Deterministic and stochastic methods for molecular simulation

Kimiya Minoukadeh

UNIVERSITÉ PARIS-EST

THÈSE

pour obtenir le titre de

DOCTEUR EN SCIENCES DE L'UNIVERSITÉ PARIS-EST

Spécialité : MATHÉMATIQUES

École Doctorale : MSTIC

Présentée et soutenue par

Kimiya MINOUKADEH

Méthodes déterministes et stochastiques pour la simulation moléculaire

---

Soutenance le 24 novembre 2010 devant le jury composé de

Rapporteurs :	Jean DOLBEAULT	-	Université Paris-Dauphine
	Erwan FAOU	-	INRIA (Rennes)
Directeur :	Eric CANCÈS	-	École des Ponts ParisTech
Co-directeur :	Tony LELIÈVRE	-	École des Ponts ParisTech
Examinateurs :	Chris CHIPOT	-	Université Nancy
	Denis TALAY	-	INRIA (Sophia-Antipolis)
	Graeme HENKELMAN	-	University of Texas at Austin

Méthodes déterministes et stochastiques pour la simulation moléculaire

Résumé

La simulation moléculaire est un outil indispensable pour comprendre le comportement de systèmes complexes pour lesquels les expériences s'avèrent coûteuses ou irréalisables à l'heure actuelle. Cette thèse est dédiée aux aspects méthodologiques de la simulation moléculaire et comprend deux volets. Le premier volet porte sur la recherche de chemins de réaction et de points col d'une surface d'énergie potentielle. Nous proposons, dans le chapitre 3, une amélioration d'une des méthodes de cette classe, appelée *Activation Relaxation Technique* (ART). Nous donnons également une preuve de convergence pour un algorithme prototype. Le deuxième volet porte sur le calcul d'énergie libre pour les transitions caractérisées par une coordonnée de réaction. Nous nous plaçons dans le cadre d'une méthode d'échantillonnage d'importance adaptative, appelée *Adaptive Biasing Force* (ABF). Ce volet comprend deux sous-parties. La première partie (chapitre 5) s'attache à montrer l'applicabilité à un système biomoléculaire, d'une nouvelle mise en oeuvre parallèle d'ABF, nommée *multiple-walker ABF* (MW-ABF), consistant à utiliser plusieurs répliques. Cette mise en oeuvre s'est avérée utile pour surmonter des problèmes liés à un mauvais choix de coordonnée de réaction. Nous confirmons ensuite ces résultats numériques en étudiant la convergence théorique d'un algorithme d'ABF adapté dans le chapitre 6. L'étude de convergence en temps long utilise des méthodes d'entropie relative et des inégalités de Sobolev logarithmiques.

Deterministic and stochastic methods for molecular simulation

Abstract

Molecular simulation is an essential tool in understanding complex chemical and biochemical processes as real-life experiments prove increasingly costly or realistically infeasible. This thesis is devoted to methodological aspects of molecular simulation, with a particular focus on computing transition paths and their associated free energy profiles. The first part is dedicated to computational methods for reaction path and transition state searches on a potential energy surface. In Chapter 3 we propose an improvement to a widely-used transition state search method, the Activation Relaxation Technique (ART). We also present a local convergence study of a prototypical algorithm. The second part is dedicated to free energy computations. We focus in particular on an adaptive importance sampling technique, the Adaptive Biasing Force (ABF) method. The first contribution to this field, presented in Chapter 5, consists in showing the applicability to a large molecular system of a new parallel implementation, named multiple-walker ABF (MW-ABF). Numerical experiments demonstrated the robustness of MW-ABF against artefacts arising due to poorly chosen or oversimplified reaction coordinates. These numerical findings inspired a new study of the longtime convergence of the ABF method, as presented in Chapter 6. By studying a slightly modified model, we back our numerical results by showing a faster theoretical rate of convergence of ABF than was previously shown.

Publications dans des revues internationales

1. T. Lelièvre, K. Minoukadeh. *Long-time convergence of an Adaptive Biasing Force method: the bi-channel case.* Soumis au journal Arch. Rational Mech. Anal. (2010). Preprint: hal-00477302.

2. K. Minoukadeh, C. Chipot, T. Lelièvre. *Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach.* J. Chem. Theory Comput. 6(4), 1008 – 1017 (2010).

3. E. Cancès, F. Legoll, M.-C. Marinica, K. Minoukadeh, F. Willaime. *Some improvements of the ART method for finding transition pathways on potential energy surfaces.* J. Chem. Phys. 130, 114711 (2009).

Proceedings

1. K. Minoukadeh. *Une méthode adaptative pour le calcul d'énergie libre.* Revue Femmes et Mathématiques, n°9 (2010).

Communications orales

1. Multiscale Molecular Modelling: Molecular Dynamics, Computational Statistical Mechanics, and Simulation Algorithms, Edinburgh, Scotland, June 30-July 3, 2010. *Adaptive biasing force simulations: overcoming local metastability.*

2. SIAM: Emerging Topics in Dynamical Systems and Partial Differential Equations, Barcelona, Spain, May 31-June 4, 2010. *Long-time convergence of an Adaptive Biasing Force method: the bi-channel case.*

3. Groupe de travail des thésards CEREMADE, Université Paris-Dauphine, France, May 20, 2010. *Long-time convergence of an Adaptive Biasing Force method: the bi-channel case.*

4. Neuvième forum des jeunes mathématiciennes, Institut Henri Poincaré, Paris, France, November 6-7, 2009. *An adaptive method for free energy computations.*

5. Workshop LN3M 2009, ENS Lyon, France, September 28-30, 2009. *An adaptive method for free energy computations.*

6. Réunion HYBRID 2009, ENS Cachan Bretagne, France, April 28-29, 2009. *An Adaptive Method for Free Energy Calculations.*

Posters

1. Foundations of Modeling and Molecular Simulation (FOMMS) 2009, Semiahmoo Resort, Blaine, WA, July 12-16, 2009. *Poster: Adaptive Biasing Force Simulations with Multiple Replicas.*

2. Molecular Simulations: Algorithms, Analysis and Applications, IMA, University of Minnesota, Minneapolis, MN, May 18-22, 2009. *Poster: Some improvements of the Activation-Relaxation Technique (ART).*

3. Rare Events in High-Dimensional Systems, IPAM, UCLA, Los Angeles, CA, February 23-27, 2009. *Poster: Some improvements of the Activation-Relaxation Technique (ART).*

Contents

1 Introduction **1**

1.1 Introduction à la simulation moléculaire . . . 1

1.1.1 Calcul de moyennes . . . 3

1.1.2 Les ensembles thermodynamiques . . . 4

1.1.3 Énergie potentielle . . . 6

1.2 Méthodes d'échantillonnage . . . 10

1.2.1 Algorithme de Metropolis-Hastings . . . 10

1.2.2 Dynamique moléculaire . . . 12

1.3 Simuler les systèmes sur des temps longs . . . 17

1.3.1 Recherche de chemins de transition . . . 18

1.3.2 Simuler sur les temps plus longs . . . 19

I Finding transition paths **21**

2 Finding transition paths **23**

2.1 Simulations on large time scales . . . 23

2.2 Double-ended methods . . . 24

2.2.1 Nudged Elastic Band method . . . 25

2.2.2 The String method . . . 28

2.2.3 Transitions at finite temperature . . . 29

2.3 Single-ended methods . . . 30

2.3.1 Locating transition states in nonconvex regions . . . 31

2.3.2 Climbing out of convex regions . . . 37

3 An improvement to the Activation-Relaxation Technique **39**

3.1 Introduction . . . 39

3.2 A new type of ART-like algorithms . . . 40

3.3 Numerical results: Migration of point defects in α-iron . . . 43

3.A Appendix: Local convergence analysis . . . 47

II Free Energy Computations **51**

4 Free Energy Computations **53**

4.1 Introduction . . . 53

4.1.1 Describing a transition . . . 55

4.1.2 Choosing a good reaction coordinate . . . 58

4.2 Computing free energy differences . . . 60

4.2.1 Thermodynamic integration . . . 60

4.2.2 Other equilibrium free energy methods . . . 64
4.2.3 Nonequilibrium free energy methods . . . 67
4.3 Adaptive sampling methods . . . 68
4.3.1 Adaptive Biasing Potential . . . 69
4.3.2 Adaptive Biasing Force . . . 71
4.4 Longtime convergence of dynamics . . . 73
4.4.1 Entropy and the Logarithmic Sobolev Inequality . . . 73
4.4.2 Convergence of the overdamped dynamics . . . 76
4.4.3 Convergence of ABF dynamics . . . 77

5 A Multiple-Walker ABF method 81
5.1 Introduction . . . 81
5.2 General setting . . . 83
5.3 Adaptive biasing force methods . . . 84
5.3.1 Framework . . . 84
5.3.2 Calculating the bias . . . 85
5.3.3 Enhancing sampling through selection . . . 86
5.4 Implementation details . . . 89
5.5 Numerical results . . . 93
5.5.1 Reaction coordinate range: 12-32 Å . . . 94
5.5.2 Reaction coordinate range: 4-16 Å . . . 94
5.5.3 Selection . . . 97
5.6 Discussion . . . 98

6 Long-time convergence of an ABF method 101
6.1 Introduction . . . 101
6.1.1 The adaptive biasing force method . . . 103
6.1.2 Existing convergence results, and the multiple channel scenario 104
6.2 The bi-channel model and statement of the main results . . . 105
6.2.1 The bi-channel model . . . 106
6.2.2 A partial differential equation formulation . . . 107
6.2.3 Entropy and Fisher information . . . 111
6.2.4 The free energy as a bias in each channel . . . 114
6.2.5 Main result . . . 115
6.3 Proof of main result . . . 116
6.3.1 Preliminary computations on the total entropy E . . . 116
6.3.2 Controlling E_M . . . 122
6.3.3 Controlling E_m . . . 122
6.3.4 Controlling E_c . . . 125
6.3.5 Completing the proof . . . 130

Bibliography 133

Chapter 1

Introduction

Sommaire

1.1 Introduction à la simulation moléculaire 1
1.1.1 Calcul de moyennes 3
1.1.2 Les ensembles thermodynamiques 4
1.1.3 Énergie potentielle 6
1.2 Méthodes d'échantillonnage 10
1.2.1 Algorithme de Metropolis-Hastings 10
1.2.2 Dynamique moléculaire 12
1.3 Simuler les systèmes sur des temps longs 17
1.3.1 Recherche de chemins de transition 18
1.3.2 Simuler sur les temps plus longs 19

1.1 Introduction à la simulation moléculaire

La simulation moléculaire est un outil indispensable pour comprendre le comportement des systèmes physiques (matériaux) ou biologiques (protéines) lorsque les expérimentations sont difficiles ou coûteuses. La simulation moléculaire donne un moyen de calculer les propriétés moyennes de tels systèmes en utilisant les lois de la physique statistique.

On considère un système de N atomes, décrit par les positions $q = (q_1, \ldots, q_N)$ et impulsions $p = (p_1, \ldots, p_N)$, où $q_i \in \mathbb{R}^3$ et $p_i \in \mathbb{R}^3$ désignent, respectivement, la position et l'impulsion du i-ème atome. Dans la suite on note $\Omega = \mathbb{R}^{3N} \times \mathbb{R}^{3N}$ l'espace des phases. Les particules interagissent selon la fonction d'énergie potentielle $V : \mathbb{R}^{3N} \to \mathbb{R}$, définie sur le domaine des configurations $\mathcal{D} = \{q \in \mathbb{R}^{3N} \mid V(q) < \infty\}$. L'énergie totale du système est donnée par l'Hamiltonien

$$H(q,p) = \frac{1}{2} p^T M^{-1} p + V(q), \tag{1.1}$$

où $M = \mathrm{diag}(m_{11}, m_{11}, m_{11}, ..., m_{NN}, m_{NN}, m_{NN})$ et m_{ii}, désignent la masse de l'atome i. Le premier terme correspond à l'énergie cinétique et le deuxième à l'énergie potentielle. Il existe de nombreuses façons de calculer cette énergie potentielle; cf. la Section 1.1.3.

Temps	Échelle (s)	Observations physiques	Simulation moléculaire
fs	10^{-15}	Vibration atomique	Pas de temps
ps	10^{-12}	Extension de liaisons	
ns	10^{-9}	Repliement peptidique	Échelle de temps accessible (sur machines 'ordinaires')
μs	10^{-6}	Réplication de l'ADN	Échelle de temps accessible (sur machines 'performantes')
ms	10^{-3}	Réactions chimiques	
s	1	Repliement de protéines	

Table 1.1: Les phénomènes physiques se produisent à des échelles plus longues que celles accessibles par la simulation moléculaire habituelle.

L'évolution en temps d'un système $(q(t), p(t))_{t\geq 0}$ peut être décrite, une fois l'énergie potentielle donnée, par des dynamiques appropriées. Des exemples et des méthodes d'intégration numériques sont présentés dans la Section 1.2.2. En pratique, les échelles de temps accessibles par des simulations habituelles sont limitées à quelques nanosecondes (quelques microsecondes sur les machines les plus puissantes) par journée de simulation. Ceci est dû aux vibrations atomiques (liaisons covalentes) qui limitent le pas de temps de la discrétisation numérique. Cette description microscopique du système est donc rarement suffisante pour décrire toutes les transitions importantes d'un système. Le tableau 1.1 recense, pour différents types d'observations ou phénomènes physiques, les échelles de temps correspondantes, ainsi que les éventuelles limitations numériques qui en découlent. Les phénomènes physiques qui se manifestent au-delà de quelques microsecondes (par exemple le repliement de protéines) sont donc rarement capturés par la simulation moléculaire habituelle. La restriction du pas de temps (1 fs) fait en sorte qu'au moins 10^{12} pas de la dynamique moléculaire soient nécessaires pour observer des phénomènes tels que la réplication de l'ADN (~ 1 μs) ou le repliement de protéines (~ 1 ms).

Il est important de noter que les échelles de temps accessibles à la simulation moléculaire – soit quelques microsecondes par jour – augmentent grâce aux progrès algorithmiques et informatiques. Le travail effectué au cours de cette thèse est essentiellement relatif au premier aspect. Il concerne plus particulièrement des méthodes pour la recherche de chemins de réactions et pour le calcul d'énergie libre. Nous proposons, dans un premier temps, une amélioration d'une méthode déjà existante de recherche de chemins de réaction, et en présentons une étude de convergence (cf. chapitre 3). Nous étudions ensuite l'applicabilité d'une méthode probabiliste de calcul d'énergie libre, récemment proposée, à un système biomoléculaire de grande dimension (cf. chapitre 5), dont les résultats prometteurs inspirent une étude mathématique de convergence (cf. chapitre 6).

1.1.1 Calcul de moyennes

La physique statistique permet de faire le lien entre un modèle à l'échelle moléculaire et la quantité macroscopique via le calcul de la moyenne d'une observable $A(q,p)$, en fonction de l'état microscopique du système. Cette moyenne est calculée par rapport à une mesure de probabilité $d\mu$ donnée. La mesure $d\mu$ est associée à un ensemble thermodynamique, dont quelques exemples seront donnés dans la Section 1.1.2. Pour le moment, on ne précise pas l'ensemble statistique utilisé; les notions suivantes sont valables pour toutes les mesures $d\mu$ présentées plus loin.

La moyenne de toute observable $A \in L^1(\mu)$ s'écrit sous la forme

$$\langle A \rangle = \int_\Omega A(q,p)\ d\mu(q,p). \tag{1.2}$$

On peut s'intéresser par exemple à l'énergie moyenne du système, donné par $\langle H \rangle$, ou bien à la capacité calorifique $C = \langle H^2 \rangle - \langle H \rangle$. Le calcul analytique de l'intégrale (1.2) n'est, en général, pas possible en particulier à cause de la taille du domaine Ω (la plupart des systèmes d'intérêt sont constitués de plus de $N \sim 10^5$ particules). En outre, il existe des régions du domaine, $D \subset \Omega$ (de taille importante), telles que $\mu(D) \approx 0$ et qui en conséquence n'influencent pas la valeur de l'intégrale. En pratique, on est donc contraint d'approcher l'intégrale en utilisant des méthodes probabilistes, qui fassent en sorte que seules les régions les plus importantes (de forte probabilité) soient échantillonnées. Les méthodes probabilistes se déclinent en deux catégories principales.

Méthode de Monte Carlo standard

Cette méthode purement probabiliste consiste à considérer une suite de points $(q_n, p_n)_{n\in\mathbb{N}}$, indépendantes et identiquement distribuées (i.i.d.) selon la mesure de probabilité $d\mu$, en utilisant, par exemple, un algorithme de rejet. La méthode de rejet consiste à simuler une loi de proposition $d\tilde{\mu}$ plus simple et accepter ou rejeter les points selon un certain critère. La loi des grands nombres et le théorème de la limite centrale nous permettent alors d'approcher la moyenne $\langle A \rangle$ par

$$\langle A \rangle \approx \frac{1}{N} \sum_{n=1}^{N} A(q_n, p_n). \tag{1.3}$$

L'efficacité de l'algorithme de rejet (c'est-à-dire la proportion des points acceptés) dépend fortement de la proximité des mesures $d\tilde{\mu}$ et $d\mu$. La difficulté de choisir une loi de proposition proche de la mesure canonique (surtout en grande dimension) fait que cette méthode est rarement utilisée en pratique.

Dynamique moléculaire

L'approche dynamique moléculaire est celle à laquelle on s'intéresse dans le cadre de cette thèse. Cette approche consiste à définir une dynamique sur $(q(t), p(t))$ qui

est ergodique par rapport à la mesure $d\mu$, c'est-à-dire pour toute fonction $f \in L^1(\mu)$ et pour toute condition initiale $q(0) = q_0, p(0) = p_0$, on a

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(q(t), p(t))\ dt = \int_\Omega f(q,p)\ d\mu(q,p).$$

Ainsi, sous l'hypothèse d'ergodicité, la moyenne $\langle A \rangle$ est estimée par

$$\langle A \rangle \approx \frac{1}{T} \int_0^T A(q(t), p(t))dt, \tag{1.4}$$

pour $\mu-$presque toute condition initiale (q_0, p_0). En pratique, la dynamique est discrétisée et intégrée en utilisant un schéma d'intégration numérique fiable (voir la Section 1.2.2.2). Notons $\{(q_{i\Delta t}, p_{i\Delta t})\}_{i=0,\dots,N}$ la trajectoire induite par la dynamique discrétisée, où Δt désigne le pas de temps de discrétisation et $T = N\Delta t$. L'intégrale intervenant dans (1.4) est alors approchée par la somme de Riemann

$$\langle A \rangle \approx \frac{1}{N+1} \sum_{i=0}^{N} A(q_{i\Delta t}, p_{i\Delta t}). \tag{1.5}$$

1.1.2 Les ensembles thermodynamiques

La mesure $d\mu$ décrit les états microscopiques d'un système dans un ensemble thermodynamique donné. Ainsi, $d\mu(q,p)$ donne la probabilité que les positions et impulsions des particules du système se trouvent dans les intervalles $[q, q+dq]$ et $[p, p+dp]$ respectivement. Nous présentons brièvement les quatre ensembles thermodynamiques principaux, tout en nous focalisant sur les deux premiers.

Ensemble microcanonique

Pour un système isolé d'énergie constante, on se place dans le cadre de l'ensemble microcanonique, ou ensemble (N, V, E), à nombre d'atomes N, volume V et énergie E constants. La mesure microcanonique est définie par

$$d\mu_{\text{mc}} = \frac{d\sigma_E}{|\nabla H|},$$

où $d\sigma_E$ est la mesure de Lebesgue induite par la mesure de Lebesgue dans Ω et le produit scalaire euclidien sur la sous-variété

$$\{(q,p) \in \Omega \mid H(q,p) = E\}.$$

Ensemble canonique

Dans le cadre de cette thèse, on s'intéresse à l'ensemble canonique aussi appelé ensemble (N, V, T), c'est-à-dire à nombre de particules N, volume V et température

T fixés. La mesure de probabilité associée aux états microscopiques du système pour cet ensemble est donnée par la *mesure canonique*

$$d\mu_{\mathrm{c}}(q,p) = Z^{-1}\exp(-\beta H(q,p))\ dq\ dp, \tag{1.6}$$

où Z, le facteur de normalisation, est défini par

$$Z = \int_{\Omega} \exp(-\beta H(q,p))\ dq\ dp,$$

et $\beta = 1/k_B T$ ($k_B \approx 1,38 \times 10^{-23}$ J/K étant la constante de Boltzmann et T la température, mesurée en Kelvin (K)). L'Hamiltonien (1.1) étant séparable, la mesure canonique peut s'écrire comme un produit de deux mesures:

$$d\mu_{\mathrm{c}}(q,p) = d\kappa(p)d\pi(q), \tag{1.7}$$

où

$$d\kappa(p) = Z_p^{-1}\exp\left(-\beta\frac{p^T M^{-1} p}{2}\right)dp, \tag{1.8}$$

$$d\pi(q) = Z_q^{-1}\exp\left(-\beta V(q)\right)dq. \tag{1.9}$$

Ici,

$$Z_p = \int_{\mathbb{R}^{3N}} \exp\left(-\beta\frac{p^T M^{-1} p}{2}\right)dp \quad \text{et} \quad Z_q = \int_{\mathbb{R}^{3N}} \exp\left(-\beta V(q)\right)dq$$

sont des facteurs de normalisation. On remarque que la mesure $d\kappa$, selon laquelle les impulsions sont distribuées, n'est rien d'autre qu'une loi normale. Cette décomposition sera utile dans la suite pour la raison suivante. Lorsque la quantité d'intérêt est la moyenne d'ensemble et non pas la dynamique du système, il suffit de tirer les impulsions selon la loi normale. On peut ainsi se contenter de décrire une dynamique uniquement sur les positions q telle que, en temps long, elles soient distribuées selon la loi $d\pi$. Ceci explique pourquoi, plus tard, on se contentera d'une dynamique sur les positions q (cf. la dynamique de Langevin *amortie* (1.29)) et non pas une dynamique sur (q,p) (cf. la dynamique de Langevin (1.25)).

Ensemble grand-canonique

Lorsque l'on considère un système ouvert, qui peut échanger de l'énergie et des particules avec un réservoir externe, on se place dans l'ensemble grand-canonique. Dans cet ensemble, à potentiel chimique, volume et température fixés, seuls l'énergie et le nombre de particules fluctuent.

Ensemble isotherme-isobare

Lorsque l'on considère un système à nombre de particules N, pression P et température T fixés, il s'agit de l'ensemble (N,P,T) ou isotherme-isobare. Le système est en contact avec un réservoir de d'énergie et peut changer de volume, afin d'assurer une pression P et une température T constantes.

1.1.3 Énergie potentielle

Il existe de nombreuses manières de modéliser les interactions atomiques des systèmes à l'échelle moléculaire. Les modèles les plus précis proviennent des modèles *ab initio* de la chimie quantique. Ces modèles se basent sur des systèmes moléculaires constitués de (i) N noyaux atomiques, dont les positions sont désignées par $q_1, ..., q_N$, $q_i \in \mathbb{R}^3$ et les charges électriques $Z_i \in \mathbb{N}^*$ et de (ii) M électrons, dont les positions sont notées $x_1, ..., x_M$, $x_j \in \mathbb{R}^3$, décrit par une fonction d'onde ψ. Afin de simplifier l'écriture des équations, on travaille habituellement avec les unités atomiques, obtenues en imposant

$$m_e = 1, \quad e = 1, \quad \hbar = 1, \quad \frac{1}{4\pi\epsilon_0} = 1,$$

où m_e est la masse de l'électron, e la charge électrique élémentaire, $\hbar$ la constante de Planck réduite et ϵ_0 la permittivité diélectrique du vide. Dans ce système d'unité, l'unité de masse est $9,11 \times 10^{-31}$ kg, l'unité de longueur est le rayon de Bohr $a_0 = 5,29 \times 10^{-11}$ m, l'unité de temps est $2,42 \times 10^{-17}$ s, et l'unité d'énergie est le Hartree Ha $= 4,36 \times 10^{-18}$ J $= 627$ kcal/mol.

D'après l'hypothèse de Born et Oppenheimer, on suppose que les positions des noyaux q_i de charge Z_i sont fixes, et l'énergie du système est la somme de deux termes: le potentiel de Coulomb entre les noyaux et l'énergie de l'état électrique fondamental (l'état de plus basse énergie):

$$V(q_1, ..., q_N) = V_{\text{coulomb}}(q_1, ..., q_N) + V_{\text{elec}}(q_1, ..., q_N), \tag{1.10}$$

où le potentiel de Coulomb décrit l'interaction entre les noyaux chargés

$$V_{\text{coulomb}}(q_1, ..., q_N) = \sum_{1 \leq i < j \leq N} \frac{Z_i Z_j}{|q_i - q_j|}, \tag{1.11}$$

et le potentiel V_{elec} est la fonction qui minimise le problème électronique sur l'espace de Hilbert $\mathcal{H} \subset \bigwedge_{m=1}^{M} L^2(\mathbb{R}^3, \mathbb{C})$ de fonctions d'onde admissibles

$$V_{\text{elec}}(q_1, ..., q_N) = \inf \left\{ \langle \psi, \hat{H}_{q_1,...,q_N} \psi \rangle_{\mathcal{H}} \mid \psi \in \mathcal{H}, \|\psi\|_{L^2} = 1 \right\} \tag{1.12}$$

où $\hat{H}$ est l'Hamiltonien électronique qui s'écrit

$$\hat{H}_{q_1,...,q_N} = -\sum_{m=1}^{M} \frac{1}{2} \Delta_{x_m} - \sum_{m=1}^{M} \sum_{i=1}^{N} \frac{Z_i}{|x_m - q_i|} + \sum_{1 \leq n < m \leq M} \frac{1}{|x_n - x_m|}.$$

En raison de la taille de l'ensemble de fonctions d'onde admissible, on ne peut résoudre le problème de minimisation (1.12) que pour les systèmes de petite taille (un ou deux électrons). Pour des systèmes plus complexes on a recours à des modèles d'approximation, telles que le modèle de Hartree-Fock ou les modèles de Kohn-Sham [Kohn 1965] (issus de la Théorie de la Fonctionnelle de la Densité), voir [Cancès 2003] pour plus de détails.

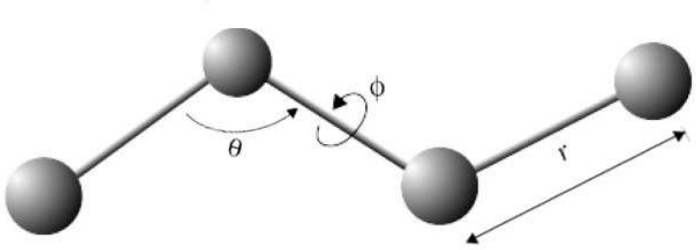


Figure 1.1: Les interactions interatomiques: r désigne la distance entre deux atomes, θ est l'angle entre deux liaisons reliant trois atomes et ϕ est l'angle dièdre qui séparent deux plans contenant chacun trois atomes.

Pour des systèmes de très grande dimension, le potentiel à N-corps quantique n'est pas utilisé en pratique puisqu'il est le plus souvent impossible de le calculer en un temps raisonnable. Pour cette raison, des efforts ont été menés afin de développer des fonctions d'énergie potentielle (ou *champs de force*) empiriques. Les potentiels empiriques, dont quelques exemples sont donnés ci-dessous, ont pour but de reproduire – au moindre coût de calcul – des données expérimentales et calculs précis issus de la mécanique quantique.

1.1.3.1 Champs de force pour les biomolécules

Dans le domaine de la biologie moléculaire, il est commode d'exprimer le champ de force comme une somme de trois termes: l'énergie due à la présence du potentiel extérieur V_{ext} (électrostatique, magnétique,...); l'énergie $V_{\text{liés}}$ provenant de l'interaction entre les atomes liés par une liaison covalente et l'énergie $V_{\text{non-liés}}$ issue de l'interaction entre atomes non-liés

$$V(q) \quad = \quad V_{\text{ext}}(q) + V_{\text{liés}}(q) + V_{\text{non-liés}}(q), \tag{1.13}$$

où $q = (q_1, ..., q_N) \in \mathbb{R}^{3N}$, représente les positions des N atomes du système. En pratique le potentiel modélisant les interactions des atomes liés est beaucoup moins coûteux à évaluer puisqu'il ne s'agit que de termes locaux. L'interaction entre deux atomes liés par une liaison covalente peut par exemple être exprimée par un potentiel harmonique

$$V_2(q_i, q_{i+1}) = \frac{k}{2}(|q_i - q_{i+1}| - r_{\text{eq}})^2, \tag{1.14}$$

où $k > 0$ est une constante à déterminer et r_{eq} désigne la longueur d'équilibre de la liason. On peut, de la même manière, définir un angle d'équilibre θ_{eq} entre deux liaisons covalentes consécutives de telle sorte que le potentiel à trois corps s'écrit

$$V_3(q_i, q_{i+1}, q_{i+2}) = \frac{k_\theta}{2}(\theta_i - \theta_{\text{eq}})^2, \tag{1.15}$$

où $k_\theta > 0$ est une constante et θ_i est l'angle entre la liaison reliant les atomes i et $i+1$ et celle reliant $i+1$ et $i+2$. L'angle satisfait la relation suivante

$$\cos\theta_i = \left(\frac{r_{i,i+1}}{|r_{i,i+1}|} \cdot \frac{r_{i+1,i+2}}{|r_{i+1,i+2}|} \right).$$

où $r_{i,j} = q_i - q_j$. Enfin, l'interaction de quatre atomes liés est décrite par une fonction de l'angle dièdre ϕ_i

$$V_4(q_i, q_{i+1}, q_{i+2}, q_{i+2}) = u_\phi(\cos\phi_i), \tag{1.16}$$

où

$$\cos\phi_i = -\frac{r_{i,i+1} \times r_{i+1,i+2}}{|r_{i,i+1} \times r_{i+1,i+2}|} \cdot \frac{r_{i+1,i+2} \times r_{i+2,i+3}}{|r_{i+1,i+2} \times r_{i+2,i+3}|}.$$

Le potentiel u_ϕ s'écrit typiquement sous la forme

$$u_\phi(x) = c_1(1-x) + 2c_2(1-x^2) + c_3(1+3x-4x^3),$$

où les paramètres c_i $(i = 1, 2, 3)$ sont à déterminer, de telle sorte qu'il y ait trois angles stables (dont $\phi = 0$ est le plus favorable énergétiquement). La Figure 1.1 montre les paramètres décrits ci-dessous sur une chaîne de quatre atomes. La contribution des atomes liés s'écrit ainsi

$$V_{\text{liés}}(q) = \sum_{i=1}^{N-1} V_2(q_i, q_{i+1}) + \sum_{i=1}^{N-2} V_3(q_i, q_{i+1}, q_{i+2}) + \sum_{i=1}^{N-3} V_4(q_i, q_{i+1}, q_{i+2}, q_{i+3}).$$

La modélisation des interactions non-liées constitue un des grands défis de la simulation moléculaire de nos jours. En faisant intervenir tous les atomes du système, le coût de calcul croît quadratiquement avec la taille du système. L'interaction de paire dans un gaz monoatomique (contenant des atomes neutres) est souvent modélisée par le potentiel de Lennard-Jones [Lennard-Jones 1931] (voir la Figure 1.2)

$$V_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^{6} \right], \tag{1.17}$$

où r est la distance entre deux atomes, ϵ est l'amplitude du puits du potentiel et r_0 la distance à laquelle V_{LJ} vaut zéro. Ces paramètres sont choisis afin de reproduire des résultats expérimentaux. Le premier terme traduit l'interaction répulsive entre deux atomes neutres à courte distance (les nuages électroniques ne se superposent pas) et le deuxième traduit l'attractivité du potentiel à longue distance (interaction de van der Waals). Remarquons que le premier terme domine lorsque la distance entre les deux atomes est petite, $r < r_0$, le deuxième lorsque $r > r_0$. Afin de réduire le coût de calcul, ce potentiel est en pratique tronqué à une distance critique lorsque sa valeur devient très petite.

Finalement la contribution des atomes non-liés s'écrit

$$V_{\text{non-liés}}(q) = V_{\text{coulomb}}(q) + \sum_{1 \le i < j \le N} V_{\text{LJ}}(|q_i - q_j|),$$

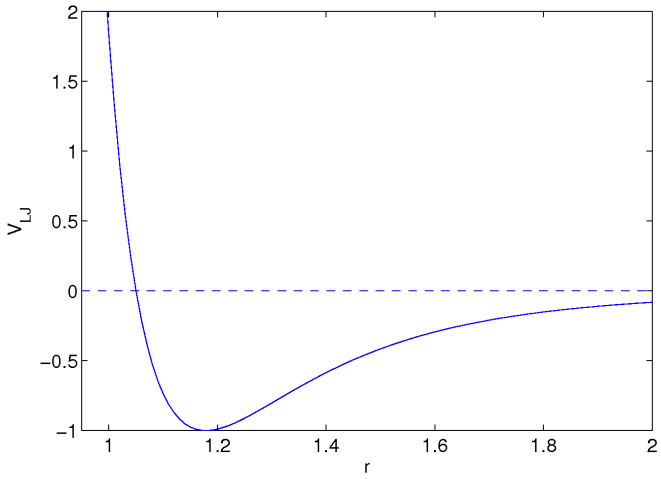


Figure 1.2: Potentiel de Lennard-Jones (1.17), avec $\epsilon = 1$, $r_0 = 1.05$.

où le potentiel Coulombien est défini par (1.11).

