they allow the estimation of several free energies, such as the binding free energy of a ligand with a receptor, or even the solvation free energy of a compound in solvent. When coupled to the λ -dynamics method, which deals with the dynamic evolution of the parameter λ , the *Orthogonal Space Random Walk* (OSRW) sampling method may permit a better and quicker sampling of the configuration space. Drawing inspiration from this algorithm, we will implement an adaptive biasing method coupled to the λ -dynamics, and compare it with the original OSRW algorithm. This work led to the implementation of a new interface between the Tinker-HP program and the Collective Variables module software.

Keywords: molecular dynamics, diffusion process, adaptive biasing algorithm, functional inequality, free energy, alchemical transition

Résumé

L'objet de cette thèse porte sur l'étude tant théorique que numérique de méthodes de biais adaptatifs pour la dynamique moléculaire. Etant donnée une description microscopique d'un système de particules, le but de la dynamique moléculaire est d'en déduire des informations macroscopiques. Les méthodes de biais adaptatifs se révèlent être un outil efficace, notamment lorsqu'il est question de déterminer l'énergie libre d'un système. Nous accorderons une attention toute particulière à la méthode de l'*Adaptive Biasing Force*, dont le principe est de biaiser la force d'interaction entre les particules afin d'accélérer l'échantillonnage de l'espace des configurations. En particulier, nous étudierons sa robustesse dans le cas où les forces d'interaction s'avèrent être non-conservatives. Nous procéderons ensuite à la construction d'un algorithme d'accélération d'échantillonnage dans le cadre de transitions alchimiques, où l'évolution du système d'un état initial à un état final prédéfinis est indexée par un paramètre λ compris entre 0 et 1. De telles transitions s'avèrent utiles en pharmacologie, leur étude permettant de déterminer différentes énergies libres, telles que l'énergie libre de liaison d'un ligand avec un récepteur donné ou bien encore l'énergie libre de solvatation d'une molécule ou d'un ion dans un solvant. Couplée à la méthode de la λ -dynamique, qui traite l'évolution dynamique du paramètre λ , la méthode d'échantillonnage de l'*Orthogonal Space Random Walk* (OSRW) pourrait permettre une accélération de l'échantillonnage des configurations. En s'inspirant de cette dernière, nous implémenterons une méthode de biais adaptatif couplée à une méthode de λ -dynamique, que nous comparerons à la méthode de l'OSRW originelle. Ce travail d'implémentation permettra la mise en place d'une nouvelle interface entre le logiciel Tinker-HP et le module Collective Variables.

Mots clés : dynamique moléculaire, processus de diffusion, méthode de biais adaptatif, inégalité fonctionnelle, énergie libre, transitions alchimique



Laboratoire Jacques-Louis Lions

Sorbonne Université – Campus Pierre et Marie Curie – 4 place Jussieu – 75005 Paris – France