Le calcul pour les interactions d'atomes liés requiert $O(N)$ opérations alors que pour les atomes non-liés $O(N^2)$ opérations sont requises à cause du caractère à longue portée du potentiel Coulombien. En fait le coût de calcul de ce dernier peut être réduit à $O(N \log N)$ opérations en utilisant la *Fast Multipole Method* (FMM) [Greengard 1987], ou bien dans le cas d'un système périodique, la méthode de sommation d'Ewald [Ewald 1921, Darden 1999]. Les interactions à longue portée restent toujours les calculs les plus coûteux dans l'évaluation de force. Pour cette raison de nombreuses études sont consacrées à la réduction du coût de calcul de celles-ci.

Parmi les nombreux champs de force, on peut distinguer trois catégories principales:

1. les champs de force *tout-atome* font intervenir explicitement tous les atomes, y compris l'hydrogène;

2. les champs de force *atomes unifiés* ne font pas intervenir explicitement les atomes d'hydrogène, sauf si ces derniers sont polaires;

3. les champs de force *à gros grains* sont basés sur des modèles simplifiés et construits spécifiquement pour étudier les transitions conformationnelles à large échelle. En regroupant un ou plusieurs groupements atomiques en une seule entité, on réduit considérablement le coût de l'évaluation des forces.

Les ensembles de champs de force empiriques les plus utilisés pour la simulation biomoléculaire sont *AMBER* (Assisted Model Building and Energy Refinement) [Cornell 1995], *GROMOS* et *CHARMM* (Chemistry at HARvard Macromolecular Mechanics) [Brooks 1983, MacKerel Jr. 1998]. Ce dernier regroupe en

particulier CHARMM19 (atomes unifiés), CHARMM22 (tout-atome), utilisés pour la simulation des protéines et le champ de force CHARMM27. Ce dernier sera utilisé pour les simulations effectuées dans le chapitre 5.

1.1.3.2 Champs de force pour les sciences des matériaux

Dans le domaine des sciences des matériaux, l'un des potentiels semi-empiriques les plus utilisés est celui de la *(Modified) Embedded Atom Method* (EAM), développé par Daw et Baskes [Daw 1983, Daw 1984, Baskes 1992] pour les métaux, puis par Mendelev *et al.* [Mendelev 2003, Ackland 2004] pour le cas particulier du fer liquide et cristallin. Ce dernier est utilisé dans les simulations du Chapitre 3.

Ces potentiels semi-empiriques se basent sur l'idée suivante. Puisqu'un atome peut être considéré comme un défaut parmi d'autres atomes du système, l'énergie totale du système peut s'écrire sous la forme

$$V(q) = \sum_{i=1}^{N} F_\alpha(\varrho_\alpha^i(q)) + \sum_{i=1}^{N} \varphi_i(q), \tag{1.18}$$

où F_α, appelée fonction 'embedding', désigne l'énergie requise pour placer l'atome i de type α dans la nuage électronique. La fonction ϱ_α^i est la somme des contributions à la densité de charge électrique des atomes à la position q_i, écrite sous la forme

$$\varrho_\alpha^i(q) = \sum_{j \neq i} \rho_\alpha(|q_i - q_j|), \tag{1.19}$$

et enfin φ_i est la somme des interactions de paire d'atomes i et j, de type α et β respectivement

$$\varphi_i(q) = \frac{1}{2} \sum_{j \neq i} \phi_{\alpha,\beta}(|q_i - q_j|). \tag{1.20}$$

Les fonctions F_α et ρ_α et d'autres paramètres intervenants sont optimisés afin de reproduire une base de données de référence.

1.2 Méthodes d'échantillonnage

Cette partie donne un aperçu des méthodes probabilistes les plus commodes pour échantillonner une mesure donnée. On aborde, dans la Section 1.2.1, les méthodes purement probabilistes, avant de présenter dans la Section 1.2.2 les dynamiques stochastiques pour la simulation moléculaire.

1.2.1 Algorithme de Metropolis-Hastings

Il existe des méthodes telles que les algorithmes de Metropolis datant de 1953 [Metropolis 1953], et de Metropolis-Hastings datant de 1970 [Hastings 1970], qui permettent d'échantillonner une mesure $\nu(q)dq$ (connue à une constante de normalisation près) à partir d'une densité de proposition $P(q'|q)$, représentant la probabilité de transition $q \to q'$. Dans le cas d'une densité de proposition symétrique,

$P(q'|q) = P(q|q')$ on parle d'algorithme de Metropolis. Supposons qu'à partir d'un point q^n, une nouvelle configuration $\tilde{q}^{n+1}$ a été choisie; alors la probabilité d'accepter cette dernière est

$$a(q^n, \tilde{q}^{n+1}) = \min\left(\frac{\nu(\tilde{q}^{n+1})}{\nu(q^n)}, 1\right).$$

Les fonctions de proposition qui décrivent la distribution uniforme et la loi normale, par exemple, sont symétriques. Pour cette dernière, il s'agit d'une nouvelle configuration

$$\tilde{q}^{n+1} = q^n + g \quad \text{où } g_i \sim \mathcal{N}(0, \sigma^2),$$

La distribution de proposition symétrique est, dans ce cas, donnée par

$$P(q'|q) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{|q-q'|^2}{2\sigma^2}\right).$$

Bien que l'algorithme de Metropolis soit exact, il risque de rejeter une proportion importante de configurations proposées. L'algorithme de Metropolis-Hastings généralise celui de Metropolis aux distributions de proposition P asymétriques, c'est-à-dire telle que $P(q'|q) \neq P(q|q')$ (voir l'Algorithme 1). Prenons, par exemple, la proposition de Metropolis-Adjusted Langevin Algorithm (MALA) [Roberts 1998]

$$\tilde{q}^{n+1} = q^n - h\nabla V(q^n) + \sqrt{2\beta^{-1}h}\, G^n, \tag{1.21}$$

où $G^n \sim \mathcal{N}(0, I)$, I est la matrice d'identité de taille $3N \times 3N$ et h est un paramètre de discrétisation (que l'on peut voir comme un pas de temps). Cette proposition est en fait une discrétisation de la dynamique de Langevin amortie, permettant d'échantillonner la mesure canonique, qui sera présentée en détail plus loin (voir équation (1.29) dans la Section 1.2.2). La fonction de proposition associée à (1.21) est donnée par

$$P(q'|q) = \left(\frac{\beta}{4\pi h}\right)^{3N/2} \exp\left(-\beta\frac{|q-q'+h\nabla V(q)|^2}{4h}\right).$$

Cette fonction de proposition a tendance à choisir les configurations de plus forte probabilité de la mesure canonique (pour lequel la densité de probabilité est donnée par $\nu(q) = Z_q^{-1} e^{-\beta V(q)}$, cf. (1.9)), et par conséquent à diminuer le taux de rejet de l'algorithme. La probabilité de passer de q vers q' est définie par

$$\rho(q, dq') = P(q'|q)a(q, q')dq' + \delta_q(q')\,(1 - \alpha(q))\,dq', \tag{1.22}$$

où $\alpha \in [0, 1]$ est la probabilité d'accepter un pas de l'algorithme, partant du point q:

$$\alpha(q) = \int_{\tilde{q}\in\mathcal{D}} P(\tilde{q}|q)a(q, \tilde{q})d\tilde{q}.$$

Le terme $\delta_q(q')\,(1 - \alpha(q))\,dq'$ dans (1.22) résume tous les mouvements rejetés et est alors la probabilité de rester à la position initiale q.

Une preuve de convergence pour l'algorithme de Metropolis-Hastings se trouve dans [Meyn 1993, Mengersen 1996]. Elle est dûe essentiellement au fait que

Algorithme 1 Algorithme de Metropolis-Hastings

Pour $n \geq 0$

1. proposer une nouvelle configuration $\tilde{q}^{n+1}$ à partir de la configuration q^n, avec une probabilité $P(\tilde{q}^{n+1}|q^n)d\tilde{q}^{n+1}$.

2. accepter la nouvelle configuration avec une probabilité

$$a(q^n, \tilde{q}^{n+1}) = \min\left(\frac{\nu(\tilde{q}^{n+1})P(q^n|\tilde{q}^{n+1})}{\nu(q^n)P(\tilde{q}^{n+1}|q^n)}, 1\right),$$

auquel cas $q^{n+1} = \tilde{q}^{n+1}$. Sinon $q^{n+1} = q^n$.

(i) La condition de bilan détaillé est satisfaite: $\forall q, q' \in \mathcal{D}$

$$\rho(q, q')\nu(q) = \rho(q', q)\nu(q').$$

En d'autres termes, ν est une mesure réversible pour la dynamique;

(ii) La chaîne de Markov est ν-irréductible: Notons la n-ième pas de la probabilité de transition

$$\rho^{(n)}(q, q') = \int_{\tilde{q}\in\mathcal{D}} \rho(q, \tilde{q})\rho^{(n-1)}(\tilde{q}, q')d\tilde{q},$$

avec $\rho^{(1)}(q, q') = \rho(q, q')$. Alors, pour tout ensemble A mesurable tel que $\nu(A) > 0$ et pour $\nu-$presque toute condition initiale q, il existe $n_0 > 0$ tel que $\forall n \geq n_0$

$$\rho^{(n)}(q, q') > 0.$$

Ainsi, partant d'un point q, tout ensemble A peut être atteint en n pas avec une probabilité non nulle.

1.2.2 Dynamique moléculaire

Rappelons que la moyenne d'ensemble peut être estimée par une moyenne temporelle (1.4) lorsque la position et l'impulsion $(q(t), p(t))$ d'un système suivent une dynamique appropriée. Cette section a pour but de présenter les dynamiques souvent employées pour échantillonner la mesure canonique (1.6). Ces dernières sont construites à partir de la dynamique d'un système classique isolé (d'énergie constante), décrite par les équations de Hamilton. Dans la suite, $q = (q_1, q_2, ..., q_{3N})$ et $p = (p_1, p_2, ..., p_{3N})$.

1.2.2.1 Dynamique hamiltonienne

À partir d'une condition initiale $(q(0), p(0))$ donnée, l'évolution en temps d'un système isolé est décrite par les équations de Hamilton:

$$\begin{cases} \dfrac{dq}{dt} &= M^{-1}p = \nabla_p H, \\ \dfrac{dp}{dt} &= -\nabla V = -\nabla_q H, \end{cases} \tag{1.23}$$

où H est l'Hamiltonien défini par (1.1), $\nabla_p H = (\partial H/\partial p_1, ..., \partial H/\partial p_{3N})$ et $\nabla_q H = (\partial H/\partial q_1, ..., \partial H/\partial q_{3N})$. Il est facile de vérifier que l'énergie totale du système est conservée. On obtient, en dérivant $E(t) = H(q(t), p(t))$ par rapport au temps et en substituant (1.23),

$$\frac{dE}{dt} = \nabla_q H \cdot \frac{dq}{dt} + \nabla_p H \cdot \frac{dp}{dt} = \sum_{i=1}^{3N} \frac{\partial V}{\partial q_i}\frac{dq_i}{dt} + \sum_{i=1}^{3N} m_i^{-1} p_i \frac{dp_i}{dt} = 0.$$

En pratique, puisque l'on s'intéresse à des intégrations en temps longs, la dynamique (1.23) est intégrée en utilisant des méthodes numériques symplectiques telles que l'algorithme de Störmer-Verlet [Verlet 1967]. Pour un pas de temps Δt donné, cet algorithme consiste à faire un demi-pas en temps sur les impulsions (calculer $P_{t+\Delta t/2}$), ensuite un pas en temps sur les positions (calculer $Q_{t+\Delta t}$ en utilisant $P_{t+\Delta t/2}$), et enfin un demi-pas à nouveau sur les impulsions. Partant d'une configuration initiale (Q_0, P_0),

$$\begin{cases} P_{t+\Delta t/2} &= P_t - \nabla V(Q_t)\dfrac{\Delta t}{2}, \\ Q_{t+\Delta t} &= Q_t + \Delta t M^{-1} P_{t+\Delta t/2}, \\ P_{t+\Delta t} &= P_{t+\Delta t/2} - \nabla V(Q_{t+\Delta t})\dfrac{\Delta t}{2}. \end{cases} \tag{1.24}$$

Cet algorithme est en effet plus adapté que d'autres tels que les méthodes d'Euler puisqu'il garantit la (presque) conservation de l'énergie totale H en temps longs (voir [Hairer 2003]).

1.2.2.2 Dynamique de Langevin

Nous nous intéressons dans le cadre de cette thèse, à l'étude de systèmes moléculaire à température constante (ensemble canonique). La température du système est gardée constante grâce à l'échange d'énergie avec un bain thermal. Lorsqu'un système interagit avec un bain thermal externe, les particules subissent des petites fluctuations aléatoires. Ces fluctuations sont souvent décrites par un mouvement brownien [Brown 1828]. Le modèle de Langevin [Langevin 1908] intègre une force fluctuante et un terme de frottement proportionnel à la vitesse du système. La

dynamique de Langevin est donc une perturbation de la dynamique hamiltonienne, décrite par un système d'équations différentielles stochastiques (EDS):

$$\begin{cases} dQ_t = M^{-1}P_t\ dt, \\ dP_t = -\nabla V(Q_t)\ dt - \gamma M^{-1}P_t\ dt + \sqrt{2\gamma\beta^{-1}}dW_t, \end{cases} \tag{1.25}$$

où $(Q_t, P_t) \in \mathbb{R}^{6N}$ représente les positions et impulsions des particules du système au temps t, $\gamma > 0$ le coefficient de frottement et W_t est un mouvement Brownien. On note dans la suite ψ_t la densité de la loi de (Q_t, P_t). Sous certaines hypothèses sur V, notamment (1.26) et (1.27) ci-dessous, le système d'EDS admet une solution unique (Q_t, P_t) et est ergodique par rapport à la mesure $d\mu_c$. Ceci permet d'estimer la moyenne $\langle A\rangle$ par une moyenne trajectorielle (1.4). Le générateur infinitésimal $\mathcal{L}$ associé à la dynamique est donné par

$$\mathcal{L}\varphi = M^{-1}p\cdot\nabla_q\varphi - (\gamma M^{-1}p + \nabla V(q))\cdot\nabla_p\varphi + \gamma\beta^{-1}\Delta_p\varphi.$$

L'évolution de la densité de probabilité de (Q_t, P_t) est donnée par l'équation de Fokker-Planck

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

où $\mathcal{L}^*$ est l'opérateur adjoint de $\mathcal{L}$, défini par

$$\mathcal{L}^*\psi_t = -\mathrm{div}_q\left(M^{-1}p\ \psi_t\right) + \mathrm{div}_p\left([\gamma M^{-1}p + \nabla V(q)]\psi_t\right) + \gamma\beta^{-1}\Delta_p\psi_t.$$

On vérifie aisément que la densité $Z^{-1}\exp(-\beta H(q,p))$ de la mesure canonique (1.6) est bien une solution stationnaire de l'équation Fokker-Planck.

Il est important de noter que l'existence et l'unicité d'une solution (Q_t, P_t) au système (1.25) ne peuvent être assurées que sous certaines hypothèses sur le potentiel V. Les hypothèses, que l'on retiendra dans la suite, sont (i) $V \in C^1$ et (ii) ∇V globalement Lipschitz, c'est à dire

$$\exists C > 0, \forall (q_1, q_2) \in \mathbb{R}^{3N}\times\mathbb{R}^{3N},\quad |\nabla V(q_1) - \nabla V(q_2)| \le C|q_1 - q_2|. \tag{1.26}$$

Afin d'assurer l'ergodicité de la dynamique, on suppose également

$$Z_q = \int_{\mathbb{R}^{3N}} e^{-\beta V(q)}dq < \infty. \tag{1.27}$$

Remarquons que dans le cas de systèmes périodiques (où l'espace de configuration est $\mathbb{T}^{3N}$, par exemple), l'hypothèse (1.27) n'est pas nécessaire.

En pratique, une discrétisation de la dynamique (1.25) est donné par le schéma de Brünger-Brooks-Karplus (BBK):

$$\begin{cases} P_{t+\Delta t/2} &= P_t - \nabla V(Q_t)\dfrac{\Delta t}{2} - \gamma M^{-1}P_t\dfrac{\Delta t}{2} + \sqrt{\dfrac{\gamma\beta^{-1}\Delta t}{2}}\ G^{3N}, \\ Q_{t+\Delta t} &= Q_t + \Delta t M^{-1}P_{t+\Delta t/2}, \\ P_{t+\Delta t} &= P_{t+\Delta t/2} - \nabla V(Q_{t+\Delta t})\dfrac{\Delta t}{2} - \gamma M^{-1}P_{t+\Delta t}\dfrac{\Delta t}{2} + \sqrt{\dfrac{\gamma\beta^{-1}\Delta t}{2}}\ G^{3N}, \end{cases} \tag{1.28}$$

où les G^{3N} sont des vecteurs gaussiens centrés réduits de dimension $3N$. Notons que la troisième équation peut être réécrite sous une forme explicite:

$$P_{t+\Delta t} = \frac{2M}{2M+\gamma\Delta t}\left[P_{t+\Delta t/2} - \nabla V(Q_{t+\Delta t})\frac{\Delta t}{2} + \sqrt{\frac{\gamma\beta^{-1}\Delta t}{2}}\, G^{3N}\right].$$

Notons que si $\gamma = 0$, on retrouve le schéma Störmer-Verlet (1.24). L'erreur globale de ce schéma est usuellement de l'ordre de $O(\Delta t)$. On peut en effet montrer que

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=0}^{N-1} A(Q_{t+n\Delta t}, P_{t+n\Delta t}) = \int_\Omega A(p,q)\; d\mu_c(q,p) + O(\Delta t).$$

Bien que cette approche ne conduise pas à un échantillonnage exact de la mesure canonique, contrairement à l'algorithme de Metropolis-Hastings, l'erreur associée à la discrétisation de l'EDS reste en général petite devant la moyenne d'ensemble calculée $\int A\; d\mu_c$. De plus, puisque la méthode de Metropolis-Hastings risque de rejeter une proportion importante des configurations proposées, la dynamique de Langevin n'est pas sans avantage.

Il est important de noter que la dynamique de Langevin est surtout utilisée pour décrire précisément la dynamique d'un système physique. Lorsqu'il est question d'échantillonnage, on se contente de faire appel à des méthodes purement probabilistes pour les impulsions et de décrire une dynamique simplifiée pour les configurations.

Dynamique de Langevin amortie

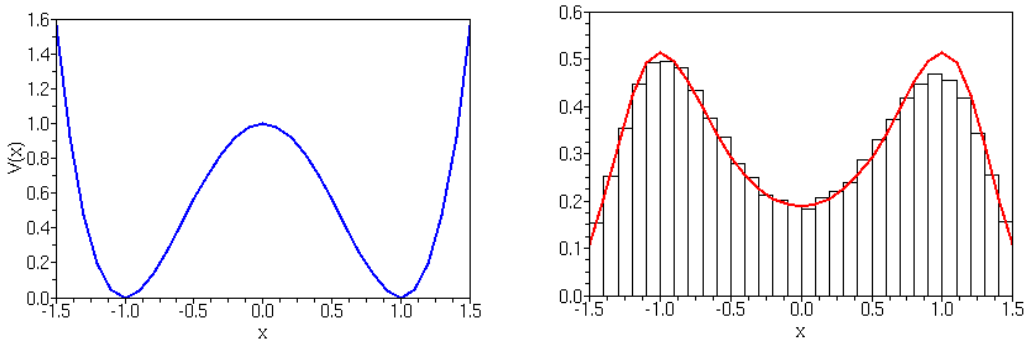
Pour échantillonner la mesure canonique les impulsions p peuvent être tirées selon la loi normale $d\kappa$ (1.8) et les positions q selon la loi $d\pi$ (1.9). La dynamique de Langevin *amortie* propose une dynamique sur q, ergodique par rapport à $d\pi$. Elle provient de la dynamique de Langevin (1.25) dans les limites $M \to 0$ et $\gamma = 1$ et est décrite par une équation différentielle stochastique faisant donc uniquement intervenir les positions Q_t:

$$dQ_t = -\nabla V(Q_t)\; dt + \sqrt{2\beta^{-1}} dW_t. \tag{1.29}$$

L'existence et l'unicité de la solution de cette EDS sont encore une fois assurées sous l'hypothèse (1.26) sur le gradient ∇V. Cette dynamique peut être discrétisée par le schéma d'Euler, ce qui donne la chaîne de Markov

$$Q_{t+\Delta t} = Q_t - \nabla V(Q_t)\; \Delta t + \sqrt{2\beta^{-1}\Delta t}\; G^{3N}.$$

Cette dynamique est simulée pour un potentiel de dimension 1, voir Figure 1.3. La Figure 1.3(a) montre le potentiel double-puits $V(x) = (x^2-1)^2$ et la Figure 1.3(b) montre la distribution empirique des positions (barres verticales) après 10^5 pas de temps et la mesure canonique associée à V (courbe noire).



(a) (b)

Figure 1.3: Échantillonnage par la dynamique de Langevin amortie (1.29). (a) Potentiel double-puits $V(x) = (x^2-1)^2$. (b) La mesure canonique associée à V (courbe rouge) et distribution des positions après 10^5 pas de temps (barres verticales). Les paramètres du modèle utilisés sont $\beta = 1$ et $\Delta t = 0.01$.

Le générateur infinitésimal $\mathcal{L}$ associé à la dynamique (1.29) est donné par

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

L'évolution en temps de ψ_t, la densité de la loi de Q_t est alors donnée par l'équation de Fokker-Planck suivante:

$$\partial_t\psi_t = \nabla \cdot (\nabla V\psi_t + \beta^{-1}\nabla\psi_t). \tag{1.30}$$

On peut facilement vérifier que $Z_q^{-1}\exp(-\beta V(q))$, la densité de la mesure π (1.9), est bien une solution stationnaire:

$$\nabla V\exp(-\beta V(q))+\beta^{-1}\nabla\exp(-\beta V(q)) = \nabla V\exp(-\beta V)+\beta^{-1}(-\beta\nabla V)\exp(-\beta V) = 0.$$

Sous certaines hypothèses sur V, on peut montrer qu'en temps long, ψ_t converge au sens de l'entropie relative exponentiellement vite vers la densité de Boltzmann-Gibbs $\psi_\infty = Z^{-1}\exp(-\beta V)$ (voir la Section 4.4.2). La *vitesse* de convergence est, pourtant, souvent lente à cause des métastabilités de la dynamique associée au potentiel sous-jacent V. La vitesse de convergence est en fait liée à la constante de Sobolev logarithmique de la mesure associée (voir la Section 4.4 pour plus de précisions).

Métastabilité

On dit d'une dynamique qu'elle est métastable si le système passe un temps important dans un état (ou région) stable avant de passer rapidement dans un autre. Ces états métastables, de basse énergie, sont typiquement séparés par des barrières de haute énergie (ou par des barrières entropiques, voir la Figure 4.1 plus loin), qui ralentissent l'exploration de l'espace. Le système passe beaucoup de temps dans un

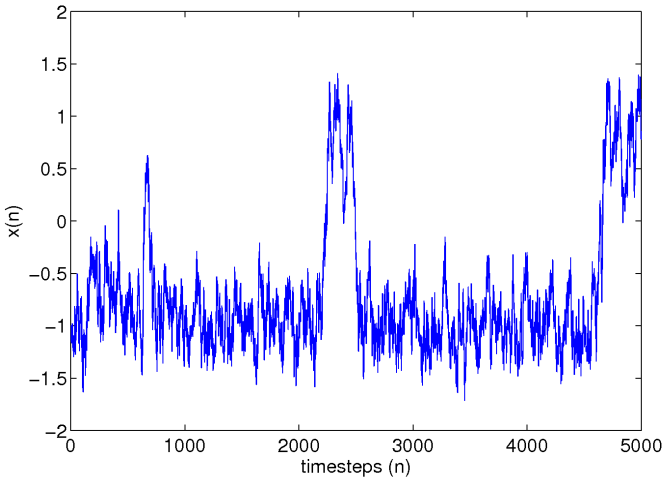


Figure 1.4: Une réalisation de la dynamique de Langevin amortie (1.29) pour le potentiel $V(x) = (x^2 - 1)^2$, présentée dans la Figure 1.3(a). Le caractère bimodal de ce potentiel nuit à l'exploration de l'espace et donc à l'échantillonnage de la mesure $d\pi = \exp(-\beta V)\, dq$.

puits, avant de franchir une barrière et de se diriger rapidement vers un autre puits. On a montré dans la Figure 1.3(a) un exemple d'un tel potentiel en dimension 1; la Figure 1.4 représente l'évolution en temps de la coordonnée x, le degré de liberté lent du système. Donnons maintenant une définition plus formelle de la métastabilité. Considérons deux régions d'intersection vide, $A \subset \mathbb{R}^{3N}$ et $B \subset \mathbb{R}^{3N}$, de basse énergie (forte probabilité) séparées par des régions de haute énergie (faible probabilité). Les régions constituent des états dits métastables du système si l'ensemble $A \cup B$ est de volume fini, mais représente les régions où le système passe la majorité de son temps:

$$Z^{-1} \int_{A \cup B} e^{-\beta V(q)}\, dq \approx 1.$$

Cette situation est malheureusement répandue pour les systèmes moléculaires complexes (sans compter que le nombre d'états métastables est typiquement bien plus grand que deux et augmente rapidement avec la taille du système). Pour cette raison, les dynamiques de Langevin ou de Langevin amorties sont le plus souvent couplées à d'autres outils permettant d'accélérer l'exploration de l'espace.

1.3 Simuler les systèmes sur des temps longs

Comme il a été précisé précédemment, la véritable difficulté lorsque l'on échantillonne la mesure canonique $d\pi$ est la convergence lente des moyennes trajectorielles, due à la multimodalité de $d\pi$. Ceci pose généralement problème lorsque l'on veut décrire le comportement d'un système en temps long.

La loi d'Arrhenius [Arrhenius 1889], basée surtout sur des observations empiriques (mais justifié mathématiquement dans certains cas par la théorie des grands déviations [Freidlin 1998]), nous dit qu'en moyenne le temps τ nécessaire pour franchir une barrière d'énergie de hauteur ΔV varie selon

$$\tau \propto \exp(\beta \Delta V). \tag{1.31}$$

Pour cette raison, la dynamique moléculaire usuelle est en pratique lente et est rarement utilisée pour décrire le comportement en temps long de systèmes complexes. Nous présentons maintenant des méthodes, aussi bien déterministes que stochastiques, qui ont pour but d'identifier les états métastables et les chemins de transition correspondants afin d'essayer d'accélérer l'exploration de l'espace.

1.3.1 Recherche de chemins de transition

Afin de décrire des dynamiques sur des échelles de temps plus longs, il faut d'abord pouvoir localiser les états métastables (ou minima locaux) du système et décrire le mécanisme par lequel le système subit les transitions d'un état à l'autre. Nous donnons ci-dessous un bref résumé des méthodes qui visent à localiser ces états métastables et à évaluer les taux de transition. Elles sont classées dans trois catégories.

Localisation des puits de potentiel et les points cols

Plusieurs méthodes existent qui ont pour objectif d'explorer les puits et cols du potentiel. Leur principe commun est le suivant: partant d'un minimum local donné, on force le système à sortir du bassin correspondant en le dirigeant vers un point col adjacent en suivant la direction de la courbure la plus faible de la fonction de potentiel. Ensuite, le système est poussé dans la direction de courbure négative afin qu'il franchise le point col et tombe dans un nouveau minimum local. Les algorithmes de cette catégorie, tels que la *Eigenvector-following method* [Cerjan 1981], *Dimer method* [Henkelman 1999, Heyden 2005] et *Activation Relaxation Technique* (ART) [Barkema 1996a, Mousseau 1998, Mousseau 2000b, Malek 2000], seront présentés dans la Section 2.3. Un des travaux effectués au cours de la thèse propose une amélioration pour une version d'ART adaptée au cas de la température nulle, appelée ART *nouveau*. Ce travail sera présenté dans le chapitre 3, où nous donnons également une preuve de convergence pour un algorithme prototype du modèle amélioré.

Chemins de réaction

Une fois les minima locaux identifiés, des méthodes telles que la méthode de *Nudged Elastic Band* (NEB) [H. Jonsson 1998] et la *zero temperature String method* [E 2002] permettent de tracer un chemin de transition d'énergie minimale entre deux états stables (minima locaux) en passant par le col. Dans le cas d'une surface d'énergie potentielle rugueuse (voir la Figure 1.5, à droite), les méthodes classiques ne suffisent pas et on fait appel à des méthodes telles que la *finite temperature String method*

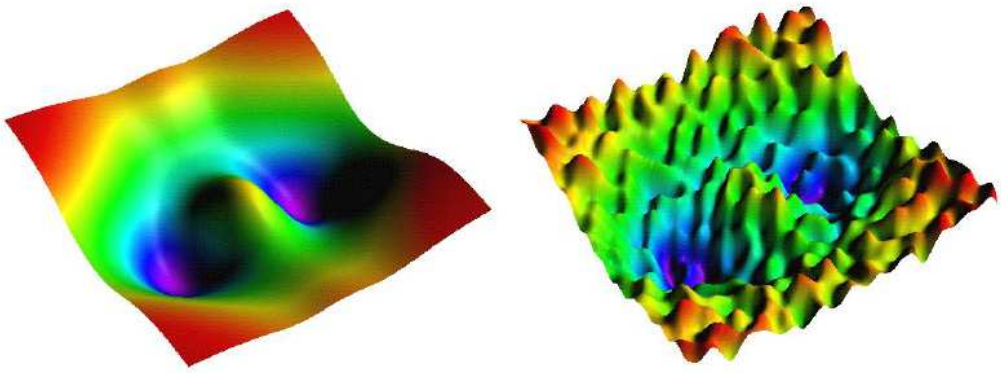


Figure 1.5: *A gauche*: Un exemple d'une surface d'énergie potentielle lisse avec deux minima locaux. Cette situation se rencontre parfois en sciences des matériaux (sauts de lacunes ou d'atomes interstitiels dans un cristal) ou en chimie (par exemple la di-alanine). *A droite*: Une surface d'énergie potentielle rugueuse, présentant de nombreux minima locaux, qui possède néanmoins deux minima locaux importants, correspondants aux énergies les plus basses. En parcourant le chemin d'un puits à l'autre, le système va tomber dans de petites régions métastables, ce qui ralentit la trajectoire. Cette surface est typique de systèmes complexes tels que les biomolécules de grande dimension.

(FTS) [E 2005a]. On peut voir cette dernière comme un couplage entre la *String method* et une méthode d'échantillonnage.

Taux de transition

Il existe des méthodes pour déterminer le taux de passage d'un puits à l'autre, une fois ces derniers identifiés. L'une des méthodes les plus célèbres s'appelle *Transition State Theory* (TST), développée simultanément par Eyring, et par Evans et Polanyi en 1935 [Eyring 1935, Evans 1935]. Une fois que les taux de transition sont connus, ils sont utilisés pour simuler l'évolution du système à des échelles de temps plus longs par les méthodes telles que la Monte-Carlo cinétique, décrite ci-dessous.

1.3.2 Simuler sur les temps plus longs

Monte-Carlo cinétique

Les taux de passage d'un puits à un autre étant supposés connus, les méthodes de Monte-Carlo cinétique (*Kinetic Monte-Carlo*, KMC) [Young 1966, Bortz 1975] permettent de décrire l'évolution du système à des échelles de temps souvent inaccessibles via des méthodes de dynamique moléculaire habituelle. Cette classe de méthodes requiert (et souvent maintient) une liste d'événements et de taux de réaction (calculés par les méthodes de la Section 1.3.1). On substitue alors à la

dynamique initiale du système, une dynamique simplifiée ne faisant intervenir que les états métastables et les probabilités de transition précalculées.

Dynamique biaisée par l'énergie libre

Une fois que l'on peut paramétrer une transition par une *coordonnée de réaction*, on peut utiliser l'énergie libre associée pour forcer le système de passer d'un état métastable à l'autre. Les méthodes de cette catégorie seront présentées dans le chapitre 4. Déterminer une bonne coordonnée de réaction est un problème d'intérêt en soi et ce sera le sujet de la Section 4.1.2.

Ce rapport de thèse se divise en deux parties. Nous présentons, dans la première partie, les méthodes de recherche de chemins de transition et les points col associés: dans le chapitre 2, nous présentons plus en détail les méthodes de recherche de chemins de transition évoquées à la Section 1.3.1; dans le chapitre 3, nous proposons une amélioration et une preuve de convergence pour la méthode *Activation-Relaxation Technique* de température nulle (ARTn). La deuxième partie est consacrée au calcul d'énergie libre et aux dynamiques biaisées: nous donnons d'abord, dans le chapitre 4, un récapitulatif des méthodes principales de cette classe; nous présentons ensuite, dans les chapitres 5 et 6, les nouveaux résultats, numériques et théoriques, obtenus au cours de cette thèse.

Part I

Finding transition paths

Chapter 2

Finding transition paths

Sommaire

2.1 Simulations on large time scales 23
2.2 Double-ended methods 24
2.2.1 Nudged Elastic Band method 25
2.2.2 The String method 28
2.2.3 Transitions at finite temperature 29
2.3 Single-ended methods 30
2.3.1 Locating transition states in nonconvex regions 31
2.3.2 Climbing out of convex regions 37

2.1 Simulations on large time scales

This chapter is dedicated to a review of computational methods commonly used in theoretical chemistry and solid state physics to find transition states on a potential energy surface and more precisely minimal energy paths. As previously discussed, classical molecular dynamics (MD) rarely allow us to capture the important transitions of molecular systems due to the limited feasible time-scales on even the fastest modern-day computers. Due to the smallness of the characteristic time scales of atomic vibrations, timesteps typically used in MD simulations are on the order of a femtosecond (fs). Important transitions (chemical reactions or conformational changes) in molecular systems, however, occur on the scale of a microsecond (μs). The advances in high-performance computing and parallel architectures have made it possible to approach these large time scales with classical MD methods. This past decade has seen the development of special purpose supercomputers such as Anton [David E. Shaw 1998, Larson 2009, Shaw 2008], whose highly-parallel architecture has been specially designed to carry out force evaluations at extremely high speeds, with the help of a Gaussian split Ewald method [Shan 2005]. Other groups have been devoted to a distributed computing project *Folding@home* [Shirts 2001, Shirts 2000] whereby common household computers around the world can donate their idle time to protein folding simulations. Efficient MD software is of course also at the heart of modern-day research. The widely-used MD code NAMD [Bhandarkar 2003, Kale 1999, Phillips 2005] and the

more recent Desmond [Bowers 2006] have been developed to run efficiently on parallel architectures. Furthermore, more recently, a platform of choice for a growing number of MD applications has been the powerful and widely available graphical processing unit (GPU) [Stone 2007, Rodrigues 2008].

Such technological advances have allowed for longer simulations using classical MD techniques. However every small perturbation or displacement of a system does not always need to be explicitly described to understand its long-time behavior. The past two decades have given rise to a multitude of computational methods aimed at pushing the system to undergo important transitions typically observed at very large time scales, overlooking the small thermal vibrations. These methods are for the most part deterministic, relying heavily on the gradient and curvature of the underlying potential energy surface, but stochastic elements are not entirely absent. This will become apparent at a later stage (see for instance Sections 2.3.1.3 and 2.2.3). These methods fall into two categories: (i) initial and final states are known and the minimal energy path between the two states is sought; (ii) a single local minimum is known a priori and one seeks adjacent local minima and corresponding transition states.

Section 2.2 will give an overview of the methods falling into category (i) and Section 2.3 will outline some methods of category (ii), for which improvements and convergence results have been proposed as a part of this work (see Chapter 3).

2.2 Double-ended methods

This section is devoted to a review of algorithms developed in the past two decades for finding minimal energy paths in the particular case where initial and final states (each a local minimum), denoted A and B respectively, are known. The minimal energy path is of particular interest when studying the dynamics of physical systems as it is known to be the most probable path taken by the system to undergo the reaction from A to B [Freidlin 1998].

A formal definition will help in motivating the algorithms discussed herein. A curve φ connecting A and B is said to be a minimal energy path if the orthogonal component of the force to the curve is zero at every point along the curve and that no critical point along the path has a Morse index larger than or equal to two (see discussion further on). More formally, if a minimal energy path is parameterized by $\alpha \in [0,1]$, where $\varphi(0) = A$ and $\varphi(1) = B$, then one has

$$(\nabla V(\varphi(\alpha)))_\perp = 0, \quad \forall \alpha \in [0,1]. \tag{2.1}$$

where $(\nabla V(\varphi(\alpha)))_\perp$ is the component of the gradient of the potential orthogonal to the curve at point $\varphi(\alpha)$. Let us denote in the following

$$(\nabla V(\varphi(\alpha)))_\parallel = \frac{\nabla V(\varphi(\alpha)) \cdot \nabla\varphi(\alpha)}{|\nabla\varphi(\alpha)|^2} \nabla\varphi(\alpha),$$

and

$$(\nabla V(\varphi(\alpha)))_\perp = \nabla V(\varphi(\alpha)) - (\nabla V(\varphi(\alpha)))_\parallel,$$

the components of the gradient parallel and orthogonal to the curve.

Once the minimal energy path is determined, the point along it with the highest energy is called the transition state. A transition state $q_{\rm ts}$ is a saddle point satisfying the following criteria: (i) $\nabla V(q_{\rm ts}) = 0$; and by Murrell and Laidler [Murrell 1968] (ii) $H(q_{\rm ts}) = \nabla^2 V(q_{\rm ts})$, the Hessian, has a single negative eigenvalue. The second criterion signifies that at the transition state, the energy is a local maximum in one direction and a local minimum in the orthogonal hyperspace. The eigenvector associated to the single negative eigenvalue at $q_{\rm ts}$ is parallel to the minimal energy path. Criterion (ii) is justified as follows. Suppose that the Hessian at a saddle point $q^*_{\rm ts}$ has two negative eigenvalues, $-\lambda_1 < 0$ and $-\lambda_2 < 0$, associated to the normal eigenvectors v_1 and v_2. By taking small steps in the direction of these eigenmodes, we find

$$\begin{aligned} V(q^*_{\rm ts} + \varepsilon_1 v_1 + \varepsilon_2 v_2) &= V^* + (\varepsilon_1 v_1 + \varepsilon_2 v_2)^T \nabla V^* + (\varepsilon_1 v_1 + \varepsilon_2 v_2)^T H^* (\varepsilon_1 v_1 + \varepsilon_2 v_2) \\ &= V^* - \lambda_1 \varepsilon_1^2 v_1^2 - \lambda_2 \varepsilon_2^2 v_2^2 \\ &< V^*, \end{aligned}$$

where $V^* = V(q^*_{\rm ts})$, $\nabla V^* = \nabla V(q^*_{\rm ts}) = 0$ and $H^* = H(q^*_{\rm ts}) = \nabla^2 V(q^*_{\rm ts})$. In other words, in the region of $q^*_{\rm ts}$, the hypersurface in the two eigenmodes would a hill, thus a lower-energy path may be constructed by walking around it. The Hessian at a transition state, therefore, must always be of index one; a generic saddle point differs in that it typically has one *or more* negative eigenvalues, and at least one positive eigenvalue.

With these notions in mind, we are now ready to present two of the leading double-ended methods used to estimate paths of minimal energy: the Nudged Elastic Band method (Section 2.2.1) and the String Method (Section 2.2.2). The idea, common to both of these algorithms, is to begin with an initial discretized path between the initial and final states and to evolve the path in such a way as to estimate the minimal energy path. It should be emphasized that the methods discussed below provide a means to find *one* minimal energy path, determined almost entirely by the initial path. In the case where more than one reaction path exists (see for example Figure 4.2(b) in Chapter 4), the others may be obtained for example by applying the algorithms to a number of different initial paths.

2.2.1 Nudged Elastic Band method

The Nudged Elastic Band (NEB) method [H. Jonsson 1998] was one of the earliest proposed to estimate transition pathways given initial and final states, A and B. The idea of the NEB method is to evolve a discretized path consisting of a set of $N+1$ *images*, $A = x_0, x_1, ..., x_N = B$, in such a way that the final discretized path approximates the minimal energy path. The end points x_0 and x_N typically remain fixed throughout the algorithm. To obtain the orthogonality condition (2.1), each image is pushed (or *nudged*) in the steepest descent direction orthogonal to the curve. Depending on the choice of the initial path, the images would tend to form clusters in

the basins of attraction, leaving the region about the transition state undersampled. In order to tackle this potential clustering of images, the NEB method introduces a spring-like (or *elastic*) force, to encourage equal spacing. The NEB method owes its name to these key features of the algorithm.

The evolution of each image i is described by

$$\dot{x}_i = F_i = -\nabla V(x_i)_\perp + F_i^s|_\parallel, \tag{2.2}$$

where $\dot{x}_i$ denotes the derivative in time of the evolving image $x_i(t)$, $-\nabla V(x_i)_\perp$ denotes the projection of the force onto the hypersurface perpendicular to the (normal) tangent τ_i of the curve at the point x_i (the precise definition of τ_i will be given below). The term $F_i^s|_\parallel$ represents the parallel component of the force to the curve exerted on the images. The perpendicular projection of the force is defined as follows

$$-\nabla V(x_i)_\perp = -\nabla V(x_i) + (\nabla V(x_i) \cdot \tau_i)\tau_i$$

and the spring-like force is evaluated as [H. Jonsson 1998]

$$F_i^s|_\parallel = k(|x_{i+1} - x_i| - |x_i - x_{i-1}|)\tau_i, \tag{2.3}$$

or, as suggested in [E 2007],

$$F_i^s|_\parallel = k[(x_{i+1} - 2x_i + x_{i-1}) \cdot \tau_i]\tau_i, \tag{2.4}$$

where $k > 0$ is the spring constant. These terms favor equal spacing between image i and its adjacent counterparts.

Remark 1 *Note that in the continuous limit ($N \to \infty$), the evolution of the strings (2.2) could be interpreted as*

$$\dot{\varphi}(\alpha) = -\nabla V(\varphi(\alpha))_\perp + k\varphi_{\alpha\alpha}(\alpha),$$

where φ is the continuous string characterized by parameter $\alpha \in [0, 1]$ and $\varphi_{\alpha\alpha}$ denotes the double derivative of the curve with respect to parameter α.

The tangent τ_i is often calculated as the average of the local first order approximations:

$$\tau_i = \frac{x_{i+1} - x_i}{|x_{i+1} - x_i|} + \frac{x_i - x_{i-1}}{|x_i - x_{i-1}|}. \tag{2.5}$$

This choice, however, has been linked to kinks appearing in the path and is the subject of discussion in a paper by Henkelman *et al.* [Henkelman 2000a]. In this paper, an alternative choice for the tangent was suggested in order to obtain a smoother path:

$$\tau_i = \begin{cases} \tau_i^+ & \text{if } V(x_{i+1}) > V(x_i) > V(x_{i-1}) \\ \tau_i^- & \text{if } V(x_{i+1}) < V(x_i) < V(x_{i-1}) \end{cases}$$

where

$$\tau_i^+ = \frac{x_{i+1} - x_i}{|x_{i+1} - x_i|} \quad \text{and} \quad \tau_i^- = \frac{x_i - x_{i-1}}{|x_i - x_{i-1}|}.$$

If an image is at a local minimum or maximum, a potential weighted finite difference expression is used to avoid abrupt changes in the curve. More precisely, if $V(x_{i+1}) > V(x_i) < V(x_{i-1})$ or $V(x_{i+1}) < V(x_i) > V(x_{i-1})$, the tangent would be estimated as follows

$$\tau_i = \begin{cases} \tau_i^+ \Delta V_i^{\max} + \tau_i^- & \text{if } V(x_{i+1}) > V(x_{i-1}) \\ \tau_i^- \Delta V_i^{\min} + \tau_i^+ & \text{if } V(x_{i+1}) < V(x_{i-1}) \end{cases}$$

where

$$\Delta V_i^{\max} = \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|) \text{ and } \Delta V_i^{\min} = \max(|V_{i+1} - V_i|, |V_{i-1} - V_i|).$$

In the following examples, this new implementation of the tangent has been used.

The algorithm is terminated when the maximum displacement of the string or maximum of the projected forces on each image is below a certain tolerance TOL. The algorithm is more formally given as follows

Algorithme 2 Nudged Elastic Band

N, x_0 and x_N, $n_{\max}$ given.
Set $\Delta x = (x_N - x_0)/N$.
Set $x_i = x_0 + i\Delta x$, for $i = 1, ..., N-1$.
Set $n = 0$.
while $((n < n_{\max})$ **and** $(\max_i \|\nabla V(x_i)|_\perp\| > \text{TOL}))$

for $i = 1, .., N-1$

Calculate τ_i
$x_i^{n+1} = x_i^n + \Delta t(-\nabla V(x_i)|_\perp + k(|x_{i+1} - x_i| - |x_i - x_{i-1}|)\tau_i)$,

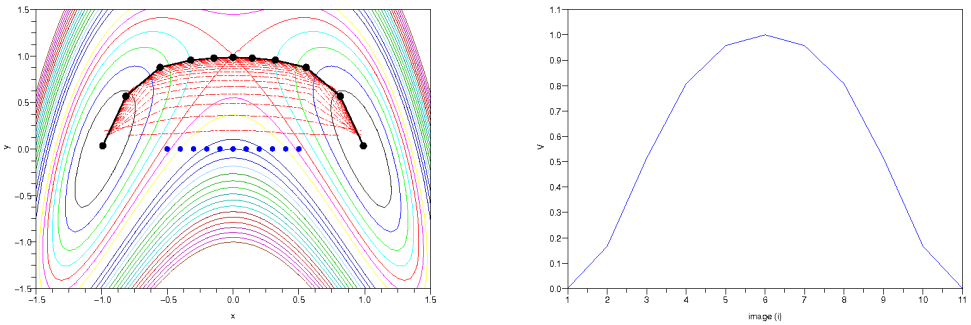
end
$n = n + 1$,

end

Although the original NEB method keeps the images on the extreme ends of the elastic band fixed at the local minima, the method can be extended to the more general case, where the end-most images are placed on either side of the dividing surface (a surface from which a system is equally likely to reach state A or B). In this case, the end-most images x_0 and x_N take steps in the direction of steepest descent and the intermediary images follow the standard algorithm described above. This method is demonstrated with 11 images in Figure 2.1(a), for the smooth double-well potential

$$V(x, y) = (x^2 - 1)^2 + (y + x^2 - 1)^2. \tag{2.6}$$

Model parameters are given in the figure caption. Once the path has converged, the saddle point may be identified as the point with the highest energy along the minimal energy path. This can be done by tracing the energy as a function of the images along the curve, as shown in Figure 2.1(b).



(a) (b)

Figure 2.1: The NEB method applied to the smooth double-well potential V defined in (2.6). The parameters used are as follows: $N = 10$, $\Delta t = 0.1$, $k = 1$, $n_{\max} = 100$, TOL$= 10^{-2}$. (a) Starting from a straight-line path from either side of the dividing surface, the images on the extreme ends of the band follow a steepest descent direction while intermediary images follow the standard NEB algorithm described in Algorithm 2.2.1. The termination criterion is satisfied after 35 steps. (b) The energy along the minimal energy path suggests that image $i = 6$ (the point of highest energy) is a good approximation of the transition state.

2.2.2 The String method

The zero-temperature String method [E 2002, E 2003, E 2007] is in essence very similar to the NEB method: given initial and final states, a path or *string* (discretized by a set of $N + 1$ images) evolves on a smooth potential according to certain rules towards the minimal energy path. The String method distinguishes itself from NEB in the second term on the right hand side of (2.2). The NEB method enforces equal spacing between images in a flexible manner, through a spring-like force, whereas the String method imposes a more rigid constraint by a reparameterization of the string. Assuming one begins with the string $\varphi = \{\varphi(\alpha), \alpha \in [0, 1]\}$, then the normal tangent to the curve is described by $\tau(\alpha) = \varphi_\alpha/|\varphi_\alpha|$, where $\varphi_\alpha = \partial\varphi/\partial\alpha$. The curve φ in the string method evolves according to the following model:

$$\frac{\partial \varphi}{\partial t} = -\nabla V(\varphi)_\perp + \lambda\tau. \tag{2.7}$$

where $\lambda = \lambda(\alpha, t)$ is the Lagrange multiplier, enforcing the reparameterization of the string $|\varphi_\alpha| = 1$. The evolution of the images discretized in time, with time step Δt gives:

$$\varphi_i(t + \Delta t) = \varphi_i(t) - \nabla V(\varphi(t))_\perp \Delta t. \tag{2.8}$$

Once the images φ_i of the string have taken a steepest descent step, the partial string segment between φ_1 and φ_n has length

$$S_n = \sum_{i=1}^{n} |\varphi_i - \varphi_{i-1}|,$$

and the normalized mesh is obtained by setting $\alpha_i = S_i/N$. Then, a cubic spline interpolation is carried out on $\{(\alpha_i, \varphi_i), i = 1, ..., N\}$ and the new images φ_i^* are chosen to be at equidistant points along the string:

$$\alpha_i^* = i/N, \quad \text{for } i = 0, ..., N.$$

2.2.3 Transitions at finite temperature

When studying transitions at finite (high) temperature, the methods discussed so far no longer suffice. Simulations at finite temperature may be of particular interest when studying transitions on rough energy landscapes or to take into account entropic barriers on the potential energy surface. First, when the underlying potential is rugged, *i.e.* there exist a plethora of local minima and saddle points, as depicted earlier in Figure 1.5, the notion of a minimal energy path (thus the NEB and String methods) becomes irrelevant. The shallow basins on a rugged surface typically have energy barriers in the order of 1 eV, which are easily surmountable by systems at finite temperature and are therefore almost always unseen in the main transition path from A to B. Second, entropic effects may result in a higher energy path being favored over another: a high energy barrier around which the energy is flat may lead to a lower free energy (see Chapter 4) than a narrow lower energy barrier, and thus is favored over the latter.

Figure 2.2(a) shows an example of a rough energy landscape, where a standard NEB algorithm fails to find the most probable path between the two global minima. The NEB and string methods fail in such situations due to the simple projected force term $-(\nabla V)_\perp$ driving the evolution of the images (cf. (2.2) and (2.7)). As shown in Figure 2.2(a), such algorithms drive the images into local basins, which, unless pulled out by spring-like forces along the curve or reparameterization of the string, remain trapped. Figure 2.2(b) shows the energy along the images of the falsely converged path.

To tackle such situations, one may consider effective transition pathways, introduced in [E 2005b]. The idea is to consider isoprobability surfaces, *i.e.* surfaces from which the probability of a trajectory reaching the final state B before the initial state A is uniform. By weighting these surfaces with the equilibrium Boltzmann-Gibbs measure, one can determine the regions of the surface where reactive trajectories are most likely to pass through. In practice this may be done using the Finite Temperature String method, described below.

Finite Temperature String method

The Finite Temperature String method [E 2005a] combines the string method and equilibrium sampling techniques to tackle rough energy landscapes and account for entropic effects. The idea of the method is to replace the gradient term in (2.7) by a term involving an averaged quantity, defined in (2.9) below. At each time step t of the string method and for each image i of the string, one defines a Voronoi cell $B_i(t)$, defined by

$$B_i(t) = \{x \ : \ |x - \varphi_i(t)| < |x - \varphi_j(t)|, \ \forall i \neq j\},$$

and calculates in each cell the conditional average

$$\langle x \rangle_{B_i(t)} = \frac{\displaystyle\int_{B_i(t)} q \ e^{-\beta V(q)} \ dq}{\displaystyle\int_{B_i(t)} e^{-\beta V(q)} \ dq}. \tag{2.9}$$

Equation (2.8) is then replaced by

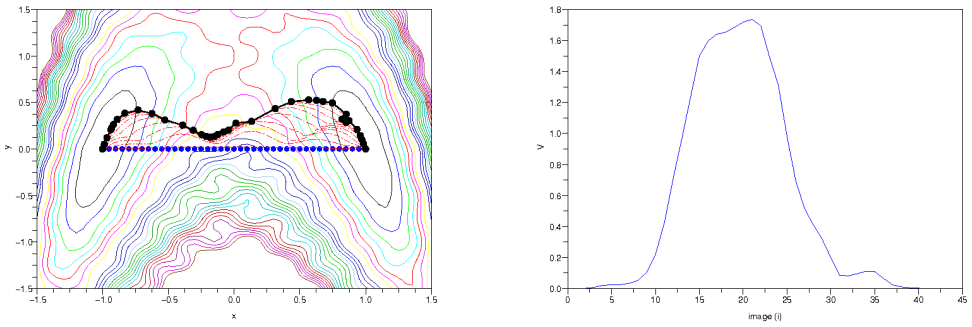
$$\varphi_i(t + \Delta t) = \varphi_i(t) - (\varphi_i(t) - \langle x \rangle_{B_i(t)}) \ \Delta t.$$

In this way, the images of the string are pulled towards the most probable position of the system within the Voronoi cell and are less likely to be trapped in shallow local minima regions. In practice, (2.9) is estimated by simulating Langevin dynamics (with timestep typically smaller than Δt) in each Voronoi cell, and imposing appropriate boundary conditions [Vanden-Eijnden 2009a, Vanden-Eijnden 2009b].

2.3 Single-ended methods

This section will give an overview of algorithms designed to search transition states around a local minimum, when no information is known about the final state. The NEB and string methods provide means for recovering the most probable path of a system for a *particular* reaction, however they may not be applied in the case where the final product state is unknown. Methods such as kinetic Monte Carlo (KMC) [Young 1966, Bortz 1975] require a list of local minima, transition states and transition rates.

Furthermore, although double-ended methods may also be used to locate transition states along a minimal energy path, they are not the most efficient for this task. A climbing-image NEB method was proposed [Henkelman 2000b] as an alternative to the classical NEB for such cases, whereby the highest image does not feel the spring forces and is driven instead towards the saddle point. The string method also allows for refinement of its mesh, but it would not be without the cost of redundant work. In such double-ended methods, each time step requires force evaluations for *every* image along the path, even the low-lying images, far away from the transition state, meaning some computational effort is wasted (supposing



(a) (b)

Figure 2.2: An example of a rugged potential energy surface, for which the NEB method fails. The ruggedness is introduced by perturbing the potential V of Figure 2.1 by $V_\varepsilon(x, y) = \varepsilon \sin(4\pi x) \cos(4\pi y)$, where in this case $\varepsilon = 0.1$. The parameters used in this example are as follows: $N = 50$, $\Delta t = 0.01$, $k = 1$. The algorithm falsely converges in about 10 steps.

we are interested *only* in the transition state). Finally, and most importantly, these methods cannot be applied in the case of an unknown final state or indeed dividing surface.

For these reasons, algorithms such as the Eigenvector following method [Cerjan 1981], the Dimer method [Henkelman 1999] and the Activation Relaxation Technique (ART) [Mousseau 2000b, Malek 2000] have been developed with the goal of locating transition states and exploring the local basins in the neighborhood. Once a transition state is located, a straightforward minimization finds an adjacent local minimum. Note that the search for a transition state is more challenging than for a local minimum or maximum. As a transition state is, by definition, a local maximum in one direction and a local minimum in all others, additional information is required on the curvature of the potential and eigenmodes of the Hessian. It should be emphasized that the curvature information will help to locate the transition state, once the system is *already* within its local vicinity. A pertinent question is then: how is a system driven out of a harmonic basin efficiently? This will be discussed in Section 2.3.2. Before, we begin by outlining a few popular single-ended methods used in nonconvex regions of a potential energy surface.

2.3.1 Locating transition states in nonconvex regions

In order to accurately locate a transition state, one needs information about the Hessian in the nonconvex region around it. Attaining such a region is by no means trivial, and its discussion is postponed to Section 2.3.2. In the following, we place ourselves in the nonconvex region local to the transition state. We begin by briefly introducing the prevailing ideas behind the methods discussed in this section, namely

that of climbing the potential surface in one direction while minimizing in all others.

For a point $q \in \mathbb{R}^n$ in configuration space, consider the Taylor expansion of the potential energy $V = V(q)$ about a point q_0, denoting the difference between the two $h = q - q_0$:

$$V(q) = V_0 + \nabla V_0^T h + \frac{1}{2} h^T H_0 h, \tag{2.10}$$

where $V_0 = V(q_0)$, $\nabla V_0 = \nabla V(q_0)$ and $H_0 = H(q_0) = \nabla^2 V(q_0)$. By solving $\nabla_h V = 0$ for h and expressing it in the basis of the orthonormal eigenvectors $(u_i)_{1 \leq i \leq n}$ of H_0, we have

$$h = -\sum_{i=1}^{n} \frac{\langle \nabla V_0, u_i \rangle u_i}{\lambda_i}, \tag{2.11}$$

where $(\lambda_i)_{1 \leq i \leq n}$ are the corresponding eigenvalues of H_0. The change in energy induced by this step is then

$$V(q) - V(q_0) = \Delta V = -\sum_{i=1}^{n} \frac{\langle \nabla V_0, u_i \rangle^2}{2\lambda_i}.$$

Therefore $\lambda_i < 0$ results in an energy increase in the direction u_i and $\lambda_i > 0$ an energy decrease.

In order to approach the transition state, the system should be pushed in the direction of the lowest eigenmode (the eigenvector corresponding to the single negative eigenvalue) and be minimized in all orthogonal directions. Recall that we are assuming there exists only one negative eigenvalue in the sufficiently close vicinity of the transition state. The use of the Hessian in such algorithms does not come without computational burden:

- For large systems, typically consisting of $N > 1000$ atoms, calculating the Hessian is very costly: in total $O(N^2)$ operations are needed. Alternatives are discussed hereafter to estimate the Hessian (in fact a Hessian-vector product) using finite difference approximations, see for example Section 3.2.
- Supposing a Hessian matrix has been obtained for a large system. Determining the full spectrum of the matrix through diagonalization causes further computational burden: a complexity of $O(N^3)$. When a full spectrum is not required, its extreme values (smallest or largest eigenvalue in absolute value) may be extracted more efficiently via methods such as the Power Iteration, Lanczos method or the minimization of the Rayleigh-Ritz quotient.

In order to bypass these potential bottlenecks, the Dimer method [Henkelman 1999], briefly discussed in Section 2.3.1.2, proposes an algorithm depending only on first derivatives of the potential energy function, making it applicable to larger systems.

These single-ended methods may in fact be coupled with schemes such as the KMC, where saddle points and local minima are calculated on the fly. This has been applied to study the diffusion of an adatom on an Al(100) surface, where KMC has been complemented with the dimer method [Henkelman 2001].

2.3.1.1 Eigenvector-following method

The Eigenvector-following method, developed in [Cerjan 1981], was one of the first algorithms proposed to locate saddle points of index one. This method consists in pushing the system towards a saddle point by taking steps as defined in (2.11), only with a restriction in the step size: $|h| = r$ for some predefined $r > 0$. The function then considered is the Lagrangian function

$$L(h, r) = V_0 + \nabla V_0^T h + \frac{1}{2} h^T H_0 h + \frac{\eta}{2}(r^2 - |h|), \tag{2.12}$$

where η is the Langrange multiplier. The extrema are located by solving $\partial L/\partial h = 0$ and $\partial L/\partial \eta = 0$. For fixed r, the second equation can be solved straightforwardly for η, determining the step taken by the system:

$$h = -\sum_i \frac{\langle \nabla V_0, u_i \rangle u_i}{(\lambda_i - \eta)}. \tag{2.13}$$

In fact, to allow for more flexibility, a different Langrange multiplier can be used for different eigendirections [Wales 2005]. This method is indeed robust and has been shown to converge quadratically [Banerjee 1985], however its dependence on the full spectrum of the Hessian restricts its application to systems of small sizes: the diagonalization scales as the cube of the system size. In order to reduce such costs the Hybrid Eigenvector-following method [Munro 1999, Wales 2003] was proposed whereby a step is taken only in the direction of the lowest eigenmode (calculated using standard iterative methods with shifting) and minimizing in the orthogonal hyperplane.

2.3.1.2 The Dimer method

The Dimer method, developed by Henkelman and Jónsson [Henkelman 1999, Heyden 2005] proposes a simple algorithm to converge to a transition state without computing the Hessian of the potential. This method consists in pushing uphill, towards a transition state, a pair of images (the 'dimer'), through rotation and translation. The goal of the rotation is to find the lowest curvature mode of the Hessian without calculating it explicitly. The translation of the dimer is the step needed to push the pair of images to climb the potential energy surface in the direction of the lowest mode.

The midpoint of the dimer is at $\mathbf{r}$ and the unit vector $\hat{\mathbf{n}}$ defines its orientation. The images, separated from the midpoint by a distance Δr, are thus defined by

$$\mathbf{r}_1 = \mathbf{r} + \Delta r\ \hat{\mathbf{n}}, \quad \text{and} \quad \mathbf{r}_2 = \mathbf{r} - \Delta r\ \hat{\mathbf{n}}.$$

The energies and forces of the images are given by E_1 and E_2 and $\mathbf{f}_1$ and $\mathbf{f}_2$, so that the total energy of the dimer is $E = E_1 + E_2$ and the force on the midpoint $\mathbf{f}$ is the average $(\mathbf{f}_1 + \mathbf{f}_2)/2$. Finally, the energy at the midpoint of the dimer may be

computed by equating the curvature of the potential at $\mathbf{r}$ in terms of forces $\mathbf{f}_1$,$\mathbf{f}_2$ and energies E_1,E_2. A simple rearrangement yields

$$E_0 = \frac{E_1 + E_2}{2} + \frac{\Delta r}{4}(\mathbf{f}_1 - \mathbf{f}_2) \cdot \hat{\mathbf{n}}.$$

The energy and force exerted on the dimer are thus calculated entirely from their value at the two images. This is a significant advantage over the Hessian-based methods as it reduces the total number of force evaluations. As the curvature C of the potential at the midpoint can be expressed as

$$C = \frac{E_1 - 2E_0 + E_2}{(\Delta R)^2} = \frac{E - 2E_0}{(\Delta R)^2}, \tag{2.14}$$

and E_0 remains fixed during rotation, it is clear that the lowest curvature can be found by minimizing the dimer energy E. In order to do so, the dimer is rotated along the rotational force $\mathbf{f}^{\perp} = \mathbf{f}_1^{\perp} - \mathbf{f}_2^{\perp}$, where $\mathbf{f}_i^{\perp} = \mathbf{f}_i - (\mathbf{f}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. Then, by defining a unit vector $\hat{\mathbf{o}}$ perpendicular to $\hat{\mathbf{n}}$ in the plane of rotation, the new position, $\mathbf{r}_1^*$, of image 1 after rotating about the angle $d\theta$ is given by

$$\mathbf{r}_1^* = \mathbf{r} + (\hat{\mathbf{n}} \cos d\theta + \hat{\mathbf{o}} \sin d\theta)\Delta r.$$

The new position of image 2 is thus $\mathbf{r}_2^* = \mathbf{r} - \hat{\mathbf{n}}^* \Delta r$, where $\hat{\mathbf{n}}^* = (\mathbf{r}_1^* - \mathbf{r})/|\mathbf{r}_1^* - \mathbf{r}|$. The idea is then to choose an angle $\Delta\theta$ such that the scalar force $f = \mathbf{f}^{\perp} \cdot \hat{\mathbf{o}}/\Delta r$ is made zero. A simple Newton's step gives the estimate

$$\Delta\theta \approx \frac{f}{-f'},$$

where, f may be approximated by the average $f = (\mathbf{f} \cdot \hat{\mathbf{o}} + \mathbf{f}^* \cdot \hat{\mathbf{o}}^*)/2$, and f' may be approximated using finite differences.

A saddle point is the maximum along the lowest curvature mode and a minimum in all other modes; thus once a lowest curvature mode has been identified, a translation of the dimer must pull it towards a maximum in the direction of the lowest mode and a minimum in all other modes. The translational force $\mathbf{f}_{\mathrm{tr}}$ required is therefore

$$\mathbf{f}_{\mathrm{tr}} = \mathbf{f} - 2(\mathbf{f} \cdot \hat{\mathbf{n}})\ \hat{\mathbf{n}},$$

so that the force acting in the direction of the dimer orientation is inverted. This force is in fact only used in non-convex regions of the potential surface; in convex regions one uses simply $\mathbf{f}_{\mathrm{tr}} = -(\mathbf{f} \cdot \hat{\mathbf{n}})\ \hat{\mathbf{n}}$ in order to local minima regions.

2.3.1.3 Activation Relaxation Technique (ART)

We present in this section the last of the single-ended methods, the Activation Relaxation Technique (ART) [Barkema 1996a, Mousseau 1998, Mousseau 2000b, Malek 2000], for which improvements and convergence analysis constitute a part of the contributions of this work and are presented in Chapter 3. We present herein a brief background and the main notions of the algorithm. Let us begin by noting that this technique identifies *events*, consisting of two principle components

1. *Activation*: leading the system out of its attractive basin and towards a transition state,

2. *Relaxation*: pushing the system over the transition state and relaxing to an adjacent local minimum.

An event is considered successful if both the above converge successfully.

ART: early steps

The original ART algorithm [Barkema 1996a, Mousseau 1998] differs from the previously discussed single-ended methods in that no local information on the curvature of the potential energy surface is used. Furthermore, the system was treated similarly in the convex and nonconvex regions. Following a small initial displacement Δq of the system from a local minimum q_0, it is pushed to a saddle point iteratively, according to

$$q_{n+1} = q_n - h[\nabla V - (1+\alpha)(\nabla V \cdot \Delta q_n)\Delta q_n], \quad \text{for } n > 0 \tag{2.15}$$

where $h > 0$ and $\alpha > 0$ are model parameters and, in successive iterations, Δq_n is the difference between the current configuration q_n and the initial local minimum q_0. The component parallel to Δq_n has opposite sign to the force (climbing in energy) and the orthogonal component is equal to the force (relaxing in the hyperspace). The quantity Δq_n is essential in determining the success of this algorithm. If it is close to the lowest eigenmode, then one can expect convergence (albeit, with a sufficiently small timestep). The above method has been applied successfully to amorphous semiconductors [Barkema 1999] and Lennard-Jones clusters [Malek 2000] in identifying relaxed configurations. This method was later further developed into ART *nouveau* (ARTn) to incorporate local curvature information, making it suitable for transition state searches. This is discussed below.

ARTn: The zero-temperature case

The ART *nouveau* (ARTn) [Mousseau 2000a] method has proven successful in locating saddle points of index one in a wide range of energy landscapes. The ARTn algorithm distinguishes itself from (2.15) in that the ascent direction is no longer along $\Delta q_n = q_n - q_0$, but the lowest eigenmode of the Hessian at each step. More precisely, the two stages of ARTn are as follows:

1. *Activation* – From a given local minimum x_i, we move the system towards a saddle point x_c. Principle steps in this phase are:

   (a) perform random movements until a point $\widetilde{x}$ is reached such that $\lambda_1(\widetilde{x}) < 0$, where $\lambda_1(\tilde{x})$ denotes the lowest eigenvalue of the Hessian $H(\tilde{q}) = \nabla^2 V(\tilde{q})$;

(b) replace $\widetilde{x}$ by

$$\widetilde{x}' = \widetilde{x} + \mu_v v_1(\widetilde{x}), \tag{2.16}$$

where $\mu_v > 0$ is a strictly positive constant and the sign of $v_1(\widetilde{x})$, the eigenvector of $H(\tilde{q})$ associated with $\lambda_1(\tilde{x})$ is chosen such that $v_1(\tilde{x}) \cdot \nabla V(\tilde{x}) > 0$ (the eigenvector is pointing in the direction of increasing energy);

(c) the system is relaxed in the hyperplane orthogonal to $v_1(\widetilde{x}')$, *i.e.* we minimize the energy in this hyperplane by carrying out several iterations of a gradient-based method, taking $\widetilde{x}'$ as the initial point;

(d) we let $\widetilde{x} = \widetilde{x}'$ and repeat steps (b) and (c) until the (supposed) convergence towards a saddle point x_c. Once the number of iterations has exceeded a certain threshold, reject $\widetilde{x}$ and repeat algorithm from step (a).

2. *Relaxation* – From the saddle point x_c, the system is relaxed towards another local minimum.

The last relaxation step is very fast in comparison with the activation phase and is a matter of a local minimization, for which many robust algorithms exist, namely conjugate gradients (CG) or limited-memory Broyden-Fletcher-Goldfarb-Shannon (L-BFGS). We detail, therefore, the less trivial steps involved in the *activation* stage.

Step 1(a):

1. let $x^0 = x_i$; we choose a strictly positive constant μ_a and a random unitary vector $e \in \mathbb{R}^d$ and set
$$\widetilde{x}^{n+1} = x^n + \mu_a\, e$$

2. relax the system in the hyperplane orthogonal to e by carrying out several iterations of a minimization algorithm, taking $\widetilde{x}^{n+1}$ as the initial point. Call this point x^{n+1} ;

3. repeat this process until $\lambda_1(x^{n+1}) < 0$. Then set $\widetilde{x} = x^{n+1}$.

Step 1(b): after the step in the direction of the lowest eigenmode (2.16), the first order expansion in μ_v of the energy gives

$$V(\widetilde{x} + \mu_v v_1(\widetilde{x})) = V(\widetilde{x}) + \mu_v \nabla V(\widetilde{x}) \cdot v_1(\widetilde{x}) + O(\mu_v^2) > V(\widetilde{x}) + O(\mu_v^2)$$

and

$$\begin{aligned} |\nabla V(\widetilde{x} + \mu_v\, v_1(\widetilde{x}))|^2 &= |\nabla V(\widetilde{x}) + \mu_v\, H(\widetilde{x}) v_1(\widetilde{x}) + O(\mu_v^2)|^2 \\ &= |\nabla V(\widetilde{x}) + \mu_v\, \lambda_1(\widetilde{x}) v_1(\widetilde{x}) + O(\mu_v^2)|^2 \\ &= |\nabla V(\widetilde{x})|^2 + 2\mu_v\, \lambda_1(\widetilde{x})\, \nabla V(\widetilde{x}) \cdot v_1(\widetilde{x}) + O(\mu_v^2) \\ &< |\nabla V(\widetilde{x})|^2 + O(\mu_v^2). \end{aligned}$$

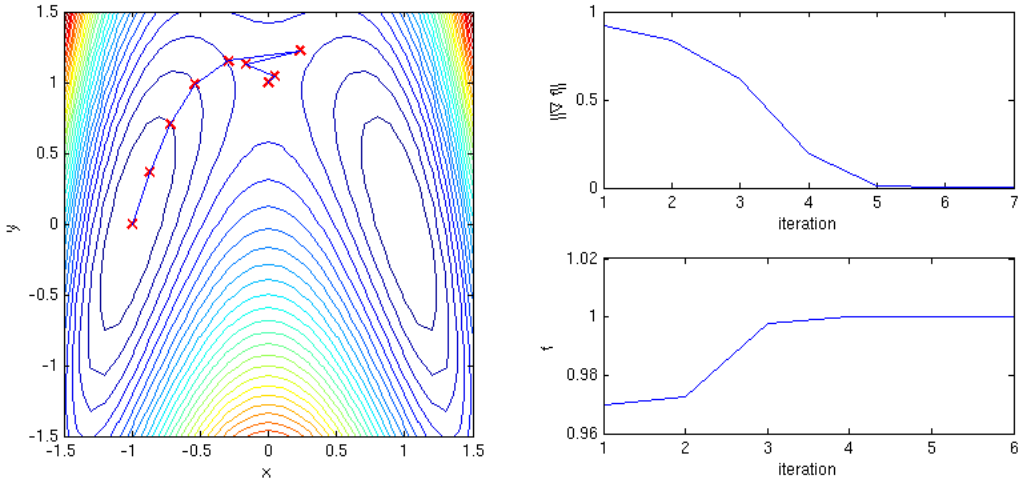


Figure 2.3: The ART algorithm applied to the potential (2.6).

where the last line is a consequence of $\lambda_1 < 0$ and $v_1 \cdot \nabla f > 0$. Therefore, step 1(b) results in an energy increase and a decrease in the norm of the gradient. The step μ_v is in practice a variable step size [Marinica 2008]: after the n^{th} step of the algorithm in the nonconvex region, the step is taken to be $\mu_v := \mu_v/\sqrt{n}$. We have shown in [Cancès 2009] that this variable step is in fact suboptimal and may be improved by incorporating gradient and curvature of the potential energy surface.

The improvement to the ARTn method is presented in Chapter 3), where the contribution is twofold:

- a prototypical algorithm m-ARTn (modified ARTn) is proposed with an optimal variable step size, depending on the gradient and curvature of the potential energy surface, see (3.1) of Section 3.2,
- local convergence analysis of the prototypical algorithm, see Appendix 3.A.

2.3.2 Climbing out of convex regions

We have, until now, detailed existing methods to converge towards a transition state once the system is already in the desired region: the algorithms rely on the existence of a single negative eigenvalue, corresponding to the lowest eigenmode. This may be considered as the reaction coordinate, describing the transition of the system from one metastable state to another. The principle bottlenecks arising in these regions concern primarily the computation of the Hessian. A key issue has been deferred until now: how a system is efficiently driven out of its convex basin. Needless to say, in a high-dimensional system, there are a large number of low-lying saddle points of interest, from which a system may escape the local basin. For methods such as

KMC, it is essential to recover as many of these as possible (in principle *all* of them are needed).

There have been attempts to drive a system out of a basin by following well-chosen directions, for example by pushing the system in the direction of the lowest eigenmodes, in order to reach a nonconvex region faster. However these techniques have shown to limit the number of relevant transition states found (typically the same transition states are found). A method that has proven successful in finding a large number of saddle points is a small *random* perturbation of the system. These small enforced perturbations on the system can be applied (i) globally, on all the atoms of the system; (ii) locally on a randomly chosen atom and (iii) locally around the system defect (vacancy or interstitial). A careful study was carried out on an iron crystal structure [Marinica 2008] to compare their impact on the number of saddle points found and the total number of force evaluations per saddle point. The global initial deformation of the system (i), proved more efficient than the random local deformation (ii), which performed the worst. The defect-centered deformation (iii) was the most efficient of the three methods, having the highest ratio of successful to unsuccessful saddle point searches, with the least number of total force evaluations. Furthermore, this last deformation technique was shown in the same work to be the least sensitive to system size. The local perturbation is essential to ensure the system is not pushed out of its current local basin. Furthermore, local perturbation has proven more robust and was the only method to remain insensitive to system size. The total number of force evaluations per successful saddle point search remained constant with growing number of atoms.

Chapter 3

An improvement to the Activation-Relaxation Technique

Ce chapitre reprend l'intégralité d'un article écrit en collaboration avec Eric Cancès, Frédéric Legoll, Mihai-Cosmin Marinica, et François Willaime, et publié dans *Journal of Chemical Physics* [Cancès 2009].

Sommaire

3.1 Introduction **39**
3.2 A new type of ART-like algorithms **40**
3.3 Numerical results: Migration of point defects in α-iron . . **43**
3.A Appendix: Local convergence analysis **47**

3.1 Introduction

The Activation-Relaxation Technique (ART) [Barkema 1996b, Barkema 1998, Mousseau 1998, Barkema 1999, Mousseau 1999, Mousseau 2000b, Barkema 2001] is a powerful method for searching saddle points and transition pathways of a given potential energy surface (PES). Search methods for saddle points and transition pathways can actually be classified in two main categories. In the first class of methods, one assumes that two local minima of the PES are known. The main objective of the methods in this class is to find the minimum energy path to go from one local minimum to the other one. The Replica Chain method [Elber 1987, Ayala 1997], the Nudged Elastic Band [Mills 1994, H. Jonsson 1998, Henkelman 2000a], the String method [E 2002, Peters 2004], the Transition Path Sampling [Dellago 1998, C. Dellago 1999, Dellago 2002, Bolhuis 2002b] and the Discrete Path Sampling [Wales 2002] are some methods belonging to this class (note that the Nudged Elastic Band method has been generalized to the finite temperature setting [Crehuet 2003], as well as the String method [E 2005a]). In the second class of methods, one assumes that *only one* local minimum of the PES is known. The aim of methods in this class is to find a saddle point of the PES, from which the exploration will be pursued toward a different local minimum, yielding a transition path. Probably the first method proposed in that class is the EigenVector Following method [Cerjan 1981]. The Dimer method [Henkelman 1999], the Hybrid EigenVector Following method [Munro 1999, Wales 2003], the Conformational Flooding method [Grubmüller 1995],

the Hyperdynamics method [Voter 1997b, Voter 1997a, Miron 2004], the Parallel Replica method [Voter 1998], the Temperature Accelerated method [Sorensen 2000, F. Montalenti 2001, Montalenti 2002], and the Scaled Hypersphere Search method [Maeda 2003, Ohno 2004, Maeda 2005] are other examples. In this article, we study the Activation-Relaxation Technique, which belongs to this second class. We focus here on the zero temperature case, the so-called *ART nouveau* (ARTn) method [Mousseau 2000b, Malek 2000], and do not consider the finite temperature case, the so-called *POP-ART* method [Vocks 2005].

The ARTn method is composed of two main steps, the activation step and the relaxation step. The activation step consists in moving the system from a local minimum to a saddle point. The relaxation step consists in relaxing the system, from the computed saddle point, to another local minimum. Of course, this relaxation step is very fast (and easy to perform) compared to the activation step.

The activation step itself can be divided into two substeps. The first substep aims at finding some region of the PES with one direction of negative curvature, which hopefully contains a first order saddle point, and that we will call the "attracting region". The basic idea for finding a point on the PES with one direction of negative curvature is to choose a random vector r, and next to repeat the two following operations: (i) move the system according to r, (ii) relax the system in the hyperplane orthogonal to r, until a point with one direction of negative curvature has been found (see Sec. 3.3 for details). The second substep consists in finding a saddle point in the reached attracting region. From a numerical viewpoint, these two substeps are of very different nature. In this article, we focus on the second substep, namely the local convergence to a saddle point, starting from a configuration with one direction of negative curvature. In Sec. 3.2, we present a simple, prototypical, algorithm, amenable to a comprehensive mathematical analysis, and that enlightens the principles of operation of a wide class of saddle point location techniques, including the ARTn and the Hybrid EigenVector Following methods. A proof of convergence and robustness of this algorithm is provided in the Appendix. Loosely speaking, this algorithm is optimal in the principal direction of negative curvature, but suboptimal in the transverse directions. Obviously, more complex numerical strategies can be elaborated on the basis of this prototypical algorithm. They are discussed at the end of Sec. 3.2. Using one of these strategies, we obtain an improved ART-like algorithm (see eq. (3.5) below), with which we study the problem of vacancy diffusion in crystalline materials. The obtained numerical results are reported on in Sec. 3.3, and clearly demonstrate the efficiency of our approach.

3.2 A new type of ART-like algorithms

From a mathematical viewpoint, a PES for an isolated molecular system with N atoms is a function $E : \mathbb{R}^{3N}/G_r \longrightarrow \mathbb{R}$, where $G_r = \mathbb{R}^3 \times \mathrm{SO}(3)$ is the group of rigid body motions which acts on $\mathbb{R}^{3N}$ in the following way: for all $g = (x_0, R) \in$

$G_r = \mathbb{R}^3 \times \mathrm{SO}(3)$, and for all $X = (x^1, \cdots, x^N) \in \mathbb{R}^{3N}$,

$$g \cdot X = (R(x^1 - x_0), \cdots, R(x^N - x_0)).$$

This viewpoint takes into account the fact that the potential energy $E(X)$ of the system is invariant upon rigid body motions. In the simulation of the condensed phase, artificial periodic boundary conditions are usually introduced. In this case, the system is translation invariant, but not rotation invariant, and a PES then has to be regarded as a function $E : \mathbb{T}^{3N}/\mathbb{R}^3 \longrightarrow \mathbb{R}$, where $\mathbb{T}^{3N}$ is a $3N$ dimensional torus.

For our purpose, namely for the analysis of saddle point location methods, there is no restriction in assuming that the PES under consideration is a function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ with isolated critical points. For $x \in \mathbb{R}^d$, we denote by $\nabla f(x)$ the gradient of f at the point x and by $H(x) = \nabla^2 f(x)$ the Hessian of f at the point x. For $x \in \mathbb{R}^d$, let $\lambda_1(x) \leq \lambda_2(x) \leq \cdots \leq \lambda_d(x)$ be the eigenvalues of $H(x)$ counted with their multiplicity, and let $(v_1(x), \cdots, v_d(x))$ be an orthonormal basis of associated eigenvectors.

Contrarily to second order methods, such as the one proposed in Ref. [Cerjan 1981], the ARTn and Hybrid EigenVector Following methods do not rely on a complete knowledge of the spectral decomposition of the Hessian matrix. Instead, they only make use of the direction of negative curvature. A prototype of such algorithm reads

$$x_{k+1} = x_k - \frac{(\nabla f(x_k), v_1(x_k))}{\min(\lambda_1(x_k), -\lambda_c)}\, v_1(x_k) - \mu_t\, \Pi_{v_1(x_k)^\perp} \nabla f(x_k), \tag{3.1}$$

where $\lambda_c > 0$ and $\mu_t > 0$ are fixed numerical parameters, and $\Pi_{v_1(x_k)^\perp} = I - (v_1(x_k), \cdot)v_1(x_k)$ is the orthogonal projector on the hyperplane $v_1(x_k)^\perp$.

In order to clarify the behavior of algorithm (3.1) and the role of the numerical parameters $\lambda_c > 0$ and $\mu_t > 0$, let us consider the simple example of a quadratic function f:

$$f(x) = \frac{1}{2} \sum_{j=1}^{d} \lambda_j |x^j|^2, \tag{3.2}$$

with $x = (x^1, \cdots, x^d)$ and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$. In this simple case, (3.1) reads as a system of d decoupled scalar equations

$$\begin{aligned} x^1_{k+1} &= \left(1 - \frac{\lambda_1}{\min(\lambda_1, -\lambda_c)}\right) x^1_k, \\ x^j_{k+1} &= (1 - \mu_t \lambda_j)\, x^j_k, \qquad 2 \leq j \leq d, \end{aligned}$$

yielding

$$\begin{aligned} x^1_k &= \left(1 - \frac{\lambda_1}{\min(\lambda_1, -\lambda_c)}\right)^k x^1_0, \\ x^j_k &= (1 - \mu_t \lambda_j)^k\, x^j_0, \qquad 2 \leq j \leq d, \end{aligned}$$

where x_0 is the initial guess of the algorithm.

Assume that all the λ_j are different from zero. In this case, f has a unique critical point (the origin), and the algorithm converges to this critical point for all choices of the initial guess if and only if

$$\lambda_1 < 0 \quad \text{and} \quad 0 < \lambda_j < 2\mu_t^{-1} \qquad \text{for all } 2 \leq j \leq d.$$

This means that if the algorithm converges, it will be toward a critical point with Morse index equal to one (a first order saddle point). Conversely, if the origin is a saddle point with Morse index equal to one (i.e. if $\lambda_1 < 0 < \lambda_2 \leq \cdots \leq \lambda_d$), the algorithm will converge to it if and only if $\lambda_d < 2\mu_t^{-1}$. The numerical parameter μ_t controls the convergence in the hyperplane $x^1 = 0$. If μ_t is too small, convergence will be slow; if μ_t is too large, the algorithm will be unstable. Note that if $\lambda_1 < -\lambda_c$, convergence in the e_1 direction (the direction of negative curvature) will be obtained in a single iteration, while linear convergence will be observed if $-\lambda_c < \lambda_1 < 0$. The role of the parameter λ_c is to prevent the algorithm, when applied to a non-quadratic energy landscape, from becoming unstable in the region where $|\lambda_1(x_k)|$ is small.

Let us now come back to the case of practical interest when f is the PES of some molecular system. As mentioned in the Introduction, we focus here on the local convergence properties and henceforth assume that the iterates have reached the neighborhood of a first order saddle point. One can then prove (see the Appendix) that algorithm (3.1) converges to the saddle point, quadratically in the principal direction of negative curvature, and linearly in the perpendicular directions. Note that quadratic convergence is obtained under the assumption that the smallest eigenvalue $\lambda_1(x_k)$ of the Hessian matrix $H(x_k)$ and the corresponding eigenvector $v_1(x_k)$ are computed exactly. However, a key ingredient in ART-like and Hybrid EigenVector Following algorithms is that $\lambda_1(x_k)$ and $v_1(x_k)$ are computed approximately, by iterative methods. Thus, for instance, the eigenelement $(\lambda_1(x_k), v_1(x_k))$ can be computed by Lanczos or Arnoldi algorithms, as in the ARTn method, or by minimizing the Rayleigh-Ritz ratio, as in the Hybrid EigenVector Following method. For both approaches, repeated evaluations of matrix-vector products of the form $H(x_k)\, v$ are needed. In turn, such matrix-vector products can be approximately computed using a finite-difference formula, such as the first-order formula

$$H(x_k)\, v \approx \frac{1}{\epsilon} \left(\nabla f(x_k + \epsilon v) - \nabla f(x_k) \right) \tag{3.3}$$

or the second order formula

$$H(x_k)\, v \approx \frac{1}{2\epsilon} \left(\nabla f(x_k + \epsilon v) - \nabla f(x_k - \epsilon v) \right). \tag{3.4}$$

In summary, the eigenelement $(\lambda_1(x_k), v_1(x_k))$ is computed approximately by repeated evaluations of forces $-\nabla f(y)$ for a collection of configurations y close to the reference configuration x_k. One can prove (see again the Appendix) that algorithm (3.1) is robust, in the sense that it can accomodate approximate evaluations

of $(\lambda_1(x_k), v_1(x_k))$. The price to pay is a lower convergence rate in the principal direction of negative curvature.

The prototypical algorithm (3.1) is not far from being optimal in the direction of negative curvature, even in presence of numerical errors in the evaluation of $(\lambda_1(x_k), v_1(x_k))$. On the other hand, it is clearly suboptimal in the transverse directions, where it behaves as a basic fixed step-size gradient. Let us recall that the reason why we focused on algorithm (3.1) in the present section is that it allows for a both simple and comprehensive mathematical analysis, enlightening the principles of operation of a wide class of saddle point location techniques. In the applications, it is useful to resort to more advanced minimization methods in the transverse direction, such as conjugate gradients, damped molecular dynamics, quasi-Newton or trust-region methods [Nocedal 2000]. In the numerical examples reported below, we make use of damped molecular dynamics, which provided very satisfactory results in that context. Instead of using (3.1), we hence work with the following algorithm, that we call the modified ARTn (m-ARTn) algorithm:

$$\begin{aligned}&\text{Set } x_{k+1} = x_k - \frac{(\nabla f(x_k), v_1(x_k))}{\min(\lambda_1(x_k), -\lambda_c)}\, v_1(x_k),\\ &\text{where } \lambda_1(x_k) \text{ and } v_1(x_k) \text{ are estimated using the Lanczos method,}\\ &\text{then relax } x_{k+1} \text{ in the hyperplane } v_1(x_k)^\perp \text{ by damped molecular dynamics}\\ &\text{(see Section 3.3 for details).}\end{aligned} \tag{3.5}$$

It is also possible, in principle, to take into account the p lowest eigenvalues of $H(x_k)$ obtained from the Lanczos or Arnoldi partial diagonalization procedure, to construct a surrogate function that will provide a better model for f in the neighborhood of x_k. Such improvements of the current ART-like algorithms will be considered in a future work.

3.3 Numerical results: Migration of point defects in α-iron

In this section we discuss the practical implementation of the m-ARTn algorithm (3.5) in the case of basic defects in α-iron: small self interstitial (SIA) and vacancy (VAC) clusters (1 to 3 defects). The crystal of α-Fe is modeled by the EAM potential developed by Mendelev *et al.* [Mendelev 2003, Ackland 2004] which has been the most widely used potential in recent years to study interstitial loops [Terentyev 2007, Terentyev 2008]. Marinica *et al.* have previously used the same potential and the standard ARTn method to test and reveal the energy landscape of small interstitial clusters (1 to 4 self-interstitials) in α-Fe [Marinica 2008]. It therefore gives us a good basis for comparison. The crystal consists of $1024\pm n$ atoms (n=1,2,3).

Starting from a local minimum configuration, the first stage of the activation step is to push the system out of the basin. In order to do this, the system is

slightly deformed using

$$x_{k+1} = x_k + \mu_A \Delta x \tag{3.6}$$

where Δx is a random variable and μ_A is a user-defined fixed step, that are both kept constant throughout the activation step. In this paper we use defect centered deformations [Marinica 2008] instead of global deformations. This means that the components of Δx corresponding to atoms located beyond a certain cut-off radius around the defects are set to zero. The other components of Δx are drawn according to a uniform distribution. In this particular application, the defects are located using the Wigner-Seitz method [Terentyev 2007], and the cut-off radius is set to 5 Å. More sophisticated strategies for chosing Δx and μ_A will be explored in future work. The reason for choosing defect centered deformations rather than global deformations is that the efficiency of the algorithm is then independent of the size of the system. Such a choice also provides a better rate of successful to unsuccessful activation processes.

At each iteration the system is relaxed in the hyperplane orthogonal to the direction Δx. If, after this relaxation, the lowest eigenvalue is still positive, we continue the deformation. As soon as $\lambda_1(x_k)$ becomes sufficiently negative ($\lambda_1(x_k) < \lambda_d$ for some threshold $\lambda_d < 0$), we move onto the next stage of the activation process. The threshold is used in order not to be misguided by numerical errors of the eigenvalue calculation. The lowest eigenvalue is computed using the Lanczos algorithm with 15 iterations, a small number compared to the size of the Hessian matrix (recall that $H(x_k) \in \mathbb{R}^{3N\times 3N}$, where N is the number of atoms in the system).

Once the system is out of the basin, we begin to move it towards the saddle point, in the hope of following the minimum-energy reaction path. The previously used method [Marinica 2008] for this stage is:

$$x_{k+1} = x_k - \frac{\mu_a}{\sqrt{k}} v_1(x_k) \tag{3.7}$$

where μ_a is a user-defined constant and $1/\sqrt{k}$ ensures that the step size gets smaller as we approach the saddle point. The direction of the eigenvector v_1 is chosen such that it points in the same direction as the force i.e. $(-\nabla f(x_k), v_1(x_k)) > 0$. This is then followed by a relaxation in the hyperplane, which is discussed in the next paragraph. Algorithm (3.7) was an improvement to some previous methods [Mousseau 2000b, Malek 2000]. However it has several drawbacks. The constant parameter μ_a in the algorithm needs to be defined according to the PES in study, and even so may be suited for some saddle point searches but not for others (very tightly positioned saddle points may force μ_a to be small for the whole system, which may in turn impede results when the surface becomes relatively smooth). With $v_1(x_k)$ unitary and μ_a fixed, it is clear that decreasing the step size according to the number of iterations is not ideal. In fact it would be better to use the first and second derivative information of the energy surface. Taking as a simple example the function (3.2) with $d = 2$ (solution x_* at the origin), we can position ourselves at a

point x_n where the displacement from x_* is in the direction of negative curvature. The further along this direction we are positioned, the more iterations would be needed to approach x_* since algorithm (3.7) would take smaller and smaller steps. In this simple case, algorithms (3.1) and (3.5) would jump to the solution in one step.

In our current implementation of the m-ARTn algorithm, the energy minimization in the orthogonal hyperplane is performed by means of damped molecular dynamics. A friction proportional to the velocities is added to the forces. The system is then propagated with the Verlet algorithm in the hyperplane $v_1(x_k)^\perp$, using a time step of 20 fs. The dynamics is carried on until a maximum number of steps, n_k, is reached, or some convergence criterion (in our example, $\max_{i=1,N} |\nabla_i f(x_k)| < 0.01$ eV /Å) is fullfilled. The maximum number of iterations n_k is gradually increased along the iterations up to a maximum value M, according to $n_k = \min(k, M)$. The reason why the maximum number of minimization steps is limited to M is twofold. On the one hand, this ensures that the algorithm does not become too costly. On the other hand, it is unnecessary to accurately relax the system if it is still far away from the saddle point.

In the case when we reach a configuration x_k with $\lambda_1(x_k) > 0$, we restart the activation step using a different random deformation field Δx (see Ref. [Marinica 2008] for details on relaxation).

The main contribution of algorithm (3.5) is the step taken in the direction of the negative curvature. In the numerical results reported later on in this section, we have implemented this step in the attracting region. On the other hand, we continue to use equation (3.6) for leaving the basin. For both stages, we use damped molecular dynamics for the relaxation in the orthogonal hyperplane.

The efficiency of ART-type algorithms depends on two main points: the number of force evaluations required during the activation stage and the ratio of successful to unsuccessful searches. The failure to find a saddle point can be determined in several ways. If minimization in the hyperplane is not done sufficiently well, the system risks climbing the energy surface too high. Once settled at a saddle point, it could be one which is not associated with the local minimum where the activation process began. It could also be the case that we fall on a saddle point where the energy is lower than the starting point, which is an immediate indication that we have overlooked at least one adjacent saddle point of the local minimum and fallen beyond. Finally, another sign of failure is when relaxation in the hyperplane yields a positive $\lambda_1(x_k)$, in which case we have reached another local minimum. It remains a challenge to be certain that a saddle point falls in the first of the three categories mentioned. For the purposes of this study, we will only reject stationary configurations if the energy is below that of the initial local minimum or if we are in fact at another minimum configuration.

Comparisons between algorithm (3.7) and the m-ARTn algorithm (3.5) are done on interstitial and vacancy defects using the parameters shown in Table 3.1. The

	n SIA		n VAC	
	Ref. [Marinica 2008]	This work	Ref. [Marinica 2008]	This work
λ_c	-	0.5	-	0.5
λ_d	-2	-2	-2	-2
μ_A	0.6	0.6	0.2	0.2
μ_a	0.24	-	0.08	-
M	18	18	18	18

Table 3.1: Parameters used in implementation. The parameter μ_A is taken from studies by Marinica *et al.* [Marinica 2008], with $\mu_a/\mu_A = 0.4$.

number of defects		SIA ARTn [Marinica 2008]	SIA This work	VAC ARTn [Marinica 2008]	VAC This work
1	$\langle f \rangle$	462	298	780	291
	η	4.6	4.7	1.8	7.9
2	$\langle f \rangle$	548	328	705	323
	η	4.2	4.4	2.6	7.1
3	$\langle f \rangle$	691	320	667	321
	η	2.6	4.4	2.8	7.4

Table 3.2: Comparison of a previous ARTn approach [Marinica 2008] and the algorithm m-ARTn (3.5) presented in this article for interstitial and vacancy defects. The new algorithm reduces the average number of force evaluations ($\langle f \rangle$) by about 40% and 55% for the self-interstitial atoms (SIA) and vacancies (VAC) case respectively. In the case of SIA, the ratio of successful to unsuccessful searches (η) is almost constant. However, in the case of vacancies, η is increased by over 260%.

results are shown in Table 3.2. Over a total number of 1000 successful events, $\langle f \rangle$ is the average number of force evaluations per activation process and η is the ratio of successful events to unsuccessful events. We observe that the proposed algorithm (3.5) improves performance by a large margin both in terms of the average number of force evaluations and the proportion of successful events. The elimination of the constant factor μ_a in algorithm (3.7) not only makes the algorithm more efficient but also more versatile. It may be applied to a wide range of potential energy surfaces without the need for parameter manipulation.

Aknowledgements

This work was partially supported by the Agence Nationale de la Recherche (LN3M project, contract No. ANR-05-CIGC-0003). The authors are grateful to Normand

Mousseau for helpful discussions.

3.A Appendix: Local convergence analysis

In this Appendix, we prove that algorithm (3.1) is locally convergent, even when the eigenelement in the direction of negative curvature is approximately computed.

Let x_* such that $\nabla f(x_*) = 0$ and $\lambda_1(x_*) < 0 < \lambda_2(x_*) \leq \cdots \leq \lambda_d(x_*)$. We introduce the notation $v_1^* = v_1(x_*)$, $\lambda_1^* = \lambda_1(x_*)$, $H_* = \nabla^2 f(x_*)$,

$$e_k = x_k - x_*, \quad z_k = (x_k - x_*) \cdot v_1^*, \quad y_k = \Pi_{v_1^{*\perp}}(x_k - x_*).$$

Note that

$$x_k - x_* = z_k v_1^* + y_k, \quad \text{hence} \quad |x_k - x_*|^2 = |z_k|^2 + |y_k|^2.$$

In the analysis below, we often use that $z_k = O(|e_k|)$.

We consider algorithm (3.1), and assume that the eigenelement $(\lambda_1(x_k), v_1(x_k))$ is *approximately* computed. The resulting algorithm, that we analyze below, reads

$$x_{k+1} = x_k - \frac{(\nabla f(x_k), \tilde{v}_1(x_k))}{\min(\tilde{\lambda}_1(x_k), -\lambda_c)} \tilde{v}_1(x_k) - \mu_t \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k), \tag{3.8}$$

where $\tilde{\lambda}_1(x_k)$ and $\tilde{v}_1(x_k)$ are approximations of $\lambda_1(x_k)$ and $v_1(x_k)$:

$$\tilde{v}_1(x_k) = v_1(x_k) + \alpha_k, \quad \tilde{\lambda}_1(x_k) = \frac{\lambda_1(x_k)}{1 + \beta_k},$$

where the errors α_k and β_k are supposed to be small (i.e. $|\alpha_k| \ll 1$ and $|\beta_k| \ll 1$). We assume that $|\tilde{v}_1(x_k)| = 1$. Note that we have made no assumption on the Hessian matrix $H(x_k)$. Hence, the errors α_k and β_k take into account both a possible approximation in the computation of $H(x_k)$ (see (3.3) and (3.4)), and an approximate partial diagonalization of this matrix (by a Lanczos, Arnoldi, or Rayleigh-Ritz ratio minimization algorithm).

We assume that $\lambda_1(x_*) < -\lambda_c$ and that x_k is close enough to x_* such that $\lambda_1(x_k) \leq -\lambda_c$ for all k sufficiently large. We also assume that the error β_k is small enough such that $\tilde{\lambda}_1(x_k) \leq -\lambda_c$ for all k sufficiently large.

It follows from (3.8) that

$$\begin{aligned} z_{k+1} &= z_k - \frac{(\nabla f(x_k), \tilde{v}_1(x_k))}{\tilde{\lambda}_1(x_k)} (\tilde{v}_1(x_k), v_1^*) - \mu_t (v_1^*, \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k)) \\ &= z_k - \frac{(\nabla f(x_k), v_1(x_k) + \alpha_k)}{\lambda_1(x_k)} (1 + \beta_k)(v_1(x_k) + \alpha_k, v_1^*) \\ &\quad - \mu_t (\Pi_{\tilde{v}_1(x_k)^\perp} v_1^*, \nabla f(x_k)) \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} y_{k+1} &= y_k - \frac{(\nabla f(x_k), \tilde{v}_1(x_k))}{\tilde{\lambda}_1(x_k)} \Pi_{v_1^{*\perp}} \tilde{v}_1(x_k) - \mu_t \Pi_{v_1^{*\perp}} \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k) \\ &= y_k - \frac{(\nabla f(x_k), v_1(x_k) + \alpha_k)}{\lambda_1(x_k)} (1 + \beta_k) \Pi_{v_1^{*\perp}} \tilde{v}_1(x_k) - \mu_t \Pi_{v_1^{*\perp}} \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k) \end{aligned} \tag{3.10}$$

Assuming that f is $C^2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ with bounded first and second derivatives, it holds

$$\nabla f(x_k) = H_*(x_k - x_*) + O(|x_k - x_*|^2) = \lambda_1^* z_k v_1^* + H_* y_k + O(|e_k|^2). \tag{3.11}$$

Besides, using perturbation theory, one obtains

$$\begin{aligned} \lambda_1(x_k) &= \lambda_1^* + (v_1^*, (\nabla H(x_*) \cdot e_k)\ v_1^*) + O(|e_k|^2), \\ v_1(x_k) &= v_1^* - \Pi_{v_1^{*\perp}} \left((H_* - \lambda_1^*)\,|_{v_1^{*\perp}} \right)^{-1} \Pi_{v_1^{*\perp}} (\nabla H(x_*) \cdot e_k)\ v_1^* + O(|e_k|^2). \end{aligned} \tag{3.12}$$

From (3.11) and (3.12), we deduce

$$\begin{aligned} (\nabla f(x_k), v_1(x_k)) &= \lambda_1^* z_k + O(|e_k|^2), \\ (v_1(x_k), v_1^*) &= 1 + O(|e_k|^2), \\ \Pi_{\tilde{v}_1(x_k)^\perp} v_1^* &= O(|e_k|) + O(|\alpha_k|), \\ \Pi_{v_1^{*\perp}} \tilde{v}_1(x_k) &= O(|e_k|) + O(|\alpha_k|), \\ \Pi_{v_1^{*\perp}} \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k) &= H_* y_k + O(|e_k|^2) + O(|e_k|\,|\alpha_k|). \end{aligned}$$

Inserting these equations in (3.9) and (3.10), we obtain

$$\begin{aligned} z_{k+1} &= z_k - \frac{(\nabla f(x_k), v_1(x_k))}{\lambda_1(x_k)} (v_1(x_k), v_1^*) - \mu_t (\Pi_{\tilde{v}_1(x_k)^\perp} v_1^*, \nabla f(x_k)) \\ &\quad - \frac{(\nabla f(x_k), \alpha_k)}{\lambda_1(x_k)} (1 + \beta_k)(v_1(x_k) + \alpha_k, v_1^*) \\ &\quad - \frac{(\nabla f(x_k), v_1(x_k))}{\lambda_1(x_k)} \beta_k (v_1(x_k) + \alpha_k, v_1^*) \\ &\quad - \frac{(\nabla f(x_k), v_1(x_k))}{\lambda_1(x_k)} (\alpha_k, v_1^*) \\ &= O(|e_k|^2) + O(|e_k|\,|\alpha_k|) + O(|e_k|\,|\beta_k|) \end{aligned} \tag{3.13}$$

on the one hand, and, on the other hand,

$$\begin{aligned} y_{k+1} &= y_k - \frac{(\nabla f(x_k), v_1(x_k))}{\lambda_1(x_k)} \Pi_{v_1^{*\perp}} \tilde{v}_1(x_k) - \mu_t \Pi_{v_1^{*\perp}} \Pi_{\tilde{v}_1(x_k)^\perp} \nabla f(x_k) \\ &\quad - \frac{(\nabla f(x_k), \alpha_k)}{\lambda_1(x_k)} \Pi_{v_1^{*\perp}} \tilde{v}_1(x_k) \\ &\quad - \frac{(\nabla f(x_k), v_1(x_k) + \alpha_k)}{\lambda_1(x_k)} \beta_k \\ &= (I - \mu_t H_*) y_k + O(|e_k|^2) + O(|e_k|\,|\alpha_k|) + O(|e_k|\,|\beta_k|). \end{aligned} \tag{3.14}$$

Note that $y_k \in v_1^{*\perp}$ and that H_* is positive definite on $v_1^{*\perp}$. We compute

$$\|(I - \mu_t H_*)_{|v_1^{*\perp}}\|_2 = \max(1 - \mu_t \lambda_2, \mu_t \lambda_d - 1).$$

Thus, if $\mu_t < 2/\lambda_d$, we infer from (3.13) and (3.14) that

$$|x_{k+1} - x_*| \le \gamma |x_k - x_*| + O(|x_k - x_*|^2) + O(|x_k - x_*|\,|\alpha_k|) + O(|x_k - x_*|\,|\beta_k|),$$

with $\gamma = \|(I - \mu_t H_*)_{|v_1^{*\perp}}\|_2 < 1$. Under the assumption that the errors α_k and β_k are uniformly bounded by a small constant, this proves that algorithm (3.8) locally converges, and that the convergence speed is at least linear.

In the case when the eigenelement $(\lambda_1(x_k), v_1(x_k))$ is exactly computed, algorithm (3.8) reduces to algorithm (3.1). We hence have proved that algorithm (3.1) locally converges, and that this convergence is robust with respect to errors in the computations of the lowest eigenvalue (and the associated eigenvector) of $H(x_k)$.

Estimates for the convergence of algorithm (3.1) are readily obtained from (3.13) and (3.14), by setting $\alpha_k = 0$ and $\beta_k = 0$. We obtain

$$z_{k+1} = O(|e_k|^2) \quad \text{and} \quad y_{k+1} = (1 - \mu_t H_*) y_k + O(|e_k|^2).$$

Note that the convergence for z_k (e.g. in the principal direction of negative curvature) is *quadratic*. If errors are introduced in the computation of the eigenelement $(\lambda_1(x_k), v_1(x_k))$, the rate of convergence of z_k becomes *linear*, as can be seen in (3.13).

Part II

Free Energy Computations

Chapter 4

Free Energy Computations

Sommaire

4.1 Introduction 53
4.1.1 Describing a transition 55
4.1.2 Choosing a good reaction coordinate 58
4.2 Computing free energy differences 60
4.2.1 Thermodynamic integration 60
4.2.2 Other equilibrium free energy methods 64
4.2.3 Nonequilibrium free energy methods 67
4.3 Adaptive sampling methods 68
4.3.1 Adaptive Biasing Potential 69
4.3.2 Adaptive Biasing Force 71
4.4 Longtime convergence of dynamics 73
4.4.1 Entropy and the Logarithmic Sobolev Inequality 73
4.4.2 Convergence of the overdamped dynamics 76
4.4.3 Convergence of ABF dynamics 77

This chapter gives an overview of some widely used free energy computation methods, to which a major part of this thesis is devoted. Our contribution to this field, namely new numerical and theoretical results, will be presented in Chapters 5 and 6 respectively.

4.1 Introduction

The Helmholtz free energy is a quantity of great importance in understanding chemical and biochemical processes. By the second law of thermodynamics, a system in the canonical ensemble is at equilibrium when the Helmholtz entropy is maximized or the free energy is minimized. The absolute free energy, denoted in the following by A, is defined as

$$A = -\beta^{-1} \ln Z, \tag{4.1}$$

where Z is the so-called partition function defined by

$$Z = \frac{1}{h^{3N} N!} \int_{\Omega} e^{-\beta H(q,p)} \, dq \, dp.$$

Here h is the Planck constant (4.136×10^{-15} eV s) and $N!$ accounts for the N indistinguishable particles of the system. We recall that H is the Hamiltonian of the system (as defined in (1.1)), $\Omega = \mathbb{R}^{3N} \times \mathbb{R}^{3N}$ (or $\mathbb{T}^{3N} \times \mathbb{R}^{3N}$ in the case of periodic boundary conditions) and $\beta = 1/k_B T$ is the inverse temperature, with k_B the Boltzmann constant. The free energy and partition function are thus related: the computation of A determines Z and vice-versa. Partition functions are useful in thermodynamics as many fundamental properties may be derived from them. The average energy of a system, for example, is computed as

$$\langle H \rangle_\mu = \frac{\displaystyle\int_\Omega H(q,p) e^{-\beta H(q,p)} \, dq \, dp}{\displaystyle\int_\Omega e^{-\beta H(q,p)} \, dq \, dp} = -\frac{\partial \, \ln Z}{\partial \beta},$$

and its variance is $\langle (H - \langle H \rangle_\mu)^2 \rangle_\mu = \dfrac{\partial^2 \ln Z}{\partial \beta^2}$, where μ is here, and in the following, assumed to be the canonical measure

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

as defined in (1.6). Another quantity of thermodynamic relevance is the heat capacity, defined in terms of the partition function as

$$C = -\frac{1}{k_B T^2} \frac{\partial^2 \ln Z}{\partial \beta^2}.$$

Remark 2 *We may compare (4.1) to the free energy introduced in macroscopic thermodynamics $A = U - TS$, where U is the internal energy of the system, T is the temperature and S the entropy. The internal energy is defined as the canonical average of the energy, $U = \displaystyle\int_\Omega H \, d\mu$ and the entropy is defined as $S = -k_B \int_\Omega \ln\left(\frac{d\mu}{dx}\right) d\mu$, where $x = (q,p)$ so that $dx = dq \, dp$. Using in order the relations $k_B T = \beta^{-1}$, $\ln\left(\frac{d\mu}{dx}\right) = -\ln Z - \beta H$ and $\int d\mu = 1$, we have*

$$\begin{aligned} A &= U - TS = \int H d\mu + k_B T \int \ln\left(Z^{-1} e^{-\beta H}\right) d\mu \\ &= \int (H - \beta^{-1} \ln Z - H) d\mu \\ &= -\beta^{-1} \ln Z, \end{aligned}$$

which is indeed (4.1) and is a more useful definition from a mathematical viewpoint when studying free energy differences (see Section 4.2).

Calculating the partition function, however, is in practice infeasible due to the integration over a high-dimensional domain Ω. Fortunately there exist efficient methods to compute free energy differences, from which the ratio of partition functions may be deduced. The free energy difference between two systems or states of a

system is also of interest to practitioners, as it gives the relative probability of each state. Supposing the initial and final states of the system are indexed by 0 and 1, the free energy difference ΔA is defined by

$$\Delta A = A_1 - A_0 = -\beta^{-1} \ln \left(\frac{Z_1}{Z_0} \right), \tag{4.2}$$

where Z_0 and Z_1 are the partition functions corresponding to the initial and final phases. In fact for computational reasons it is convenient to index intermediary stages of a transformation by a parameter ranging from 0 to 1 (see Section 4.2.2). The choice of parameterization is not trivial and depends entirely on the system transition.

4.1.1 Describing a transition

The choice of parameterization depends on the system transformation and typically falls into one of two categories.

(i) Two states of a system may be described by different Hamiltonian functions. The immersion of a molecule in a solvent, for example, would be characterized by Hamiltonians H_0 and H_1 (thus macroscopically by their associated canonical measures), where H_0 would account for interatomic interactions of the molecule in vacuum, and H_1 would incorporate solute-solvent interactions. The insertion of a particle in an initially closed system would also be treated similarly. This is known as an *alchemical transformation*, referring to the traditional practice of alchemy, the transmuting of a base substance into another.

(ii) An alternative parameterization, through an *order parameter* (or *reaction coordinate*) $\xi : \mathbb{R}^{3N} \to \mathbb{R}^m$ (or $\mathbb{T}^m$), $m \ll 3N$, is suitable for transitions involving internal coordinates of the system such as the conformational change of a macromolecule. In this case, the transformation is described by a coarse-grained variable $\xi(q)$, typically representing the system's slower degrees of freedom (see Section 4.1.2 for a discussion on choosing a *good* reaction coordinate). For a protein, this could represent a dihedral angle, the distance between two groups of atoms, or indeed the end-to-end distance of the protein chain. In this case, the initial and final states are described by the canonical measure μ *conditioned* to $\xi(q) = z_0$ and $\xi(q) = z_1$ respectively (see Section 4.2 for more detail).

Before discussing how free energy differences are computed in practice, let us give some mathematical precision for these two approaches.

Alchemical transformations

Suppose that the initial and final states of a system are characterized by Hamiltonians H_0 and H_1 respectively. The partition functions Z_0 and Z_1 in (4.2) are

computed based on their associated Hamiltonians, so that the free energy difference may be written as

$$A(1) - A(0) = -\beta^{-1}\ln\left(\frac{\int_\Omega e^{-\beta H_1(q,p)}\ dq\ dp}{\int_\Omega e^{-\beta H_0(q,p)}\ dq\ dp}\right). \tag{4.3}$$

In fact, intermediate stages of the transformation may be represented by indices $\alpha \in [0,1]$, where each value is associated to an (often physically meaningless) Hamiltonian H_α. The Hamiltonians differ often only in their potential function term:

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

The potential function is in turn defined as a linear combination of two reference potentials V_0 and V_1 in the following way:

$$V_\alpha = (1-\alpha)V_0 + \alpha V_1.$$

Taking as an example the insertion of a particle into a closed system of N atoms, the initial and final potential energy functions may be defined as

$$V_0 = \sum_{i=1}^{N}\sum_{j>i} V_{\mathrm{LJ}}(|q_i - q_j|) \quad \text{and} \quad V_1 = \alpha\sum_{i=1}^{N+1}\sum_{j>i} V_{\mathrm{LJ}}(|q_i - q_j|).$$

Thus V_0 represents the potential for the N-atom system and V_1 represents the potential for the $(N+1)$-atom system. The interatomic potential V_{LJ} is the Lennard-Jones potential defined in (1.17).

Transformation via an order parameter

A *reaction coordinate* $\xi : \mathbb{R}^{3N} \to \mathbb{R}^m$ (or indeed $\mathbb{T}^m$), with $m \ll 3N$, is used to describe a transformation involving only internal coordinates of the system. In the m-dimensional case, $\xi(q) = (\xi_1(q), \xi_2(q), ..., \xi_m(q))$. The free energy associated to $\xi(q) = z$ is given by

$$A(z) = -\beta^{-1}\ln\, Z_z, \tag{4.4}$$

where Z_z denotes the partition function associated to the law μ conditioned to $\xi(q) = z$. It is defined up to an additive constant by

$$Z_z = \int_{\Sigma_z} e^{-\beta V(q)}\ (\det\, G)^{-1/2}\ d\sigma_{\Sigma_z}, \tag{4.5}$$

where G is the $m \times m$ matrix with elements $G_{i,j} = \nabla\xi_i \cdot \nabla\xi_j$, $1 \le i,j \le m$ and $d\sigma_{\Sigma_z}$ is the Lebesgue measure on

$$\Sigma_z = \{q \mid \xi(q) = z\}$$

induced by the surface measure of the ambient space equipped with the standard Euclidean scalar product. Note that the momenta p do not appear in (4.5) as they are constant on hyperplanes Σ_z and (due to the separability of the Hamiltonian (1.1)) are integrated out. The definition (4.5) is a consequence of Lemma 1 and Corollary 1, presented later in Section 4.2.

In the following, we denote

$$\delta_{\xi(q)-z}(dq) = (\det G)^{-1/2}\, d\sigma_{\Sigma_z}. \tag{4.6}$$

In the case $m = 1$, the measure simplifies to $\delta_{\xi(q)-z}(dq) = |\nabla\xi|^{-1} d\sigma_{\Sigma_z}$, where we assume that ξ is such that $|\nabla\xi| \neq 0$. The free-energy difference is then given by (4.2) where the partition functions are now integrals on the submanifolds of fixed values of ξ:

$$A(z_1) - A(z_0) = -\beta^{-1}\ln\left(\frac{\displaystyle\int_{\Sigma_{z_1}} e^{-\beta V(q)}\ \delta_{\xi(q)-z_1}(dq)}{\displaystyle\int_{\Sigma_{z_0}} e^{-\beta V(q)}\ \delta_{\xi(q)-z_0}(dq)}\right). \tag{4.7}$$

A conformational change of a macromolecule would typically be parameterized by a reaction coordinate. The reaction coordinate ξ could represent the bond or dihedral angle between a sequence of atoms or indeed the distance between two atoms of the chain

$$\xi(q) = |q_\alpha - q_\beta|,$$

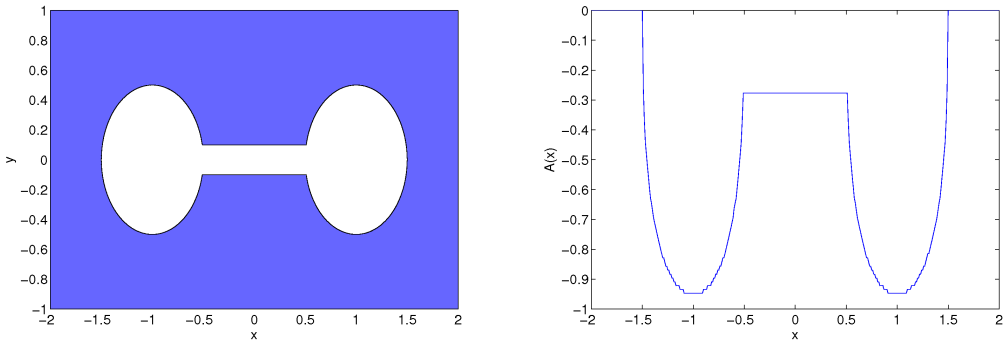
where $q_\alpha = (q_{\alpha,1}, q_{\alpha,2}, q_{\alpha,3})$ and $q_\beta = (q_{\beta,1}, q_{\beta,2}, q_{\beta,3})$ are the coordinates of atoms α and β. For covalently bonded atoms, it is common to consider instead the distance between the center of mass of two *groups* of atoms. In either case, $|\nabla\xi|$ is a non-zero constant.

The reaction coordinate may also simply be one coordinate of the system configuration $q = (q_1, ..., q_{3N})$. Without loss of generality we assume $\xi(q) = q_1$, so that $\nabla\xi = (1, 0, ..., 0)$ and $|\nabla\xi| = 1$. Then (4.7) simplifies to

$$A(z_1) - A(z_0) = -\beta^{-1}\ln\left(\frac{\displaystyle\int_{\mathbb{R}^{3N-1}} e^{-\beta V(z_1,q_2,...,q_{3N})}\ dq_2\ ...\ dq_{3N}}{\displaystyle\int_{\mathbb{R}^{3N-1}} e^{-\beta V(z_0,q_2,...,q_{3N})}\ dq_2\ ...\ dq_{3N}}\right),$$

so that the integrals are over all the *free* degrees of freedom. This choice of reaction coordinate may be relevant when studying the insertion of a protein in a membrane, where the single coordinate of interest is chosen as the z coordinate of the end-most atom entering the membrane. Needless to say this choice also makes mathematical analysis much easier, and this is the reason why it is for a demonstrative example later in this report.

Often the free energy corresponding to two different states may be equal, $\Delta A = 0$, giving little insight into the transition undergone by the system. For this reason,



(a) (b)

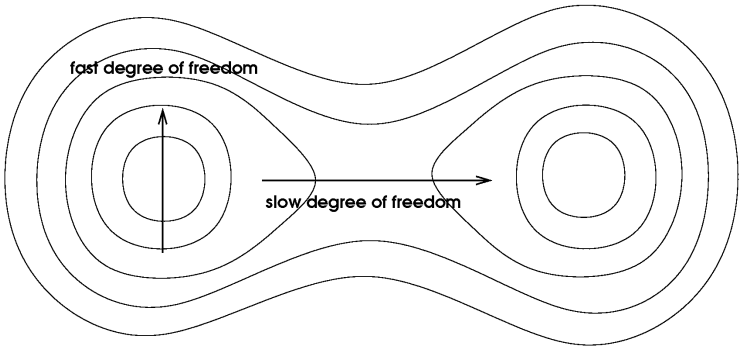
Figure 4.1: A toy example demonstrating entropic barriers. (a) The potential in the dumbell region is zero, surrounded by a wall of high energy. (b) The configurational bottleneck in the potential function is seen as a barrier in the free-energy profile.

it is informative to seek not only the total free-energy difference but its continuous evolution during the course of the transition, through the free-energy difference profile. Examining the whole profile is more revelatory in that it recovers free energy barriers encountered during the chemical reaction or conformational change. These could be due to energetic as well as entropic barriers. Entropic barriers, as depicted in Figure 4.1, refer to configurational bottlenecks, impeding transitions from one state to another.

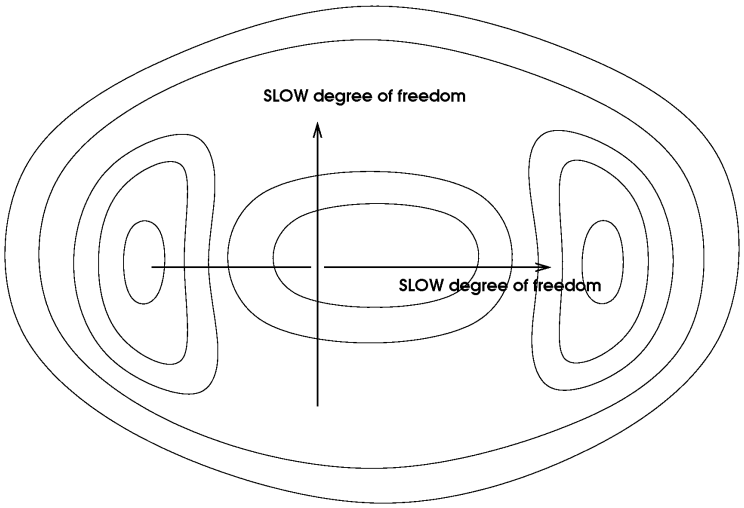
4.1.2 Choosing a good reaction coordinate

A persistent challenge in computational chemistry is that of determining a *good* or *optimal* reaction coordinate. In its simplest form, a reaction coordinate is a smooth function $\xi : \mathbb{R}^{3N} \to \mathbb{R}^m$, with $m \ll 3N$, such that (i) the initial and final states, A and B of the transition can be described well in terms of it and (ii) $\det G \neq 0$, in order for (4.5) to be well defined and in particular so that the set of submanifolds $\Sigma_z = \{q \in \mathcal{D} \mid \xi(q) = z\}$ foliates the phase space between A and B (that is, they are nonintersecting smooth dividing surfaces). From a sampling point of view, a *good* reaction coordinate is a function ξ that describes well the metastability of the underlying dynamics. In other words ξ is a good choice if the measure $\exp(-\beta(V - A \circ \xi))$ is easier to sample than $\exp(-\beta V)$. For instance, in the case where $\xi : \mathbb{R}^{3N} \to \mathbb{T}$, if $X \sim \exp(-\beta(V - A \circ \xi))$, then the law of $\xi(X)$ is uniform, thus far easier to sample.

Very often the reaction coordinate is chosen intuitively, depending on the transition taking place, but this becomes all the more difficult with increasing system size. In the one-dimensional case, $m = 1$, a good candidate for the reaction coordinate may be determined by computing the *committor function* on the configuration space [E 2005b, E 2004]. The committor function $\rho^+(q)$ is the proba-



(a) Sketch of the potential $V(x, y) = y^2 + (x^2 - 1)^2$, where the double-well feature is only visible in the direction of the variable x, the y variable is regarded as a fast degree of freedom. In this case the reaction coordinate $\xi(x, y) = x$ would be optimal.



(b) Sketch of a two-dimensional *bi-channel* potential, where metastabilities are encountered in two orthogonal directions. In this case, no one-dimensional reaction coordinate would be sufficient to describe the slow degrees of freedom.

Figure 4.2: Depending on the nature of the underlying potential energy surface, a reaction coordinate of low dimension (compared to the system size) may not always be sufficient to describe the metastabilities of the associated dynamics.

bility with which, starting from a canonically distributed initial configuration q, the system is likely to attain the final state B before A. Determining it may indeed help to compute other quantities of dynamical interest such as reaction rates [Vanden-Eijnden 2006]. The interest of the committor function as a reaction coordinate was recognized earlier and is discussed in [Dellago 2002, Bolhuis 2002a] within the framework of Transition Path Sampling. The committor function $\rho^+(q)$ is formally defined as the solution to

$$0 = -\nabla V \cdot \nabla \rho^+ + \beta^{-1} \Delta \rho^+, \quad \rho^+|_{q \in A} = 0 \text{ and } \rho^+|_{q \in B} = 1, \tag{4.8}$$

the backward Kolmogorov equation associated to the overdamped Langevin dynamics. In the two-dimensional case, the problem may be solved numerically using finite differences (see Appendix A.1 of [Metzner 2007]). For higher dimensional systems, one may resort to the String method (cf. Section 2.2.2), but typically, for most applications of practical interest, solving (4.8) is rarely possible.

It is important to note that often there are transitions for which an *optimal* one-dimensional reaction coordinate may not exist. Figure 4.2(b) shows, for instance, an example of a so-called *bi-channel* potential, where initial and final states are linked by two distinct paths, separated from each other by high energy barriers. In this particular case, no optimal one-dimensional reaction coordinate exists. Such a scenario is typical in molecular dynamics, and is a recurring theme in the works presented in Chapters 5 and 6. In this work, we will refer to any order parameter as a reaction coordinate. We describe methods for computing free energies with respect to a *given* reaction coordinate; we do not attempt to find an optimal one.

4.2 Computing free energy differences

4.2.1 Thermodynamic integration

As previously mentioned, practitioners are often more interested in free-energy differences (4.2) rather than absolute free energies (4.1). Fortunately this has an advantage from a mathematical viewpoint. Free-energy differences may be computed by calculating the derivative of the free energy A' and then numerically integrating it:

$$\Delta A = A(1) - A(0) = \int_0^1 A'(\lambda)\, d\lambda. \tag{4.9}$$

This is the idea of *thermodynamic integration* (TI), one of the earliest methods for free energy computation, dating back to [Kirkwood 1935]. The derivative A' is computed at fixed indexing values α or ξ (depending on the choice of parameterization) as presented in (4.10) and (4.15) below. For each indexing value, the canonical averaging is performed by running a relevant dynamics ergodic with respect to the appropriate measure (μ_α or $\mu(\cdot|z)$). Though the idea is the same in either setting, for the sake of mathematical precision a distinction between the two is made.

Thermodynamic integration for alchemical transformations

In the case of an alchemical transformation, the derivative of the free energy can be expressed as the canonical average of the derivative of the Hamiltonian with respect to the parameter α:

$$A'(\alpha) = \frac{\displaystyle\int_\Omega \frac{\partial H_\alpha}{\partial \alpha} e^{-\beta H_\alpha(q,p)}\, dq\, dp}{\displaystyle\int_\Omega e^{-\beta H_\alpha(q,p)}\, dq\, dp} = \left\langle \frac{\partial H_\alpha}{\partial \alpha} \right\rangle_{\mu_\alpha}, \tag{4.10}$$

with $d\mu_\alpha(q,p) = Z_\alpha^{-1} e^{-\beta H_\alpha(q,p)}\, dq\, dp$ and $\langle \cdot \rangle_{\mu_\alpha}$ denoting the average with respect to the measure μ_α. In practice this is done by evaluating A' at discrete values in the parameter range $[0,1]$. One may typically choose to divide the parameter range into n equally spaced intervals, and evaluate A' at points $\alpha_i = i/n$. Finally the free-energy difference may be computed by the Riemann sum:

$$\Delta A = \frac{1}{n} \sum_{i=0}^{n-1} A'(\alpha_i). \tag{4.11}$$

Thermodynamic integration for the reaction coordinate case

The reaction coordinate case needs a little more care due to the delta function term in the integrand (see (4.4), (4.5) and (4.6)). An important formula, *the co-area formula* will be needed to treat this case.

Lemma 1 ***(Co-area formula).*** *For any smooth function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\int_{\mathbb{R}^n} f(q) (\det G)^{1/2}(q) dq = \int_{\mathbb{R}^m} \int_{\Sigma_z} f d\sigma_{\Sigma_z} dz,$$

where G is the $m \times m$ matrix with elements $G_{i,j} = \nabla \xi_i \cdot \nabla \xi_j$, $1 \le i,j \le m$ and σ_{Σ_z} denotes the surface measure on $\Sigma_z = \{q \in \mathcal{D} \mid \xi(q) = z\}$, the Lebesgue measure on Σ_z induced by the Lebesgue measure in the ambient Euclidean space $\mathbb{R}^n$.

Finally, the following result, the proof of which depends on the lemma above (see Section 3.2.1 of [Lelièvre 2010b]), allows us to formally introduce the marginal and conditional measures referred to in the sequel.

Corollary 1 *If a continuous random variable X has law $\psi(q)\, dq$ in $\mathbb{R}^{3N}$, then $\xi(X)$ has law*

$$\left(\int_{\Sigma_z} \psi (\det G)^{-1/2} d\sigma_{\Sigma_z} \right) dz,$$

and the law of X conditioned to a fixed value z of $\xi(X)$ is

$$\frac{\psi\ (\det G)^{-1/2} d\sigma_{\Sigma_z}}{\displaystyle\int_{\Sigma_z} \psi\ (\det G)^{-1/2} d\sigma_{\Sigma_z}}.$$

In particular we have the equilibrium marginal measure in ξ

$$\mu^{\xi}(dz) = \left(\int_{\Sigma_z} e^{-\beta V(q)} \ \delta_{\xi(q)-z}(dq) \right) \ dz, \tag{4.12}$$

where $\delta_{\xi(q)-z}(dq) = (\det G)^{-1/2} d\sigma_{\Sigma_z}$, *and the equilibrium conditional measure is*

$$\mu(dq|z) = \frac{e^{-\beta V(q)} \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} e^{-\beta V(q)} \ \delta_{\xi(q)-z}(dq)}. \tag{4.13}$$

The derivative of the free energy $\nabla A(z)$ in the general case $m > 1$ has components $\partial_{z_\alpha} A$, for $\alpha = 1, ..., m$, given by the conditional canonical average

$$\partial_{z_\alpha} A(z) = \int_{\Sigma_z} \left(\sum_{\gamma=1}^{m} G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \cdot \nabla V - \beta^{-1} \mathrm{div} \left(G^{-1}_{\alpha,\gamma} \nabla \xi_\gamma \right) \right) \mu(dq|z), \tag{4.14}$$

where $G^{-1}_{\alpha,\gamma}$ is the (α, γ)-component of the inverse matrix G^{-1}. Whenever $m = 1$, which will be the case treated in the following chapters, one has

$$A'(z) = \int_{\Sigma_z} f^V(q) \ d\mu(q|z) = \mathbb{E}_\mu \left[f^V | \xi(q) = z \right], \tag{4.15}$$

where

$$f^V = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \nabla \cdot \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right). \tag{4.16}$$

The derivation of (4.16) can be found in [Ciccotti 2008, Sprik 1998, den Otter 1998]. The first term is the projection of the force onto the direction of the order parameter and the second term accounts for the change of variables. For this reason the term f^V is called the *local mean force* and the derivative of the free energy A' with respect to the order parameter is called the *mean force*. Notice that in the simple case $\xi(q) = q_1$, one has $f^V = \partial_{q_1} V$.

Remark 3 *In some literature the local mean force* (4.16) *is written as*

$$f^V = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \nabla \cdot \left(\frac{\partial \ln |J|}{\partial \xi} \right), \tag{4.17}$$

where $J = \frac{\partial(q_1, q_2 ..., q_{3N})}{\partial(\xi, q'_2, ..., q'_{3N})}$ *is the Jacobian matrix arising from a global change of variables from Cartesian to the new set of coordinates (containing* ξ *as an independent variable) and* $|J|$ *is its determinant. In order to compute the second term above, the Cartesian coordinates need to be expressed explicitly in terms of the generalized coordinates (new coordinates including* ξ*), which is very often not trivial. The equivalent definition of the local mean force* (4.16) *depends only on the derivatives of* ξ *with respect to the Cartesian coordinates* q*, thus easier to calculate. For the equivalence of* (4.17) *and* (4.16) *(in fact in its more general form), see [Darve 2006].*

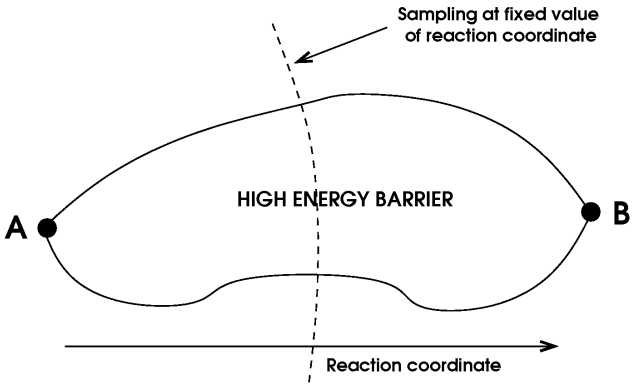


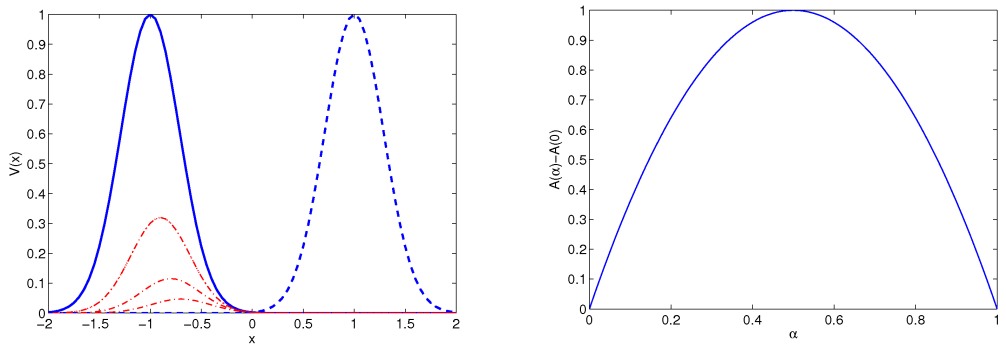
Figure 4.3: When a transition is described by a reaction coordinate, constrained sampling such as thermodynamic integration (TI) or blue moon sampling face problems when metastable regions exist at fixed values of the reaction coordinate. In this figure, two different paths linking A and B are separated by a high energy barrier, thus inhibiting the fast sampling in each bin.

The mean force A' is indeed easier to calculate than the free energy A as it is a conditional expectation and may be estimated through MD simulations using methods discussed in Section 1.2. Sampling may be achieved through unconstrained dynamics, by dividing the reaction coordinate interval into a finite number of bins. It may also be accomplished at fixed values of ξ by sampling the Blue Moon ensemble [Carter 1989, Ciccotti 2005, Ciccotti 2008, Lelièvre 2010c]. Blue Moon sampling is a constrained sampling technique, whereby the canonical average (4.15) is computed at a fixed ξ by sampling over the equilibrium distribution of the system restricted to the hypersurface Σ_z.

Once the mean force is computed, the free energy difference may be calculated once again by the Riemann sum

$$\Delta A = \sum_{i=0}^{n-1} A'(z_i)\Delta z. \tag{4.18}$$

One advantage of this method is that in practice each component in the sum (4.18) may be calculated separately through independent simulations, thus may be straightforwardly parallelized. However, as for all constrained dynamics, this method suffers from slow convergence in the case of "coupled slow degrees of freedom", as depicted in Figure 4.3. Supposing there exist two paths from A to B, separated by a high energy barrier, then the time needed to sample the space on both sides of the energy barrier via constrained dynamics scales exponentially with the barrier height (cf. Arrhenius' law (1.31)). Therefore, if the energy barrier is large enough, we face a problem of quasi-ergodicity: the constrained dynamics runs the risk of not sampling at all some important regions of the configuration space.



(a) (b)

Figure 4.4: An example of free energy perturbation (FEP) for computing free-energy difference profiles. (a) Thick lines represent probability distribution functions corresponding to potential energies $V_0 = (x+1)^2$ and $V_1 = (x-1)^2$. The dash-dotted lines correspond to distributions of the interpolated potentials $V_\alpha = (1-\alpha)V_0 + \alpha V_1$ for $\alpha = 0.05, 0.1$ and 0.15. (b) The free energy difference is calculated according to (4.20).

4.2.2 Other equilibrium free energy methods

We detail below some other popular methods for free energy computation, based on equilibrium sampling.

Free energy perturbation

The free energy perturbation (FEP) technique [Zwanzig 1954], developed some twenty years after TI, is an approach for computing free-energy differences in the alchemical setting (4.3). The underlying idea is to avoid computing two partition functions and instead compute the average of a perturbed function with respect to a reference probability measure. In this case the reference probability measure is $d\mu_0 = Z_0^{-1} e^{-\beta H_0(q,p)}\, dq\, dp$, and (4.3) is rewritten as

$$A_1 - A_0 = -\beta^{-1} \ln \left(\int_\Omega e^{-\beta(H_1 - H_0)(q,p)}\, d\mu_0(q,p) \right). \tag{4.19}$$

In practice, the integral in the above is approximated by a sampling procedure of μ_0, to approximate

$$\mathbb{E}_{\mu_0} \left[e^{-\beta(H_1 - H_0)} \right].$$

Clearly, for the variance of the estimator to be small, the Hamiltonians H_0 and H_1 should be sufficiently close: they need to overlap in some regions of the domain. When this is not immediately the case, intermediate Hamiltonians are constructed corresponding to discrete parameterization points $\alpha_i \in [0,1]$. Figure 4.4(a) gives an

example of such a situation, where canonical measures corresponding to two potentials V_0 (left) and V_1 (right) do not share any common regions of high probability. Equation (4.19) then may be approximated by a Riemann sum:

$$\begin{aligned} \ln\left(\frac{Z_1}{Z_0}\right) &= \ln \int_\Omega e^{-\beta(H_1-H_0)(q,p)}\, d\mu_0(q,p) \\ &= \sum_{i=0}^{n-1} \ln \int_\Omega e^{-\beta(H_{\alpha_{i+1}}-H_{\alpha_i})(q,p)}\, d\mu_{\alpha_i}(q,p) \qquad (4.20) \\ &= \sum_{i=0}^{n-1} \ln\left(\mathbb{E}_{\mu_{\alpha_i}}\left[e^{-\beta(H_{\alpha_{i+1}}-H_{\alpha_i})}\right]\right) \qquad (4.21) \end{aligned}$$

where typically $\alpha_i = i/n$ and

$$H_{\alpha_i} = (1-\alpha_i)H_0 + \alpha_i H_1.$$

Importance sampling

For both transformations parameterized by a reaction coordinate and alchemical transformations, free-energy difference computations involve computing an average of an observable with respect to a given probability measure (cf. (4.15) and (4.21)). As previously mentioned, the multimodality of the given measures often lead to metastability of the associated dynamics, thus leading to slow convergence. Importance sampling is a variance reduction technique helping to improve sampling of a given probability measure. This is achieved by first sampling a biased distribution, where the exploration of 'important' regions of the state space is encouraged. Next, the observable (simulation output) is weighted, in order to unbias the estimator. To illustrate this, suppose that there exists a bounded function U such that the measure μ_U with density proportional to $e^{-\beta(V-U)}$ is easier to sample than μ (with density proportional to $e^{-\beta V}$). That is, the potential $V-U$ possesses fewer metastable states than V. The idea is then to write the canonical average of some function $\Phi \in L^1(\mu)$ as

$$\begin{aligned} \mathbb{E}_\mu[\Phi] = \int_\Omega \Phi(q) Z^{-1} e^{-\beta V(q)}\, dq &= \frac{\displaystyle\int_\Omega \Phi(q) e^{-\beta U(q)} e^{-\beta(V-U)(q)}\, dq}{\displaystyle\int_\Omega e^{-\beta U(q)} e^{-\beta(V-U)(q)}\, dq} \\ &= \frac{\mathbb{E}_{\mu_U}\left[\Phi\, e^{-\beta U}\right]}{\mathbb{E}_{\mu_U}\left[e^{-\beta U}\right]}. \end{aligned}$$

For a good candidate U to be found, it is important to know the nature of the underlying potential energy surface V. This is, unfortunately, rarely possible. Ideally, the metastable features of $e^{-\beta V}$ should be captured and eliminated when biasing the measure with $e^{\beta U}$.

When the low-dimensional reaction coordinate ξ describes well the metastable features of the probability distribution, then $U = A \circ \xi$ may indeed be a good candidate for a biasing potential, where A is the free energy as defined in (4.4). The free energy is therefore not only a quantity of thermodynamic interest but useful as a biasing potential to speed up sampling of state spaces (see [Chopin 2010] for an application in a field other than molecular dynamics). The dynamics which is ergodic with respect to the biased measure $e^{-\beta(V-A\circ\xi)}\ dq$ is defined simply by replacing the potential V in the overdampled Langevin dynamics (1.29) by $\mathcal{V} = V - A \circ \xi$, which yields

$$dQ_t = -\nabla(V - A \circ \xi)(Q_t)\ dt + \sqrt{2\beta^{-1}} dW_t.$$

Then if $\psi(t, \cdot)$ is the density of the distribution of Q_t, the equilibrium measure will be the density $\psi_\infty = Z_A^{-1} e^{-\beta(V-A\circ\xi)}$.

Of course, the free energy A is normally not known *a priori*; its computation relies itself on efficient sampling techniques. It will be shown later, in Section 4.3, through *adaptive* importance sampling techniques, how an *on-the-fly* estimate of the free energy may act as a biasing potential. The success of these methods relies primarily on a good choice of the reaction coordinate.

Histogram methods

Histogram methods for free energy computations are applicable in the case where a transition is parameterized by a reaction coordinate. Recall that the probability density of finding a system in macrostate $\xi(q) = z$ is

$$\rho(z) = \frac{\displaystyle\int_{\Sigma_z} e^{-\beta V(q)} \delta_{\xi(q)-z}(dq)}{\displaystyle\int_\Omega e^{-\beta V(q)}\ dq}. \tag{4.22}$$

The free energy difference ΔA can then be written in terms of ρ as (see (4.7))

$$\Delta A = A(z_1) - A(z_0) = -\beta^{-1} \ln\left(\frac{\rho(z_1)}{\rho(z_0)}\right).$$

The simplest of histogram methods consists in (i) discretizing the reaction coordinate space into bins of width Δz; (ii) sampling the canonical measure μ using *unconstrained* sampling; (iii) measuring the number of times $f(z)$ a system visits the bin $[z - \Delta z/2, z + \Delta z/2]$:

$$f(z) = \mathbb{E}_\mu\left[\mathbf{1}_{\ |\xi(q)-z|\le \Delta z/2}\ \right],$$

and (iv) estimating the probability (4.22) by

$$\rho(z) \approx \frac{f(z)}{\Delta z \displaystyle\sum_{z'} f(z')}.$$

In other words one measures the proportion of time a certain macrostate $\xi(q) = z$ is visited compared to the whole reaction coordinate range. The free-energy difference may in this way be approximated as

$$A(z_1) - A(z_0) = -\beta^{-1} \ln \left(\frac{f(z_1)}{f(z_0)} \right).$$

This method is often very inefficient, however, due to metastable features of the unconstrained dynamics sampling μ. In practice one computes averages with respect to modified measures $\mu_i \propto e^{-\beta V_i}$, where

$$V_i(q) = V(q) + \frac{1}{2\varepsilon_i} \left(\xi(q) - z_i \right)^2.$$

The second term on the right-hand side is a confining potential, restricting the system to remain within a close vicinity of the reaction coordinate $\xi(q) = z_i$ and $\varepsilon_i > 0$ is a small parameter determining the width of the confining potential. This set of 'local' averages with respect to μ_i may then be assembled together to obtain averages with respect to the desired measure μ through techniques such as *(multistate) BAR* ((M)BAR) [Bennett 1976, Shirts 2008].

4.2.3 Nonequilibrium free energy methods

The methods presented thus far have all relied on equilibrium sampling. Nonequilibrium techniques, due to Jarzynski [Jarzynski 1997], also exist for free energy computations. The idea of the method is presented below only for the case of Hamiltonian dynamics (for a more in-depth introduction to nonequilibrium free energy methods, in particular for Langevin dynamics, see Chapter 4 of [Lelièvre 2010b]).

Jarzynski equality

We begin by choosing a transition schedule $\alpha(t)$ such that $\alpha(0) = 0$ and $\alpha(T) = 1$, and assume the system follows the Hamiltonian dynamics

$$\begin{cases} \dfrac{dq_t}{dt} = \nabla_p H_{\alpha(t)}(q_t, p_t), \\ \dfrac{dp_t}{dt} = -\nabla_q H_{\alpha(t)}(q_t, p_t), \end{cases} \tag{4.23}$$

with initial conditions $(q_0, p_0) = (q, p) \sim \mu_0$. Let us now define the associated flow by Λ_t, so that at time t, $\Lambda_t(q, p) = (q_t, p_t)$. Then it can be shown

$$H_1(\Lambda_T(q,p)) - H_0(q,p) = \int_0^T \alpha'(t) \frac{\partial H_{\alpha(t)}(\Lambda_t(q,p))}{\partial \alpha} \, dt, \tag{4.24}$$

since

$$\frac{d}{dt} \left(H_{\alpha(t)}(\Lambda_t(q,p)) \right) = \frac{\partial H_{\alpha(t)}}{\partial \alpha}(\Lambda_t(q,p))\alpha'(t) + \begin{pmatrix} \nabla_q H_{\alpha(t)}(\Lambda_t(q,p)) \\ \nabla_p H_{\alpha(t)}(\Lambda_t(q,p)) \end{pmatrix} \cdot \partial_t \Lambda_t(q,p),$$

where the second term vanishes in view of (4.23). The first term can be interpreted as the force $(\partial H_\alpha/\partial\alpha)$ applied to the system, multiplied by the distance moved along α in time dt. The integrand in (4.24) is therefore the total work done in time T to move the system from phase 0 to 1

$$W_T = \int_0^T \alpha'(t) \frac{\partial H_\alpha(\Lambda_t(q,p))}{\partial\alpha}\, dt. \tag{4.25}$$

Using the above and (4.24), we have

$$\begin{aligned} \mathbb{E}_{\mu_0}\left[e^{-\beta W_T}\right] &= \int_\Omega e^{-\beta(H_1(\Lambda_T(q,p))-H_0(q,p))} d\mu_0(q,p) \\ &= Z_0^{-1} \int_\Omega e^{-\beta H_1(\Lambda_T(q,p))}\, dq\, dp, \end{aligned}$$

since we recall $d\mu_0(q,p) = Z_0^{-1} e^{-\beta H_0(q,p)}\, dq\, dp$. Furthermore, as Λ_T defines a change of variables of Jacobian determinant 1 and by the definition of the free energy difference (4.2), we have

$$\mathbb{E}_{\mu_0}\left[e^{-\beta W_T}\right] = \frac{Z_1}{Z_0} = e^{-\beta\Delta A}.$$

Depending on the scheduling $\alpha(t)$, the lowest values of the work can quickly dominate the average, giving rise to similar numerical issues as in free energy perturbation discussed previously.

4.3 Adaptive sampling methods

Adaptive methods provide a tool to circumvent difficulties arising in standard importance sampling methods, namely that of choosing *a priori* a good biasing potential. The key idea behind adaptive methods is to bias the potential driving the dynamics by a time-dependent function $A_t \circ \xi$ and to update A_t so that it converges to the free energy A. That is, the dynamics is governed by the biased potential

$$\mathcal{V}_t(q) = V(q) - (A_t \circ \xi)(q) \tag{4.26}$$

and the corresponding dynamics is then the overdamped Langevin dynamics (1.29) with a modified drift term:

$$dQ_t = -\nabla(V - A_t \circ \xi)(Q_t)dt + \sqrt{2\beta^{-1}} dW_t. \tag{4.27}$$

Recall that if ξ "described the metastabilities of the potential V" (see Figure 4.2(a)), then the biased dynamics with $A_t = A$ would eliminate them. The idea is therefore to update the adaptive bias A_t such that it converges to the free energy A as efficiently as possible. This may be done in two different manners.

(i) The biasing potential A_t itself is updated during the simulation in such a way that it converges to A up to some additive constant. Methods of this type are called Adaptive Biasing Potential (ABP) methods, and include famous instances such as the Wang-Landau algorithm [Wang 2001b] and metadynamics [Iannuzzi 2003]. This class of methods is discussed in Section 4.3.1.

(ii) The derivative of the bias with respect to the reaction coordinate A'_t is updated. This approach, known as the Adaptive Biasing Force (ABF) [Darve 2001] method, is presented in Section 4.3.2 and is the focus of the last two chapters of this thesis.

Note that, even though both ABF and ABP approaches give $A'_t \to A'$, their associated dynamics are not equivalent (see Remark 5 further on).

Remark 4 *Note that for any biasing force that has any longtime limit $A'_t \to A'_\infty$, the average of the local mean force f^V with respect to the equilibrium measure $d\mu_\infty = Z_\infty^{-1} e^{-\mathcal{V}_\infty}\, dq\, dp$ gives the desired mean force, irrespective of the function A'_∞. More precisely, we have*

$$\mathbb{E}_{\mu_\infty}\left[f^V | \xi(q) = z\right] = \frac{\displaystyle\int_{\Sigma_z} f^V \; e^{-\beta(V - A_\infty \circ \xi)(q)} \; \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} e^{-\beta(V - A_\infty \circ \xi)(q)} \; \delta_{\xi(q)-z}(dq)} \tag{4.28}$$

$$= \mathbb{E}_{\mu}\left[f^V | \xi(q) = z\right] = A'(z), \tag{4.29}$$

where the terms in $A_\infty(z)$ are eliminated from the numerator and denominator as they are constant on the submanifold Σ_z. The biasing function A_t serves only as a means to push systems out of metastable regions along the reaction coordinate and does not affect the computation of the mean force.

4.3.1 Adaptive Biasing Potential

Adaptive biasing potential (ABP) methods update directly the biasing potential A_t in such a way that it converges to A *up to an additive constant*. To motivate the ideas behind the algorithms of this class, we begin by considering the case where the system is instantaneously at equilibrium. That is, if $\psi(t, \cdot)$ is the density of the distribution of Q_t, the solution to the adaptive dynamics (4.27), we consider $\psi(t, \cdot) = \psi^{\mathrm{eq}}(t, \cdot) = Z_t^{-1} e^{-\beta(V - A_t \circ \xi)}$. The observed free energy at equilibrium is then given by

$$A_t^{\mathrm{eq}}(z) = -\beta^{-1} \ln \int_{\Sigma_z} \psi_t^{\mathrm{eq}} \delta_{\xi(q)-z}(dq) \tag{4.30}$$

$$= A(z) - A_t(z) + \beta^{-1} \ln Z_t.$$

In this case, using a fixed point strategy (see [Lelièvre 2008]), a general update formula for A_t would be

$$\frac{dA_t}{dt} = \mathcal{F}_t\left(A_t^{\mathrm{eq}}(z)\right), \tag{4.31}$$

where $(\mathcal{F}_t)_{t\geq 0}$ is a set of strictly increasing functions. Of course we are typically not instantaneously at equilibrium, but these results may be used as a guideline for the developing a method of the ABP class. See Chapter 5 of [Lelièvre 2010b] for further detail and in particular the multi-dimensional case: $m > 1$.

Let us first consider the simplest (time-independent) choice $\mathcal{F}_t(x) = x$ in (4.31). Furthermore, by replacing the equilibrium density in (4.30) by the density $\psi(t,\cdot)$ of Q_t, we obtain the *non-equilibrium* update scheme :

$$\frac{dA_t(z)}{dt} = -\beta^{-1}\ln\left(\int_{\Sigma_z} \psi(t,q)\ \delta_{\xi(q)-z}(dq)\right), \tag{4.32}$$

where we recall $\psi(t,q)$ is the density of the distribution of Q_t. Using this, a typical ABP dynamics is given by

$$\begin{cases} dQ_t = -\nabla(V - A_t\circ\xi)(Q_t)dt + \sqrt{2\beta^{-1}}dB_t, \\ A_t(z) = \displaystyle\int_0^t -\beta^{-1}\ln\left(\int_{\Sigma_z} \psi(s,q)\ \delta_{\xi(q)-z}(dq)\right)\ ds, \end{cases} \tag{4.33}$$

where $\psi(t,q)\ dq$ is the law of Q_t. In other words, the underlying potential V is increased in well-sampled regions of the reaction coordinate, encouraging exploration of under-sampled regions. This technique, however, does have a drawback: the biasing potential A_t never formally converges. Even when the process Q_t is at equilibrium ($\psi(t,\cdot) = \psi_\infty$), A_t is likely to continue to evolve.

The Wang-Landau method [Wang 2001b, Wang 2001a], based on a flat-histogram technique, is a popular ABP method that updates A_t in such a way that it *does* eventually converge. The Wang-Landau method uses the update function

$$\mathcal{F}_t(x) = -\gamma(t)\ \exp(-\beta x), \tag{4.34}$$

where $\gamma(t)$ is a positive, decreasing function of time with $\gamma(0) = 1$. The biasing potential is updated as

$$\frac{dA_t(z)}{dt} = \gamma(t)\int_{\Sigma_z} \psi(t,q)\ \delta_{\xi(q)-z}\ dq.$$

Here the vanishing parameter $\gamma(t)$ is essential to the convergence of $A_t(z)$. The update strategy for parameter $\gamma(t)$ is delicate. It must not decrease too quickly otherwise A_t would converge prematurely (and very often incorrectly); neither must it converge too slowly, leading to slow convergence of the algorithm.

Another method to be cited is the nonequilibrium Metadynamics [Laio 2002, Bussi 2006], where a similar update function (4.34) is used to bias a newly-introduced external variable. This method amounts to adding small gaussian potentials on the reaction coordinate space to bias the dynamics. In this way, the method introduces another parameter, namely the width (or variance) of the gaussian potential.

The bias A_t in the Wang-Landau and Metadynamics methods does indeed converge, however it does so *artificially* through the decreasing function $\gamma(t)$. The

latter method introduces a further parameter, the width of the gaussian, which also remains to be chosen. An alternative method based on the self-healing umbrella sampling method [Marsili 2006], using a 'mollified' free energy, has been developed [Dickson 2010] whereby the bias A_t converges without the need for an artificially vanishing parameter. This is achieved thanks to an additional parameter that the authors introduce to unbias the free-energy estimator on the fly.

4.3.2 Adaptive Biasing Force

The Adaptive Biasing Force (ABF) method [Darve 2001, Hénin 2004] updates the derivative with respect to z of the biasing potential $A_t(z)$. The function A'_t is updated in such a way as to estimate the mean force A' given by (4.15), and is used to bias the dynamics in the direction of ξ. To motivate the choice of A'_t, in the same spirit as for the ABP methods, we consider a system instantaneously at equilibrium. The observed mean force when $\psi(t,\cdot) = \psi^{\mathrm{eq}}(t,\cdot) = Z_t^{-1} e^{-\beta(V - A_t \circ \xi)}$ is

$$\begin{aligned} \Gamma_t^{\mathrm{eq}}(z) &= \frac{\displaystyle\int_{\Sigma_z} f^V e^{-\beta(V - A_t \circ \xi)}\ \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} e^{-\beta(V - A_t \circ \xi)}\ \delta_{\xi(q)-z}(dq)} \\ &= A'(z). \end{aligned} \tag{4.35}$$

So by choosing

$$A'_t(z) = \Gamma_t^{\mathrm{eq}}(z),$$

the result $A'_t = A'$ is immediate. In the nonequilibrium case, the density $\psi^{\mathrm{eq}}(t,\cdot)$ in (4.35) is replaced with ψ

$$A'_t(z) \approx \frac{\displaystyle\int_{\Sigma_z} f^V(q)\ \psi(t,q)\ \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} \psi(t,q)\ \delta_{\xi(q)-z}(dq)}. \tag{4.36}$$

By denoting the measure conditioned to the submanifold Σ_z by

$$\psi(t, dq|z) = \frac{\psi(t,q)\ \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} \psi(t,q)\ \delta_{\xi(q)-z}(dq)},$$

the ABF dynamics may be written as

$$\begin{cases} dQ_t = -\nabla(V - A_t \circ \xi)(Q_t)dt + \sqrt{2\beta^{-1}}dB_t, \\ A'_t(z) = \displaystyle\int_{\Sigma_z} f^V(q)\ \psi(t, dq|z) = \mathbb{E}\left[f^V\,|\xi(Q_t) = z\right]. \end{cases} \tag{4.37}$$

The conditional expectation in (4.37) may be calculated differently depending on the implementation of the method, see (4.38) and (4.39) below.

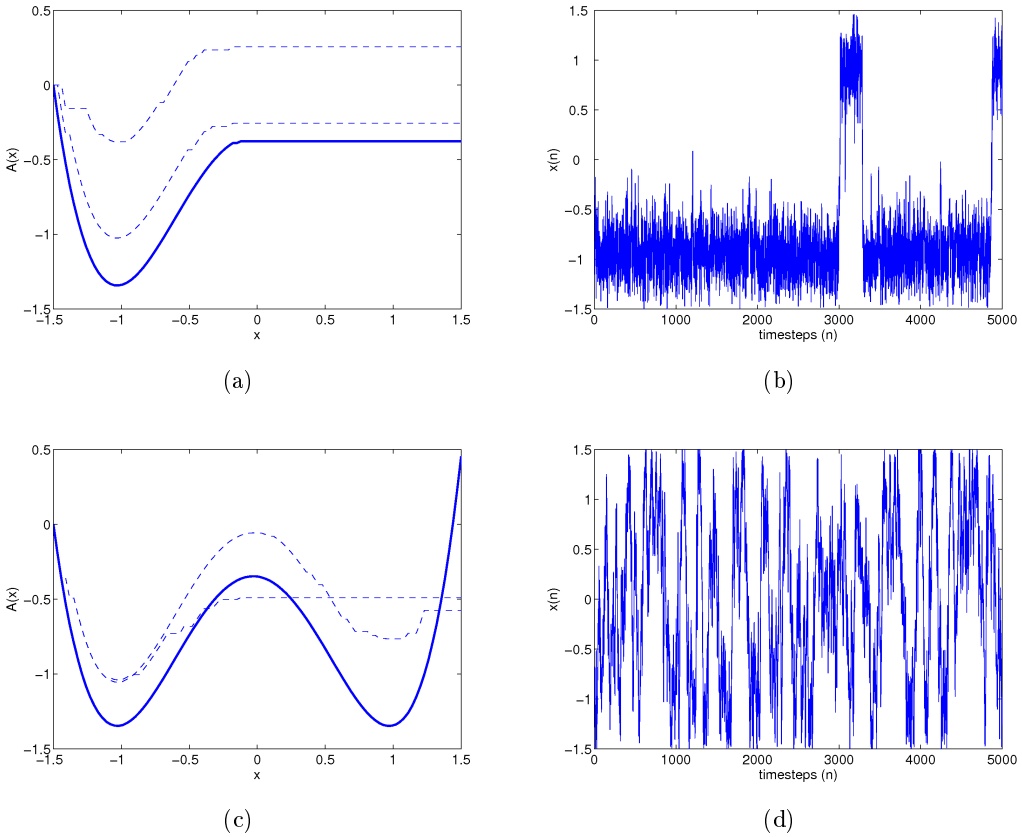


Figure 4.5: Comparing standard overdamped Langevin dynamics (1.29) and ABF dynamics (4.37). We use the potential $V(x,y) = y^2 + (x^2-1)^2$ as sketched in Figure 4.2(a), $\xi(x,y) = x$ and discretized step size $\Delta t = 0.1$. (a) Overdamped dynamics: the estimated potential of mean force profile at timesteps $n = 50, 200$ and 1000. (b) The reaction coordinate space is sampled very slowly. (c) ABF dynamics: the estimated potential of mean force profile at timesteps $n = 50, 200$ and 1000. (d) Metastability is eliminated by the biased dynamics.

Remark 5 *Notice that in both the ABP and ABF dynamics, (4.33) and (4.37), one only needs to evaluate* $\nabla(A_t \circ \xi) = (A_t' \circ \xi)\nabla\xi$, *therefore only the derivative* A_t' *is used by the dynamics. However, by differentiating the biasing potential as computed in (4.33), it is clear that it does not match* A_t' *in (4.37), due to the fact that the distribution density* $\psi(t,\cdot)$ *is not* $\psi^{\mathrm{eq}}(t,\cdot)$. *Thus, the two dynamics are indeed different.*

In standard computations, the conditional expectation is computed as a trajectorial average. The distribution ψ in (4.37) is approximated by a time average

$$\psi(t,q)\, dq \approx \frac{1}{T}\int_0^T \delta_{q-X_s}(dq)\, ds. \tag{4.38}$$

In the case of $R > 1$ walkers $(Q_t^i)_{i=0,..,R-1}$ running in parallel, as proposed in [Lelièvre 2007b], we use in practice a distribution computed over time and walkers:

$$\psi(t,q)\ dq \approx \frac{1}{R}\sum_{i=0}^{R-1}\frac{1}{T}\int_0^T \delta_{q-Q_s^i}(dq)\ ds. \tag{4.39}$$

This implementation was later called *multiple-walker ABF* (MW-ABF) and was successfully tested on the reversible folding of the deca-alanine peptide using the molecular dynamics code NAMD [Minoukadeh 2010]. This parallel implementation improved convergence of the ABF method, not only at constant wall time but at constant CPU time (see [Minoukadeh 2010] and Chapter 5). The speed-up was in particular due to MW-ABF helping to resolve issues of local metastability depicted in Figure 4.3. Results are presented in Chapter 5.

4.4 Longtime convergence of dynamics

In this section, we address an important issue that has so far been set aside: the rate of convergence to equilibrium of the dynamics presented. We study in particular the convergence *in distribution* of Q_t, following the Adaptive Biasing Force dynamics (4.37). We begin, in Section 4.4.1, by recalling some well-known results, which will be essential in defining 'distances' between probability measures and thus studying their convergence to equilibrium. In Section 4.4.2, we study the convergence of the overdamped Langevin dynamics and justify theoretically why metastabilities in the potential energy function lead to slow convergence. Finally, in Section 4.4.3, we briefly present existing convergence results, which prove sub-optimal in the particular bi-channel scenario (cf. Figure 4.2(b)).

4.4.1 Entropy and the Logarithmic Sobolev Inequality

Let us recall some well-known results for defining 'distances' between two probability measures through notions of relative entropy and related quantities. The reader is referred to the excellent books by the research group in Toulouse [Ané 2000] (written in French) and Cédric Villani [Villani 2003] for a comprehensive introduction to this topic.

Definition 1 ***(Entropy).*** *For any two probability measures μ and ν such that μ is absolutely continuous with respect to ν (denoted as $\mu \ll \nu$), the relative entropy is defined as*

$$H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu.$$

In fact for two probability measures μ and ν such that $\mu \ll \nu$, $H(\mu|\nu) \geq 0$. The non-negativity can be shown by writing $H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu + \int \left(\frac{d\nu}{d\mu}\right)\ d\mu - \int d\mu$ and using the inequality

$$\ln(1/x) + x - 1 \geq 0.$$

Furthermore, $H(\mu|\nu) = 0$ if and only if $\mu = \nu$. Although the relative entropy has these important properties, it is not strictly speaking a distance, as it fails to satisfy symmetry and the triangle inequality. It is nevertheless a useful functional as it bounds from above the total variation norm of the difference between two probability measures

$$\|\mu - \nu\|_{\mathrm{TV}} = \int \left| \frac{d\mu}{d\nu} - 1 \right| d\nu, \tag{4.40}$$

as given by Theorem 1 below. Note that since we assume that the measures have densities with respect to the Lebesgue measure, this is just the L^1 norm of the difference of their densities.

Theorem 1 ***(Csiszar-Kullback inequality).*** *For measures μ and ν which have densities with respect to the Lebesgue measure, the following holds*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \sqrt{2H(\mu|\nu)}.$$

The proof of this theorem is straightforward using (4.40), the Pinsker inequality $3(u-1)^2 \leq (2u+4)(u \ln u - u + 1)$, $\forall u \geq 0$ and the Cauchy-Schwarz inequality. The theorem (stated in its more general form) and its proof may be found in [Ané 2000].

Let us now define an important quantity, the Fisher information, that will be essential in studying longtime convergence.

Definition 2 ***(Fisher information).*** *For any probability measure μ absolutely continuous with respect to ν, the Fisher information is given by*

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

Definition 3 ***(Logarithmic Sobolev inequality).*** *The probability measure ν is said to satisfy a logarithmic Sobolev inequality with constant $\rho > 0$ (in short: LSI(ρ)) if for all probability measures μ such that $\mu \ll \nu$,*

$$H(\mu|\nu) \leq \frac{1}{2\rho} F(\mu|\nu).$$

With a slight abuse of language, in the following, we may say that a measure or its density (with respect to the Lebesgue measure) satisfies an LSI.

Logarithmic Sobolev Inequalities (LSIs) are largely due to the pioneering work by Gross [Gross 1975], who first showed it for Gaussian measures. Later, works by Bobkov and Götze [Bobkov 1999] provided a necessary and sufficient condition for a *one-dimensional* probability measure to satisfy an LSI. According to their criterion, a function $V(x)$ that behaves as $|x|^\alpha$ for $|x| \to \infty$ has an associated Gibbs probability density $e^{-V(x)}$, which satisfies an LSI if and only if $\alpha \geq 2$. Therefore we typically expect measures that decay very fast as $|x| \to \infty$ to satisfy an LSI. Such a criterion does not hold in general in a higher-dimensional case. However thanks to the work of Bakry and Emery [Bakry 1984], and Holley and Stroock [Holley 1987], logarithmic

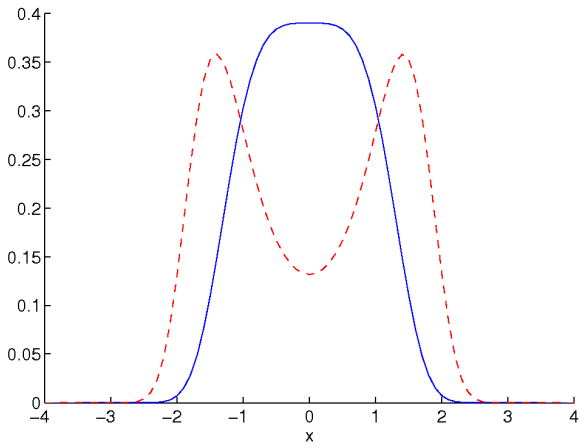


Figure 4.6: The probability densities $\phi_1(x) \propto \exp\left(-\frac{\beta x^4}{4}\right)$ (solid line) and $\phi_2 \propto \exp\left(-\beta\left(\frac{x^4}{4} - x^2\right)\right)$ (dashed line). In the latter, two regions of high probability emerge, separated by a region of low probability around $x = 0$. The height of the barrier introduced in ϕ_2 between $x = 0$ and $x = \pm 1$ results in a considerably smaller LSI constant.

Sobolev inequalities may be shown to be satisfied by a large class of probability measures. The first of the two results states that any Boltzmann probability measure ν associated to an α-convex potential function satisfies an LSI.

Theorem 2 ***(Bakry-Emery theorem).*** *Let V be a twice-differentiable function with $\int e^{-V} = 1$. Then if $\nabla^2 V \geq \alpha I$ for some $\alpha > 0$, then e^{-V} satisfies LSI(α).*

The second shows that perturbing the underlying potential function by a bounded function leads to an exponentially decayed LSI constant.

Theorem 3 ***(Holley-Stroock Perturbation).*** *Let $\tilde{\nu}$ be defined by $\dfrac{d\tilde{\nu}}{d\nu} = e^U$, where ν satisfies LSI(ρ) and U is a bounded function such that $\displaystyle\int e^U d\nu = 1$. Then $\tilde{\nu}$ satisfies LSI($\tilde{\rho}$), where*

$$\tilde{\rho} = \rho \exp(-2\mathrm{osc}(U)), \quad \mathrm{osc}(U) = \sup(U) - \inf(U).$$

A straightforward proof may again be found in [Ané 2000], in Theorem 3.4.3. An important feature of the LSI constant is that it degenerates to zero in the case where the underlying potential is highly multimodal. For example, if $d\nu(x) = Z^{-1}\exp(-\beta W(x))\,dx$ and $W(x) = x^4/4 - x^2$ is the double-well potential in dimension 1, then the LSI constant scales as $\exp(-\beta\Delta W)$ where $\Delta W = W(0) - W(\sqrt{2}) > 0$ is the height of the barrier, see Figure 4.6. Such inequalities thus hold under rather loose assumptions, but the constant ρ is very small for a multimodal measure.

Let us now define the Wasserstein distance between two probability measures.

Definition 4 ***(Wasserstein distance).*** *The Wasserstein distance with linear cost between probability measures μ and ν is defined as*

$$W(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} |y-y'|\ \pi(dy,dy'),$$

where $\Pi(\mu,\nu)$ denotes the set of coupling probability measures on $\mathcal{D}\times\mathcal{D}$, with marginals μ and ν:

$$\forall\pi\in\Pi(\mu,\nu), \forall\varphi,\psi,\quad \int_{\mathcal{D}\times\mathcal{D}} (\varphi(y)+\psi(y'))\pi(dy,dy') = \int_{\mathcal{D}} \varphi d\mu + \int_{\mathcal{D}} \psi d\nu.$$

Definition 5 ***(Talagrand inequality).*** *The probability measure ν is said to satisfy a Talagrand inequality with constant $\rho>0$ (or $T(\rho)$) if for all probability measures μ such that $\mu \ll \nu$,*

$$W(\mu,\nu) \le \sqrt{\frac{2}{\rho} H(\mu|\nu)}. \tag{4.41}$$

Logarithmic Sobolev inequalities and Talagrand inequalities are related (see [Otto 2000]):

Lemma 2 *If ν satisfies $LSI(\rho)$, then ν satisfies $T(\rho)$.*

As we shall see in the next two sections, the notions and results above provide powerful tools in determining the rate of convergence of Boltzmann-type equations.

4.4.2 Convergence of the overdamped dynamics

Let us first study the longtime convergence to equilibrium of the simple overdamped Langevin dynamics and its limitations. As discussed in the previous section, the square of the L^1 distance of the probability density ψ from its equilibrium ψ_∞ can be bounded above by twice their relative entropy (see Theorem 1). Therefore, by showing that the relative entropy $H(\psi|\psi_\infty)$ decays exponentially quickly, then so does their total variation distance.

Recall that if the process Q_t follows the dynamics

$$dQ_t = -\nabla V(Q_t)\ dt + \sqrt{2\beta^{-1}} dW_t, \tag{4.42}$$

then the probability density $\psi(t,q)$ of the distribution of Q_t satisfies the *linear* Fokker-Planck equation

$$\partial_t\psi = \nabla\cdot(\nabla V\psi + \beta^{-1}\nabla\psi),\quad t>0, \tag{4.43}$$

with $X_0\sim\psi(0,q)\,dq$. We aim to show that the density ψ converges to its equilibrium $\psi_\infty = Z^{-1}e^{-\beta V}$ *exponentially* quickly, with a rate depending on the LSI constant of ψ_∞. To do so, we make the assumption that $\exists R>0$ such that

$$\psi_\infty \text{ satisfies LSI}(R).$$

In order to study its longtime convergence, it is convenient to rewrite (4.43) in terms of the equilibrium density ψ_∞

$$\partial_t \psi = \nabla \cdot \left(\psi_\infty \nabla \ln \left(\frac{\psi}{\psi_\infty} \right) \right). \tag{4.44}$$

The evolution in time of the relative entropy is

$$\frac{d}{dt} H(\psi|\psi_\infty) = \int \ln \left(\frac{\psi}{\psi_\infty} \right) \partial_t \psi = - \int \left| \nabla \ln \left(\frac{\psi}{\psi_\infty} \right) \right|^2 \psi = -F(\psi|\psi_\infty).$$

Furthermore, using the fact that ψ_∞ satisfies LSI(R), we have

$$\frac{d}{dt} H(\psi|\psi_\infty) = -F(\psi|\psi_\infty) \leq -2R H(\psi|\psi_\infty).$$

Therefore, ψ tends to ψ_∞ exponentially fast with rate $2R$:

$$H(\psi(t,\cdot)|\psi_\infty) \leq H(\psi(0,\cdot)|\psi_\infty) e^{-2Rt}.$$

By the Csiszar-Kullback inequality: $\int |\psi - \psi_\infty| \leq \sqrt{2H_0} e^{-Rt}$.

This result is satisfying at first glance, however in reality V possesses metastable regions, separated by high energy barriers, thus as a result of Theorem 3, R may in fact be very small. The ABF dynamics, presented in Section 4.3.2, helps to obtain a better rate of convergence.

4.4.3 Convergence of ABF dynamics

Using the same tools as above, a study of the ABF dynamics (4.37) can show that (i) the ABF method does indeed help to remove metastable features of the standard dynamics (4.42) and (ii) the adaptive bias A'_t converges to the mean force A' exponentially fast. We summarize below the results obtained in [Lelièvre 2008], where technical details (omitted below) may be found.

Let us begin by stating a number of assumptions. Consider the $3N$-dimensional configuration space $\mathcal{D} \subset \mathbb{R}^{3N}$, and the one-dimensional reaction coordinate $\xi : \mathcal{D} \to \mathcal{M}$, where typically $\mathcal{M} = \mathbb{R}$ or $\mathcal{M} = \mathbb{T} = \mathbb{R}/\mathbb{Z}$, the one-dimensional torus. We assume, as before, that the ξ submanifolds $\Sigma_z = \{q \mid \xi(q) = z\}$ are non-intersecting hyperplanes foliating the configuration space. This is given by the following assumption:

[A1] The function ξ is a smooth function and $|\nabla \xi(q)| > 0$ for all $q \in \mathcal{D}$.

In fact, convergence results for the exact dynamics presented in (4.37) depend on two further assumptions: (i) the reaction coordinate space $\mathcal{M} = \mathbb{T}$ and (ii) $|\nabla \xi|$ is constant (for $|\nabla \xi| \neq 1$ a change of time is necessary). We present below the slightly modified dynamics studied in [Lelièvre 2008], for a more general ξ and $\mathcal{M}$

$$\begin{cases} dQ_t = -\nabla(V - A_t \circ \xi + W \circ \xi - \beta^{-1}\ln(|\nabla\xi|^{-2}))|\nabla\xi|^{-2}dt + \sqrt{2\beta^{-1}}|\nabla\xi|^{-1}dB_t, \\ A_t'(z) = \mathbb{E}\left[f^V \,|\xi(Q_t) = z\right]. \end{cases} \tag{4.45}$$

To lighten notation, the dependence on Q_t has been omitted on the left-hand side of first line of (4.45). The function $W \circ \xi$ has been added to treat the case of an unbounded domain $\mathcal{M}$ (this term may be removed whenever $\mathcal{M} = \mathbb{T}$ for example). This potential is chosen in such a way as to confine the dynamics to the range of values of ξ of interest and in particular to ensure that the marginal law in ξ converges exponentially fast to equilibrium (see assumption [A4]). We emphasize that in the case $\mathcal{M} = \mathbb{T}$ and $|\nabla\xi| = 1$, the results of the paper hold for the dynamics (4.37). In the following, $\psi(t,\cdot)$ denotes the density of the law of Q_t, and $\psi^\xi(t,\cdot)$ the density of the law of $\xi(Q_t)$.

The nonlinear Fokker-Planck equation satisfied by the density $\psi(t,\cdot)$ is then given by

$$\partial_t\psi = \operatorname{div}\left(\frac{\nabla(V - A_t \circ \xi + W \circ \xi)\psi + \beta^{-1}\nabla\psi}{|\nabla\xi|^2}\right),$$

where we recall that the case of a one-dimensional reaction coordinate ξ, the biasing force is expressed in terms of the density $\psi(t,\cdot)$ in the following way

$$A_t'(z) = \frac{\displaystyle\int_{\Sigma_z} f^V|\nabla\xi|^{-1}\psi(t,\cdot)\, d\sigma_{\Sigma_z}}{\displaystyle\int_{\Sigma_z} |\nabla\xi|^{-1}\psi(t,\cdot)\, d\sigma_{\Sigma_z}}. \tag{4.46}$$

It can be shown indeed that if the potential A_t and the law of Q_t have stationary states A_∞ and ψ_∞ respectively, then

$$\psi_\infty \propto \exp(-\beta(V - A_\infty \circ \xi + W \circ \xi)).$$

Thus, by replacing $\psi(t,\cdot)$ by ψ_∞ in (4.46), one has $A_t' = A'$ and therefore $A_t = A$ up to an additive constant. This proves the uniqueness of the stationary state ψ_∞.

Let us now give the main assumptions needed to prove the exponential convergence of ψ to ψ_∞, in the relative entropy sense. For technical reasons, some regularity assumptions are required on ξ and the local mean force f^V:

[A 2] $$\begin{cases} \text{There exists } m > 0 \text{ and } M > 0 \text{ such that} \\ \sup\limits_{q\in\mathcal{D}} |\nabla\xi(q)| \le m < \infty \text{ and } \sup\limits_{q\in\mathcal{D}} |\nabla_{\Sigma_z} f^V(q)| \le M < \infty, \end{cases}$$

where ∇_{Σ_z} is the surface gradient on Σ_z, defined by

$$\nabla_{\Sigma_z} = \left(I - \frac{\nabla\xi \otimes \nabla\xi}{|\nabla\xi|^2}\right)\nabla,$$

with $u \otimes v$ being the $n \times n$ matrix with components $(u \otimes v)_{i,j} = u_i v_j$. A further fundamental assumption, which governs the rate of convergence of the ABF dynamics presented in (4.45) is on the conditional measures $\mu(\cdot|z)$.

[A3] $\exists\ \rho > 0$, such that $\forall z \in \mathcal{M}$, $\mu(\cdot|z)$ satisfies LSI(ρ).

This assumption, on both the potential V and reaction coordinate ξ, ensures that if one was to sample the measure $\mu(\cdot|z)$ using projected overdamped Langevin dynamics (such as [Ciccotti 2008]), the convergence to equilibrium would be exponentially fast, with rate ρ. Generally speaking, if ξ is a 'well-chosen' reaction coordinate (see Section 4.1.2), it is expected that ρ is sufficiently large (and in particular that $\rho > R$, where we recall R is the LSI constant of $Z^{-1}e^{-\beta V}$). The final assumption that is needed is that the Fisher information (see Definition 2) for the law of $\xi(Q_t)$ decays exponentially fast. This indeed depends on the confining potential W introduced in the dynamics (4.45).

[A4] $\begin{cases} \text{Suppose } W \text{ and the initial marginal distribution } \psi^\xi(0,\cdot) \text{ are such that} \\ \exists\ F_0, r > 0 \text{ such that for all } t \geq 0,\ F(\psi^\xi(t,\cdot)|\psi^\xi_\infty) \leq F_0 \exp(-2\beta^{-1} r t). \end{cases}$

Remark 6 *Assuming [A1], the above assumption holds true for two particular cases. In the simplest case, if $\mathcal{M} = \mathbb{T}$, then [A4] holds with $r = 4\pi^2$. If on the other hand $\mathcal{M} = \mathbb{R}$ and W is an α-convex function and $\exists \bar{r} > 0$, such that $Z_W^{-1}\exp(-\beta W)$ satisfies LSI($\bar{r}$), then [A4] holds with $r = \bar{r} - \varepsilon$, for all $\varepsilon > 0$.*

We are now ready to present the main result of [Lelièvre 2008].

Theorem 4 *Assuming [A1]-[A4], the following hold.*

(i) The biasing force A'_t converges to the mean force A' in the following sense: there exists $C > 0$, $\lambda > 0$ such that for all $t \geq 0$,

$$\int_{\mathcal{M}} |A'_t - A'|^2(z)\psi^\xi(t,z)\, dz \leq C\exp(-2\lambda t). \tag{4.47}$$

More precisely, when $\rho m^{-2} \neq r$, then

$$\lambda = \beta^{-1}\min(\rho m^{-2}, r).$$

In the case $\rho m^{-2} = r$, then for all $\lambda < \beta^{-1}\min(\rho m^{-2}, r)$, there exists a positive constant C such that (4.47) is satisfied.

(ii) $\|\psi(t,\cdot) - \psi_\infty\|_{L^1(\mathcal{D})}$ converges exponentially fast with the same rate as in (4.47).

The rate of convergence of the ABF method is thus limited by the minimum of the rate of convergence of the projected dynamics in the submanifolds Σ_z and the rate of convergence to equilibrium of the marginal distribution in ξ. In light of Remark 6, it is reasonable to assume that r is often large enough so that ρ, the LSI constant

of $\mu(\cdot|z)$, is the real limiting factor. In fact, ρ may indeed be very small in situations such as the one depicted in Figures 4.2(b) and 4.3. This gives rise to similar limitations encountered in constrained sampling methods (such as Thermodynamic Integration or Blue Moon Sampling). However, as an unconstrained technique, ABF has one advantage in a multi-channel or bi-channel scenario (cf. Figure 4.2(b)): sampling the two metastable regions at fixed $\xi(q) = z$ can be achieved by means other than overcoming the high-energy barrier separating them (e.g. by taking a longer, but energetically favorable path going through one of the global minima).

The next two chapters of this thesis are devoted to the ABF method, with the aim of showing that in situations where so-called *local* metastabilities are encountered (i.e. the constant ρ is very small), the rate of convergence of ABF may in fact be faster than that proposed in Theorem 4:

(i) This is justified *numerically* in Chapter 5. A novel implementation of the ABF method was tested on a small biomolecular system, whereby many walkers were simulated in parallel, each following the ABF dynamics (4.37), sharing a common biasing function A'_t. This multiple-walker implementation (so-called MW-ABF) showed faster convergence compared to a standard single-walker ABF simulation at constant wall *and* CPU time. The results can be explained by the fact that in compact states of a peptide chain, a great number of low-energy conformations are associated to a value of ξ of the RC, which are not fully explored by a standard single-walker ABF simulation due to their separation by high free-energy barriers.

(ii) This is justified *mathematically* in Chapter 6. By studying a simplified ABF model (representing the particular *bi-channel* scenario of Figure 4.2(b)), an improved convergence rate is obtained. More precisely, for the new model, the rate of convergence is shown to depend on the LSI constant of new canonical measures, conditioned to the submanifold Σ_z *and* to a particular channel. In view of the results discussed so far, this LSI constant is typically larger than ρ, the LSI constant of the conditional measure $\mu(\cdot|z)$.

Chapter 5

A Multiple-Walker ABF method

Ce chapitre reprend l'intégralité d'un article écrit en collaboration avec Christophe Chipot et Tony Lelièvre, et publié dans *Journal of Chemical Theory and Computation* [Minoukadeh 2010].

Sommaire

5.1 Introduction 81
5.2 General setting 83
5.3 Adaptive biasing force methods 84
5.3.1 Framework 84
5.3.2 Calculating the bias 85
5.3.3 Enhancing sampling through selection 86
5.4 Implementation details 89
5.5 Numerical results 93
5.5.1 Reaction coordinate range: 12-32 Å 94
5.5.2 Reaction coordinate range: 4-16 Å 94
5.5.3 Selection 97
5.6 Discussion 98

5.1 Introduction

Central to the understanding of most processes of either physical, chemical or biological interest, the determination of the underlying free-energy change occupies a prominent position in the arena of numerical simulations. Over the past decades, a variety of methods has been devised to compute free-energy differences efficiently (see for example references [Chipot 2007] and [Lelièvre 2010b]). Roughly speaking, these methods can be classified into two main categories: (i) The free energy is computed directly, or (ii) its first derivative is determined and subsequently integrated. Perturbation techniques [Zwanzig 1954], probability density function-based methods such as histogram methods [Bennett 1976, Kumar 1992, Shirts 2008], non-equilibrium computations [Jarzynski 1997] and adaptive biasing potential methods [Wang 2001b, Iannuzzi 2003], for example, fall into the first category. Thermodynamic integration [Kirkwood 1935], and adaptive biasing force methods [Darve 2001, Hénin 2004], which are the core of the present work, belong

to the second category. Adaptive methods are designed to compute free-energy profiles and favor transitions between metastable states by using a current estimate of the free energy as a biasing potential.

In this contribution, we are interested in a particular class of adaptive methods, referred to as adaptive biasing force methods [Darve 2001, Hénin 2004, Hénin 2010]. Specifically, we endeavor to investigate a novel implementation of this class of methods, using a number of walkers simulated in parallel, in the spirit of the ideas put forth by Lelièvre *et al.* [Lelièvre 2007a] The advantage of the present, novel implementation is threefold. First, the parallelization is straightforward and its theoretical parallel efficiency is very good since the only shared information is the biasing force, or the marginal law, namely low-dimensional functions. In turn, this yields efficient, scalable algorithms to compute free-energy differences, well adapted to the massively parallel architecture of high-performance computers. Second, we will show that the implementation relying upon many walkers is particularly interesting when the reaction coordinate does not describe well all the metastabilities of the system, which, quite unfortunately, constitutes a generic situation for the vast majority of non-trivial molecular systems. This is typically the case, for instance, of the so-called bi-channel scenario — namely the free-energy landscape features two parallel valleys, which are orthogonal to the isocontours of the reaction coordinate — or, more generally, when several transition mechanisms are associated to a single reaction coordinate, which is, therefore, not sufficient to parameterize fully the transformation. The underlying idea is that when many walkers are involved, they can visit more efficiently in parallel all the valleys in the direction of the reaction coordinate. A mathematical proof is currently underway to show that, in the limiting case of a very large number of walkers, and with suitable assumptions, the rate of convergence of the ABF method is in fact not limited by free-energy barriers orthogonal to the RC direction. Third, as will be detailed below, this new implementation allows selection mechanisms to be introduced, consisting in duplicating effective walkers, while deleting poor ones, according to a fitness function that ought to be chosen. An example of such a fitness function, which favors rapid exploration of the reaction coordinate, will be provided hereafter.

As a proof of concept, the present approach was probed on a realistic test case, using a high-level Tcl implementation of the algorithm in the scalable molecular dynamics program NAMD [Bhandarkar 2003, Kale 1999, Phillips 2005]. The efficiency of the overall procedure is, however, expected to be enhanced by embedding and optimizing the algorithm at a deeper level of the molecular-dynamics platform.

In the following section, the mathematical framework of the method is introduced and the adaptive biasing force method reviewed. Next, the discretization and implementation details are presented. The present contribution closes with a discussion of the numerical results obtained for the reversible folding of the paradigmatic deca-alanine peptide.

5.2 General setting

In the canonical ensemble, a system of dimension d is equipped with the Boltzmann-Gibbs probability measure, *i.e.* the canonical measure

$$\mu(dq) = \phi(q)\ dq = Z^{-1}\exp(-\beta V(q))\ dq, \tag{5.1}$$

where ϕ is the density of the measure, $\beta = 1/(k_B T)$ is proportional to the inverse temperature, $q \in \mathbb{R}^d$ is the system configuration, $V : \mathbb{R}^d \to \mathbb{R}$ is the potential energy function and $Z = \int_{\mathbb{R}^d} \exp(-\beta V(q))dq$ is the normalization constant or the so called partition function. To sample this measure, one can use the overdamped Langevin dynamics [Langevin 1908]:

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t \tag{5.2}$$

where $(X_t)_{t\geq 0}$ is the system trajectory and W_t is an $\mathbb{R}^d$-valued standard Brownian motion (or Wiener process). Under suitable regularity assumptions on the potential, the dynamics (6.1) is ergodic and admits the canonical measure as its unique invariant measure. It must be emphasized that, for the sake of simplicity, the method is described in the framework of the overdamped dynamics. The method can, nevertheless, be generalized to the Langevin dynamics as is done in the numerical simulations at the end of the paper.

The canonical measure (5.1) gives us microscopic information about the system, the probability that it is to be found at any particular point q in configuration space. A practitioner, however, is generally interested in some coarse-grained collective variable $\xi(q)$, where ξ is typically a smooth mapping from $\mathbb{R}^d$ to $\mathbb{R}$. In what follows, ξ will be referred to as the reaction coordinate (RC). The RC typically represents an end-to-end distance of a protein chain, a structural angle in a protein or a measure of the evolution of a chemical system. If X is a random variable with probability μ, then $\xi(X)$ is a random variable with law whose density ϕ^ξ is defined by

$$\phi^\xi(z) = \int_{\mathbb{R}^d} \phi(q)\delta(\xi(q) - z)\ dq, \tag{5.3}$$

and the distribution $\phi^\xi(z)dz$ is called the marginal distribution of μ in ξ. The free energy, or so-called potential of mean force (PMF), A, is related to this marginal density in the following way

$$A(z) = -\beta^{-1}\ln\ \phi^\xi(z). \tag{5.4}$$

Sampling the canonical measure using the standard overdamped Langevin dynamics (6.1) can in fact be inefficient in practice. The convergence to equilibrium can be very slow due to metastable states where the dynamics remains trapped for long periods of time. To explore the whole configuration space, one often needs to overcome very large energy barriers. From Arrhenius's law, it follows that the typical time needed to overcome these barriers scales exponentially with the barrier

height. Regular molecular-dynamics methods are, therefore, typically not used to calculate statistical averages for systems prone to metastabilities.

Several methods have been proposed to ameliorate sampling methods in these situations, such as the *blue moon* method [Carter 1989] or importance sampling methods such as *umbrella sampling* [Frenkel 1996]. More recently, adaptive importance sampling methods have been developed such as the *Wang-Landau* method [Wang 2001b] and the *adaptive biasing force* (ABF) method [Darve 2001]. The latter and its variations will be the focus of this contribution.

Before detailing the ABF method, the reader is reminded that the main quantity of interest in the study of chemical reactions is a free-energy difference and not an absolute free energy. Free energies are, therefore, computed only up to an additive constant. The free-energy difference between two coarse-grained states, labelled by the RC values z_a and z_b, can be written as

$$\Delta A = A(z_b) - A(z_a) = \int_{z_a}^{z_b} A'(z)\, dz \tag{5.5}$$

where $'$ is the derivative with respect to the collective variable value z and A' is called the mean force. The integrand can be shown to be the Boltzmann average of a real-valued function F^V, conditioned to being at a fixed point z in the reaction coordinate space

$$A'(z) = \frac{\displaystyle\int_{\mathbb{R}^d} F^V(q)\, \exp(-\beta V(q))\, \delta(\xi(q) - z)\, dq}{\displaystyle\int_{\mathbb{R}^d} \exp(-\beta V(q))\, \delta(\xi(q) - z)\, dq} = \langle F^V(q)\, |\xi(q) = z \rangle_\mu \tag{5.6}$$

where

$$F^V = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \nabla \cdot \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \tag{5.7}$$

and $\langle \cdot \rangle_\mu$ represents the canonical average — *i.e.* the average with respect to the measure μ. Note that F^V is the negative projection of the force onto the RC plus some correction term. For the derivation of (5.7) the reader is referred to references [Ciccotti 2008], [Sprik 1998] and [den Otter 1998]. The aim of the ABF method, which will be detailed hereafter, is to compute A' as efficiently as possible.

5.3 Adaptive biasing force methods

In this section we will present the framework behind ABF methods for free energy computations.

5.3.1 Framework

The basic idea of ABF is to use the mean force estimate to bias the dynamics and help the system overcome free-energy barriers. An estimate of $A'(z)$ is obtained as the statistical average of the force field F^V at specified points z along the RC

by accruing instantaneous forces $F^V(X_t)$ for a single system trajectory X_t when $\xi(X_t) = z$. In the long-time limit, one obtains a good approximation for A' and ΔA can be computed by numerical integration. The resulting biased dynamics is

$$\begin{cases} dX_t = -\nabla(V - A_t \circ \xi)(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \\ A'_t(z) = \langle F^V(X_t)|\xi(X_t) = z\rangle, \end{cases} \tag{5.8}$$

where $A_t \circ \xi$ denotes the composition of A_t with ξ, so that $A_t \circ \xi(x) = A_t(\xi(x))$, and A'_t is the estimated mean force[1]. The estimated mean force is thus defined as a conditional average of $F^V(X_t)$ at a fixed value of $\xi(X_t)$. In practice, it can be approximated as an average over many walkers or as an *on-the-fly* average over the trajectory X_t (see next section for more details). The above can be viewed as an overdamped Langevin dynamics, with the potential V replaced by the time varying potential $\mathcal{V}_t = V - A_t \circ \xi$. The consistency of the method may be justified by noticing that if a stationary state A'_∞ for A'_t is obtained, then ψ_∞ is proportional to $\exp(-\beta\mathcal{V}_\infty)$, and thus $A'_\infty = A'$ since

since

$$A'_\infty(z) = \langle F^V(q)|\xi(q) = z\rangle_{\exp(-\beta\mathcal{V}_\infty)} = \langle F^V(q)|\xi(q) = z\rangle_\mu = A'(z),$$

As a result, A_t converges to A up to an additive constant, and the equilibrium marginal density in ξ is constant, since

$$\langle \delta(\xi(q) - z)\rangle_{\exp(-\beta\mathcal{V}_\infty)} = \frac{\langle \delta(\xi(q) - z)\rangle_\mu}{\exp(-\beta A(z))}$$

is constant by the definitions of μ and A in (5.1) and (5.4) respectively. Precise convergence results can be found in references [Lelièvre 2008]. The aim of ABF is, therefore, to estimate the biasing force as efficiently as possible in order to bias the dynamics by reducing and eventually eliminating any force along ξ. It serves as an adaptive importance sampling method, driving the system out of its metastable states, using on-the-fly estimates of the mean force.

5.3.2 Calculating the bias

Different approaches have been proposed in recent literature [Darve 2001, Lelièvre 2007b, Lelièvre 2008] for calculating the biasing force. There are two principal methods for computing A'_t, which will serve as a basis of comparison in the present contribution.

Original ABF: The idea of the standard ABF method [Darve 2001], which involves one single walker, is to calculate averages using the whole trajectory of the system. The mean force is calculated by taking a trajectorial average of instantaneous forces

[1] Note that the gradient term in the biased dynamics can be rewritten as $-\nabla V + (A'_t \circ \xi)\nabla\xi$, thus only estimated mean force information is needed and not the estimated free energy.

at fixed z using one long system trajectory (see reference [Chipot 2007] for further details)

$$\langle F^V(X_t)|\xi(X_t)=z\rangle \;\simeq\; \frac{\displaystyle\int_0^t F^V(X_s)\;\delta(\xi(X_s)-z)\;ds}{\displaystyle\int_0^t \delta(\xi(X_s)-z)\;ds}. \tag{5.9}$$

The mean force estimate is only computed once a trajectory reaches the value z in the RC, therefore the denominator in the above equation is always non-zero for the RC values needed.

Multiple walker (MW-)ABF: In a recent paper [Lelièvre 2007b], a new formulation of the ABF method has been proposed, consisting in running $R > 1$ trajectories of the ABF dynamics in parallel. The R walkers of the system follow a similar dynamics driven by independent Brownian motions. These multiple walkers then exchange information at fixed time intervals. The immediate gain of this new formulation is that one can take advantage of parallel computing to speed up convergence of the ABF method. Furthermore, with the use of a small number of walkers, we are able to overcome issues related to poorly chosen or oversimplified reaction coordinates, where other important slow degrees of freedom are overlooked. In such cases, metastabilities can be found at fixed ξ, as illustrated in 5.1. With multiple walkers, it is likely that each walker will explore a different path or valley along ξ, whereas single–walker simulations could potentially take exponentially long times to fully explore the low energy states. This will be studied numerically in the final sections of the paper.

In the following, $(X_t^i)_{0\leq i\leq R-1}$ is the set of trajectories for the R walkers. Each trajectory X_t^i follows the dynamics (5.8) with the Brownian motion W_t replaced with W_t^i. The biasing force is then estimated using an average over all the trajectories and over all walkers

$$\langle F^V(X_t)|\xi(X_t)=z\rangle \simeq \frac{\displaystyle\sum_{i=0}^{R-1}\int_0^t F^V(X_s^i)\;\delta(\xi(X_s^i)-z)\;ds}{\displaystyle\sum_{i=0}^{R-1}\int_0^t \delta(\xi(X_s^i)-z)\;ds}. \tag{5.10}$$

Implementation details for the approaches discussed above will be discussed in the next section.

5.3.3 Enhancing sampling through selection

In addition to the exchange of information between walkers to compute the estimated mean force A_t', the MW-ABF method allows for resampling of the walkers according to their "importance". The success of the ABF method is strongly determined by the marginal distribution of walkers in the RC, given that the RC has been well chosen.

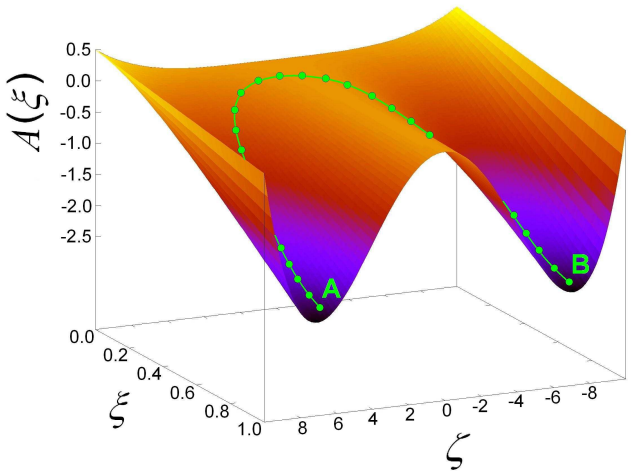


Figure 5.1: Example of a 2-dimensional free-energy surface exhibiting metastabilities at fixed ξ. The variable ζ represents another slow degree of freedom of the system, orthogonal to the RC. The standard ABF method relies on efficient sampling at fixed points in the RC, which is made difficult by the presence of such large energy barriers in the orthogonal directions. Using multiple walkers helps to overcome this issue as each one is very likely to explore a different pathway.

One would, therefore, want to encourage walkers that are exploring undersampled regions of the RC and penalize those in oversampled regions. A selection mechanism [Assaraf 2000, Doucet 2001, Lelièvre 2007b] may be used to achieve this objective. It is implemented by a system of interacting walkers, where the walkers are cloned or killed at a rate defined by $S(t,z)$ over the values taken by the RC. The function $S(t,z)$ can be chosen as

$$S(t,z) = c\frac{\partial_{zz}\psi_t^\xi}{\psi_t^\xi}, \tag{5.11}$$

where c is a positive constant and ψ_t^ξ, defined by

$$\psi_t^{\xi,I}(z) = \int_{\mathbb{R}^d} \psi_t(q)\ \delta(\xi(q) - z)\ dq,$$

represents the marginal distribution of walkers in the RC at time t. With this choice of the function S, it can be shown that the marginal density $\psi^{\xi,I}$ satisfies the partial differential equation (in fact for a slightly modified version of the original adaptive dynamics (5.8), see reference [Lelièvre 2007b])

$$\partial_t \psi_t^{\xi,I} = (\beta^{-1} + c)\partial_{zz}\psi_t^{\xi,I}. \tag{5.12}$$

The selection process thus accelerates the diffusion of the marginal distribution in the RC. The reason for this choice of S can also be understood when written in a

finite difference form

$$S(t,z) \simeq \frac{3\ c}{\Delta z^2\ \psi_t^\xi(z)} \left[\frac{\psi_t^\xi(z-\Delta z) + \psi_t^\xi(z) + \psi_t^\xi(z+\Delta z)}{3} - \psi_t^\xi(z) \right], \tag{5.13}$$

where Δz is some small displacement in the RC. The quantity S at a given value of the RC is, therefore, positive if the marginal density at this point is small compared to its local average, and negative otherwise. To implement the selection process, one can either continuously update birth and death times, initially drawn from an exponential distribution, as in reference [Lelièvre 2007b], or resample the walkers according to their weights at fixed resampling times. The latter will be used for the simulations reported herein. At a resampling time t, each walker trajectory X_t^i is given a weight [Del Moral 2004]

$$w_t^i = K_t^{-1} \exp\left[\int_0^t S\left(s, \xi\{X_s^i\}\right)\ ds \right], \tag{5.14}$$

where $K_t = \sum_{i=0}^{R-1} \exp\left[\int_0^t S\left(s, \xi\{X_s^i\}\right) ds \right]$ is the normalization constant. Replicas are initially assigned a uniform weight, $w_0^i = 1/R$, which evolves in time. From (5.13) and (5.14), it is now clear that a walker i that is often found in undersampled regions — in which case $S(t, \xi(X_t^i))$ is often positive — is given a stronger weight than walkers in oversampled regions — where $S(t, \xi(X_t^i))$ is often negative. The i^{th} walker is then replicated on average Rw_t^i times. This procedure thus accelerates the convergence to a uniform distribution of the walkers in the RC, in accordance with (5.12).

Let us now give some details about the resampling procedure. To calculate the number of times a walker is to be copied, a systematic resampling method [Carpenter 1999, Douc 2005, Kitagawa 1996] is used, described briefly by the following algorithm. At a resampling time t:

Set $u \sim U(0,1)$, $\bar{w}_0 = w_t^0$, $N_0 = \lfloor R * \bar{w}_0 + u \rfloor$,
for $i = 1, \ldots, R-1$
  $\bar{w}_i = \bar{w}_{i-1} + w_t^i$,
  $N_i = \lfloor R * \bar{w}_i + u \rfloor - \lfloor R * \bar{w}_{i-1} + u \rfloor$
end

where $U(0,1)$ denotes a uniform distribution between 0 and 1, w_t^i is the normalized weight assigned to walker i as defined in (5.14), $\bar{w}_i$ is the cumulative sum of the weights, $\lfloor \cdot \rfloor$ is the integer part and N_i is the number of copies of walker i to be generated. It is important to note that this algorithm guarantees that $\sum_{i=0}^{R-1} N_i = R$. After every resampling stage, the weights of all walkers are reset uniformly to the value $1/R$. The choice of the constant c in (5.11) is of paramount importance in the performance of the selection mechanism. This parameter should be sufficiently large to accelerate the exploration along the RC, but not too large in case one walker is

selected during the resampling stage (due to degeneracy of weights), which implies a very large variance of the estimator. This will be discussed further at the end of the next section.

5.4 Implementation details

In this section, the implementation details of the adaptive biasing force methods are provided. The simulations reported in the present contribution have been carried out using the scalable molecular-dynamics code NAMD, but the algorithmic detail is by no means specific to this software package. The ABF methods have been implemented as Tcl scripts, for which the single–walker ABF method is already available. How the method is discretized will be spelled out hereafter and the detail of the single–walker ABF method will be outlined before proceeding with the implementation of the MW-ABF method and selection.

We consider a reaction coordinate ξ taking values in the interval $[z_0, z_N]$, which is divided into N bins of size $\Delta z = (z_N - z_0)/N$. We denote by $\tilde{\xi} : \mathbb{R}^d \to \{0, \dots, N-1\}$ a mapping from a configuration onto its associated bin in the RC

$$\tilde{\xi}(\cdot) = \left\lfloor \frac{\xi(\cdot) - z_0}{\Delta z} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ again denotes the integer part. In the following, functions and trajectories will be indexed by the number of time steps k, so that A'_k will be the mean force approximation and X_k will be the configuration of the system at time $k\Delta t$, for a time step Δt. Furthermore, with a slight abuse of notation, z will now denote the bin in the reaction coordinate, $z = \tilde{\xi}(X_t)$.

Original ABF method: The reader is reminded that in the standard ABF method, the biasing force is calculated for each bin using a trajectorial average, as in (5.9). The biasing force is in practice updated to include the current force observation. For $z \in \{0, \dots, N-1\}$

$$A'_k(z) = \frac{n_{\mathrm{tot}}(k-1, z)}{n_{\mathrm{tot}}(k, z)} A'_{k-1}(z) + \frac{\mathbf{1}_{\tilde{\xi}(X_{k-1})=z}}{n_{\mathrm{tot}}(k, z)} F^V(X_{k-1}) \tag{5.15}$$

where $\mathbf{1}_{\tilde{\xi}(X_k)=z}$ denotes the indicator function — taking value 1 if $\tilde{\xi}(X_k) = z$ and 0 otherwise — and

$$n_{\mathrm{tot}}(k, z) = \sum_{s=0}^{k-1} \mathbf{1}_{\tilde{\xi}(X_s)=z} \tag{5.16}$$

is the total number of times the system trajectory has visited bin z. To justify (5.15), the expression in (5.9) is recast in its discretized form

$$A'_k(z) = \frac{\sum_{s=0}^{k-1} F^V(X_s) \mathbf{1}_{\tilde{\xi}(X_s)=z}}{\sum_{s=0}^{k-1} \mathbf{1}_{\tilde{\xi}(X_s)=z}}. \tag{5.17}$$

Developing further, one subsequently obtains

$$\begin{aligned} A'_k(z) &= \frac{\sum_{s=0}^{k-2} F^V(X_s)\mathbf{1}_{\tilde{\xi}(X_s)=z} + F^V(X_{k-1})\mathbf{1}_{\tilde{\xi}(X_{k-1})=z}}{n_{\mathrm{tot}}(k,z)} \\ &= \frac{n_{\mathrm{tot}}(k-1,z)A'_{k-1}(z) + F^V(X_{k-1})\mathbf{1}_{\tilde{\xi}(X_{k-1})=z}}{n_{\mathrm{tot}}(k,z)} \end{aligned}$$

where the last line follows from the definition (5.17) at time $k-1$.

MW-ABF: The basis for the multiple–walker implementation of ABF in NAMD can be found in a set of Tcl scripts written for parallel-tempering, walker-exchange simulations [Bhandarkar 2003, Phillips 2005]. The scripts use Tcl server and socket connections to launch NAMD processes for each individual walker. Each walker is handled by a different computing unit. The NAMD processes run for a fixed number of time steps, then wait in order for Tcl server to exchange information between walkers. 5.2 is synoptic diagram of the MW-ABF method. It is not necessary (and not desirable from a computational point of view) to exchange information at every time step. Exchange of information between walkers only occurs at every k_{ex} time steps. We therefore proceed as follows: the mean force approximation, denoted by $A'_{k,i}(\cdot)$, is evaluated locally on the computing unit, where the indices k and i represent respectively the number of time steps since the beginning of the simulation and the computing unit running walker i. This quantity therefore depends solely on the trajectory of the walker of interest. Between exchange times, $k \in [\, nk_{\mathrm{ex}}\ ,\ (n+1)k_{\mathrm{ex}}\ ]$, the mean force estimation evolves according to the update formula

$$A'_{k,i}(z) = \frac{n^i_{\mathrm{loc}}(k-1,z)}{n^i_{\mathrm{loc}}(k,z)} A'_{k-1,i}(z) + \frac{\mathbf{1}_{\tilde{\xi}(X^i_{k-1})=z}}{n^i_{\mathrm{loc}}(k,z)} F^V(X^i_{k-1}). \tag{5.18}$$

Note that this is the same as (5.15), where A'_k is replaced by $A'_{k,i}$, X_k replaced by X^i_k and $n_{\mathrm{tot}}(k,z)$ replaced by $n^i_{\mathrm{loc}}(k,z)$, the number of times walker i has entered bin z since the last exchange, defined in (5.19) below.

At every exchange time, the information gathered by each walker is collected and local variables on each processor are updated. This is formalized with the help of some further notation. We denote by

$$k_{\mathrm{last}} = \lfloor k/k_{\mathrm{ex}} \rfloor\, k_{\mathrm{ex}}$$

the time of the last exchange. Next,

$$n^i_{\mathrm{loc}}(k,z) = \sum_{s=k_{\mathrm{last}}}^{k-1} \mathbf{1}_{\tilde{\xi}(X^i_s)=z}, \tag{5.19}$$

denotes the number of times walker i has entered bin z since the last exchange and

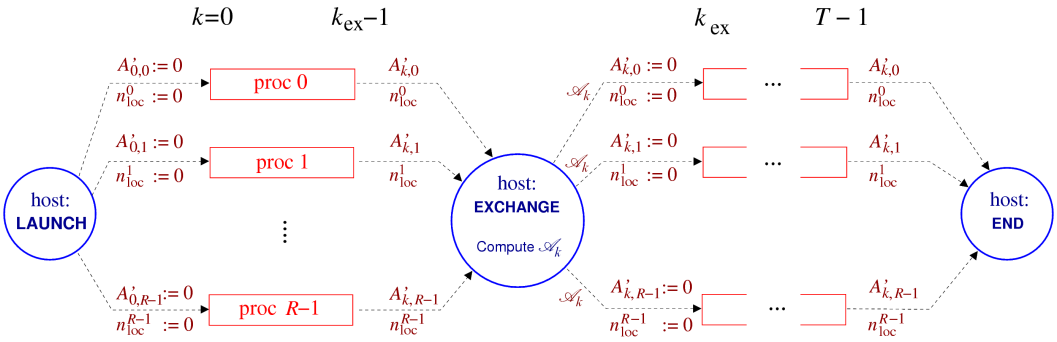


Figure 5.2: Schematic diagram of MW-ABF: The main script is executed on a host machine, which acts as the Tcl server. This machine launches the R walkers onto different processors via socket connections and, after every k_{ex} time steps, carries out exchange of information. This consists in: reading in local variables from each processor; computing the total biasing force $\mathcal{A}'_k$ by means of (5.21) and (5.22); sharing $\mathcal{A}'_k$ with all processors and setting local variables to zero. This is carried out T/k_{ex} times until the program terminates.

$$n^i_{\mathrm{tot}}(k,z) = \sum_{s=0}^{k-1} \mathbf{1}_{\tilde{\xi}(X^i_s)=z}, \tag{5.20}$$

is the total number of times walker i has entered bin z since the beginning of the simulation. Finally,

$$N_{\mathrm{loc}}(k,z) = \sum_{i=0}^{R-1} n^i_{\mathrm{loc}}(k,z) \quad \text{and} \quad N_{\mathrm{tot}}(k,z) = \sum_{i=0}^{R-1} n^i_{\mathrm{tot}}(k,z)$$

denote respectively the total number of visits to bin z since the last exchange and the beginning of the simulation, over all the walkers.

At every exchange time, a local average A'_{loc} is calculated of the mean force estimated from the run of each individual walker:

$$A'_{\mathrm{loc}}(k,z) = \frac{1}{N_{\mathrm{loc}}(k,z)} \sum_{i=0}^{R-1} n^i_{\mathrm{loc}}(k,z) A'_{k,i}(z). \tag{5.21}$$

The total biasing force, $\mathcal{A}'_k(z)$, to be shared between the walkers, is then also updated to include this new information

$$\mathcal{A}'_k(z) = \left[1 - \frac{N_{\mathrm{loc}}(k,z)}{N_{\mathrm{tot}}(k,z)}\right] \mathcal{A}'_{k-1}(z) + \frac{N_{\mathrm{loc}}(k,z)}{N_{\mathrm{tot}}(k,z)} A'_{\mathrm{loc}}(k,z). \tag{5.22}$$

The latter quantity, $\mathcal{A}'_k$, is then communicated to each one of the walkers, and the variables n^i_{loc} and $A'_{k,i}$ are reset to zero.

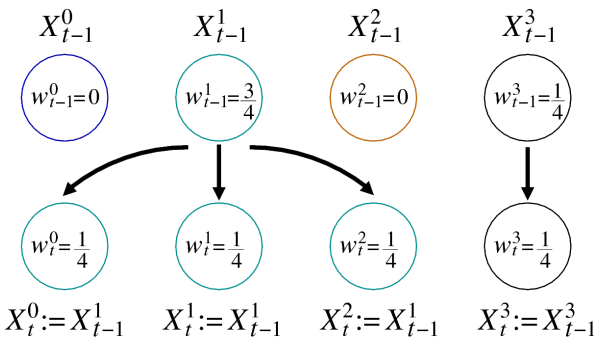


Figure 5.3: Selection mechanism for $R = 4$ walkers. If walker i has weight w^i_k at the time of selection, on average Rw^i_k copies are made of this walker at the next step. In practice, this means that Rw^i_k walkers will be launched using the configuration and velocity files of walker i. Note that in the above $k = nk_{\text{ex}}$, the time at which selection is carried out.

The total biasing force in (5.22) is utilized by each walker in the steps following the exchange, however new local information is also incorporated in order to speed up the diffusion in ξ. The biasing force applied to walker i in the simulations, therefore, writes[2]

$$F^i_{\text{bias},k}(z) = \left[1 - \frac{n^i_{\text{loc}}(k,z)}{N_{\text{tot}}(k,z)}\right] \mathcal{A}'_{k_{\text{last}}}(z) + \frac{n^i_{\text{loc}}(k,z)}{N_{\text{tot}}(k,z)} A'_{k,i}(z),$$

where $\mathcal{A}'_{k_{\text{last}}}$ is again the total mean force calculated at the preceding exchange interval and $A'_{k,i}$ is the local mean force information, as defined in (5.18).

Selection: Resampling may be carried out at most every k_{ex} time steps, when the walkers exchange information. Selection is a technically costly process as NAMD must be exited and reloaded with new configuration and velocity files. For this reason, it is even advisable for it to be carried out less frequently. The computational complexity of the process is $O(R)$ using a systematic resampling method (see the previous section for the algorithm). For purposes of illustration, the resampling will be carried out as often as the inter-processor communication, namely every k_{ex} time steps. The purpose of resampling is to improve the exploration in the RC. The weights of the walkers are adjusted according to the utility function $S(k,z)$, depending in practice upon the total distribution of the walkers:

$$S(k,z) = c\, \frac{N_{\text{tot}}(k_{\text{last}}, z+1) - 2N_{\text{tot}}(k_{\text{last}}, z) + N_{\text{tot}}(k_{\text{last}}, z-1)}{N_{\text{tot}}(k_{\text{last}}, z)}. \tag{5.23}$$

[2] In practice, this force is actually only fully applied to the dynamics after a certain number of visits have been made to the bin, that is to say after $N_{\text{tot}}(k,z) > N_{\text{min}}$, where in our simulations $N_{\text{min}} = 500$. If $N_{\text{tot}}(k,z) < N_{\text{min}}/2$, then no biasing force is added. Beyond that, the force is slowly introduced using a ramp function with scaling factor $\min\left(\frac{2N_{\text{tot}}}{N_{\text{min}}} - 1, 1\right)$.

The integral in (5.14) is calculated by summing the terms $S(k, \tilde{\xi}(X_k^i))$ over k for each walker i during each individual run. At the selection stage, when $k = k_{\text{ex}}$, the weights of the walkers are computed:

$$w_k^i = K_k^{-1} \exp\left[\sum_{s=k_{\text{last}}}^{k-1} S(s, \tilde{\xi}\{X_s^i\})\right],$$

where K_k is again the normalization constant. The walkers are then selected according to these weights using a systematic resampling method, as described above. In practice, to generate N_i copies of walker i, the configuration and velocity files for walker i are passed to NAMD as the start-up files for N_i walkers. Finally, after resampling, S is set to zero, so that all walker trajectories have equal weight.

For resampling to be effective, there are two main issues that need to be addressed. First, the constant c has to be chosen carefully: it must be large enough for the selection mechanism to be beneficial and small enough to avoid degeneration of weights, where all walkers are given zero weight except for one. Another issue to be addressed is when to stop resampling. Due to the technical costs of the selection mechanism, it is advised to impose a stopping criterion, so that selection is not applied throughout the whole simulation. The stopping criterion could depend on the sampling of the RC, or on the distribution of the weights. For the simulations herein, the latter criterion is used for once the walkers begin to be equally weighted, the selection has no effect and ought to be stopped. In order to measure the distribution of the weights we consider the relative entropy of the weights compared to a uniform distribution, defined by

$$E_w(t) = \sum_{i=0}^{R-1} w_i \log(R w_i). \tag{5.24}$$

This can be understood as the difference between the entropy of weights and the entropy for the uniform weight distribution: $\sum_{i=0}^{R-1} w_i \log(w_i) - \log(1/R)$. This quantity is bounded above by $\log(R)$, in case of degeneracy, and is bounded below by 0, in case of uniform distribution of weights. A good stopping criterion for the selection algorithm would be to end the process when the relative entropy is sufficiently small. In our simulations, the selection is stopped once

$$E_w(t) < \varepsilon \log(R) \tag{5.25}$$

where $0 < \varepsilon < 1$ is set closer to 0 for a stringent stopping criterion, or closer to 1 for a weaker threshold.

5.5 Numerical results

In this section, we present comparisons of the single–walker and multiple–walker ABF methods on the deca-alanine peptide in gas-phase, for which comprehensive

studies have already been carried out [Chipot 2005, Hénin 2004, Park 2003]. All the simulations reported herein were performed with the molecular-dynamics code NAMD [Bhandarkar 2003, Kale 1999, Phillips 2005], using the CHARMM27 [MacKerell 1998] force field. The ten-residue peptide chain has a total of 104 atoms and the RC has been chosen as the distance separating the center of mass of its first and last C-H pairs. To sample the full range of conformations from the α-helical conformation to the ensemble of extended structures, the range of values accessible to ξ varies from 12 to 32 Å. Additional tests are also carried out to study more compact conformations, where ξ varies between 4 and 16 Å. The system is kept within the assigned ranges by enforcing reflective boundary conditions.

The average forces were accumulated in bins of size $\Delta z = 0.1$ Å. The equations of motion were integrated employing Langevin dynamics with a time step $\Delta t = 0.5$ fs. Electrostatic and van der Waals interactions were truncated smoothly beyond 11 Å.

We will first present the results for the conventional range of 12 to 32 Å, which spans conformations comprised between the α-helix to more extended structures. Next, results for more compact conformations — with ξ ranging from 4 to 16 Å — are presented, where stark differences can be observed between the single– and multiple–walker ABF simulations. Finally, we will study the impact of selection on walkers.

5.5.1 Reaction coordinate range: 12-32 Å

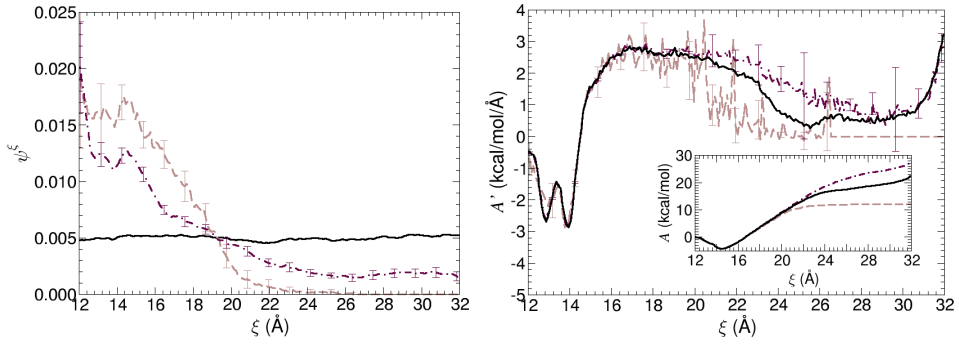
Starting from the α-helical conformation, R walkers of the system are launched with ABF dynamics, communicating at every $k_{\text{ex}} = 50,000$ time steps (25 ps). Reference curves are obtained from a 200-ns simulation using the original ABF algorithm, featuring a single walker.

5.5.1 compares the sampling distribution, mean force and free-energy profiles for single–walker and 16–walker simulations after 0.25 ns.

It may be observed from Figure 5.4(a) that the single–walker runs rarely manage to stretch beyond a distance of $\xi = 22$ Å, whereas the 16–walker simulations explore the whole reaction-coordinate space. Furthermore, in Figure 5.4(b), it is apparent that the mean force and free-energy profiles obtained by the 16–walker simulations are already qualitatively consistent with the reference curves.

5.5.2 Reaction coordinate range: 4-16 Å

As previously mentioned, convergence of the standard ABF method can be rather slow in the presence of metastabilities on the submanifold of conformations at a fixed value of ξ. This is generally the result of a poor choice of the RC, which does not capture all metastabilities of the system. In such a case — as depicted in 5.1 — several low energy conformations could be associated to a fixed value of ξ, and separated by high-energy barriers. As highlighted in reference [Chipot 2005], this shortcoming arises when studying compact conformations of the deca-alanine

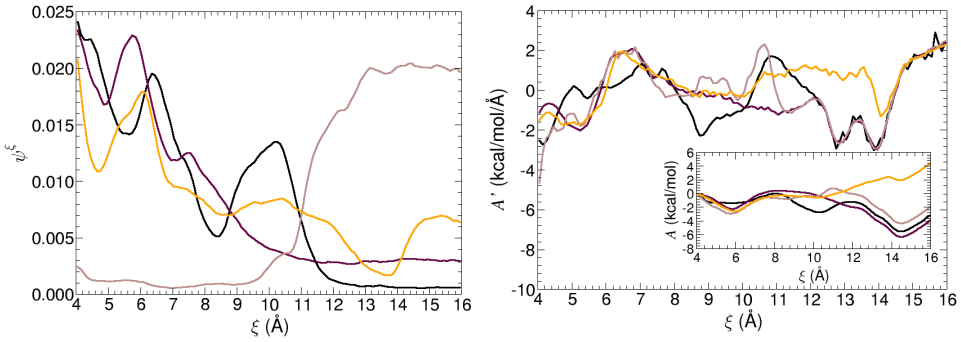


(a) (b)

Figure 5.4: **Results for ξ ranging from 12 to 32 Å (after 0.25 ns).** The curves are averages of 20 independent single–walker (dashed lines) and 16–walker (dashed-dotted lines) simulations with error bars representing the 95% confidence intervals. Solid lines represent reference profiles, obtained from a single–walker run of 200 ns. (a) Density of marginal distribution in the RC. The multiple–walker simulation has explored the whole ξ-space whereas the single–walker simulations very rarely stretch beyond 22 Å. (b) Mean force and free-energy profiles (inset). For the multiple–walker simulations we see the mean force profiles already nearly converged, whereas little information is gathered beyond 22 Å for the standard ABF simulation.

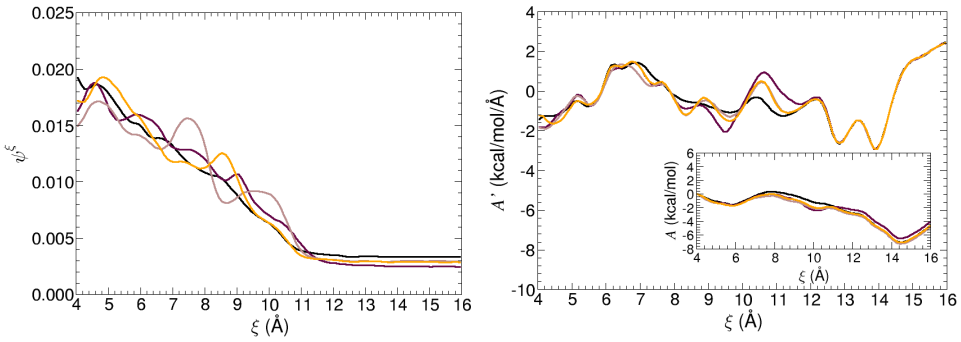
peptide. In this article, an extension of the standard sampling window reveals a free-energy profile that exhibits a wide global minimum ranging from 4 to 12 Å. It is known, however, that the global minimum of the deca-alanine peptide is the α-helical conformation at about $\xi = 14$ Å (see references [Hénin 2004] and [Park 2003]). The present results can be explained by the fact that in compact states, a great number of low-energy conformations are associated to a value of ξ of the RC, which are not fully explored by a standard, single–walker ABF simulation due to their separation by high free-energy barriers. These high free-energy barriers are generally unsurmountable from the perspective of conventional MD simulations and can be viewed as kinetic traps that preclude the exploration of the full RC space over reasonable time scales. A recent study has helped to capture the various slow degrees of freedom for these compact structures by exploring multidimensional free-energy landscapes [Hénin 2010].

The shortcomings discussed above can be advantageously circumvented using multiple walkers. The results obtained from 100–ns single– and multiple–walker simulations of the compact conformations are compared in 5.5.2 and 5.5.2 respectively. Figure 5.5(b) depicts mean force estimations for four independent single–walker simulations. Even after a 100–ns simulation, large discrepancies are observed between the mean force profiles. As can be observed in the inset of Figure 5.5(b), one free-energy profile has revealed a global minimum around $\xi = 6$ Å, which is, in most



(a) (b)

Figure 5.5: **Results for ξ ranging from 4 to 16 Å using 1 walker (after 100 ns).** Results are from four independent simulations. (a) Sampling along the RC. (b) Mean force approximations and free-energy profiles (inset): large discrepancies are observed, suggesting presence of parallel valleys along ξ. Note that one of the free-energy profiles suggests a global minimum at $\xi = 6$ Å.

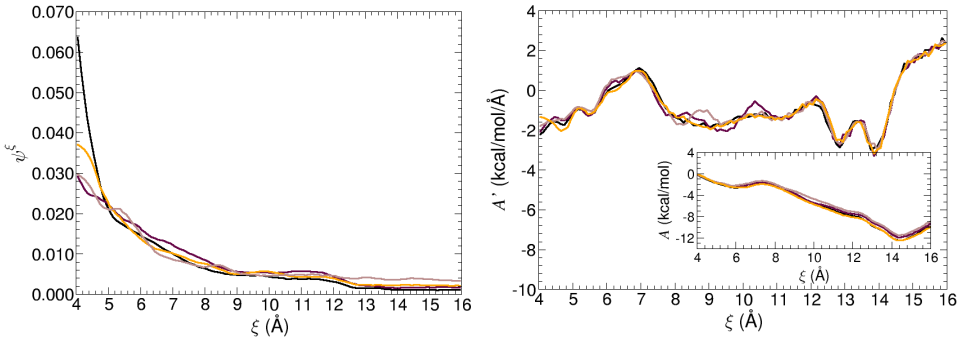


(a) (b)

Figure 5.6: **Results for ξ ranging from 4 to 16 Å using 32 walkers (results after 100 ns).** Results are from four independent simulations. (a) Sampling along the RC. (b) Mean force approximations and free-energy profiles (inset). Sampling and mean force estimations are consistent with each other and the α-helical conformation is recovered as the global free-energy minimum.

likelihood, artifactual. 5.5.2 summarizes the results obtained from four independent 32–walker simulations. A marked improvement in the convergence of the mean force profiles is immediately apparent. This supports the speculation that there exist parallel valleys along the RC, each of a different nature, separated by high free-energy barriers. The present set of results is far more promising with a multiple–walker scheme.

Due to eventual traps in the parallel valleys, it is in fact likely that a T-



(a) (b)

Figure 5.7: **Results after 3 ns using 32 walkers.** To compare results at constant total CPU time, we observe the results of a 32-walker simulation after $100/32 \sim 3$ ns. Results are from four independent simulations. (a) Sampling along the RC. (b) Mean force estimates and free-energy profiles (inset) are qualitatively very close to Figure 5.6(b). The results show that a multiple–walker simulation can outperform a single–walker simulation at constant CPU time.

nanosecond single–walker simulation will be less efficient than an R-walker simulation ran for T/R nanoseconds. This argument is supported numerically by 5.7, showing results for a 32-walker simulation after $100/32 \sim 3$ ns. The results are qualitatively consistent with 5.5.2 and offer a far more reliable set of results than a 100 ns single–walker simulation. In this way, a multiple–walker implementation not only improves results at constant wall time, but at *constant CPU time*.

5.5.3 Selection

In order to monitor the impact of the selection mechanism, we have chosen to study again the standard range of 12–32 Å. This choice is dictated by the topology of the free-energy landscape. It ought to be recalled here that the selection criterion is based solely on the position of the walkers along the RC. In the case of multiple, parallel valleys, this criterion can, therefore, impede convergence of the mean force, should many copies of one walker be generated in the same valley. It appears that selection is most effective in the presence of metastabilities in the RC only. Metastabilities in the orthogonal directions are well explored by multiple–walker simulations, albeit not always improved by selection.

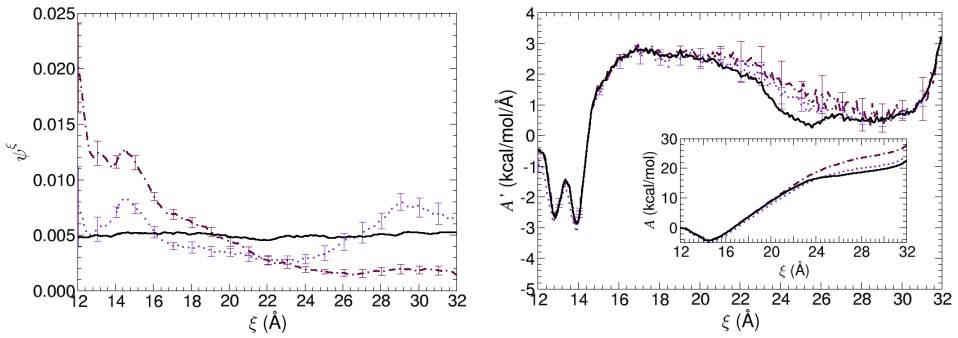
Figures 5.8(a) and 5.8(b) compare the sampling and free-energy profiles determined by a 16–walker simulation after 0.25 ns. We observe a more uniform sampling and a better potential of mean force for the simulation *with* a selection mechanism. For these simulations, we used the selection constant $c = 0.0001$ in (5.23) and a stopping criterion as in (5.25) with $\varepsilon = 0.05$.

5.9 depicts $E_w(t)$, the relative entropy of the walker weights, during the first 2.5 ns of the simulation. It may be observed that $E_w(t)$ decreases during the ABF

simulation, as the biasing force converges. This result suggests that the walkers are then more free to move along the RC and, thus, each walker is of equal "importance". Once the walkers are more or less of equal weight, the selection becomes redundant and is, therefore, switched off, avoiding unnecessary computational effort.

5.6 Discussion

In the present contribution, we have demonstrated the applicability of the MW-ABF method to a prototypical biomolecular system. Importance sampling techniques are often held back by the difficulty of choosing good reaction coordinates. If the RC is chosen poorly, one is likely to encounter parallel valleys separated by large free-energy barriers, thereby making sampling at a fixed point along the RC very difficult. In such an event, a standard single–walker ABF simulation would lead to slow convergence, as was shown here. The system is biased only in the direction of the RC and, therefore, would be likely to linger in one valley for a long time before reaching another. We have shown that such shortcomings can be elegantly overcome using multiple walkers, through the proposed MW-ABF method. We emphasize that the use of multiple walkers is particularly beneficial when the choice of the model RC is suboptimal, where improvement has been demonstrated herein at constant CPU time. For a well chosen RC, the MW-ABF is not guaranteed to outperform single–walker ABF simulations at a fixed total CPU cost, but still has the advantage



(a) (b)

Figure 5.8: **Results for ξ ranging from 12 to 32 Å using 16 walkers (results after 0.25 ns).** Comparing results between a 16–walker run with (dotted lines) and without selection (dashed-dotted lines). The curves represent averages of 20 independent ABF simulations, and the error bars are 95% confidence intervals. Reference curves are shown as solid lines. (a) The sampling along ξ shows that simulations *with* selection provide a much more uniform distribution along the RC. (b) Mean force approximations and free energy difference profiles (inset): the free-energy profile for the simulation *with* selection is already very close to the reference curve.

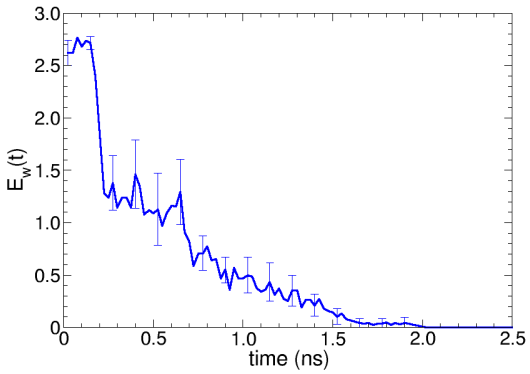


Figure 5.9: **Relative entropy of weights.** It can be seen that the $R = 16$ walkers are approximately of equal weight after about 1.5 ns of an ABF simulation. The selection is switched off after $E_w(t) < \varepsilon \log(16)$, where $\varepsilon = 0.05$. For these simulations, the selection constant in (5.23) is chosen as $c = 0.0001$.

of being easily parallelized. The selection process introduced herein can be employed profitably when encountering pronounced free-energy barriers along the RC. In the presence of parallel valleys, attention must be paid to avoid degeneration of weights, as this could lead to many walkers being kinetically trapped in the same valley, losing the main interest of the use of multiple walkers.

Chapter 6

Long-time convergence of an ABF method

Ce chapitre reprend l'intégralité d'un article écrit en collaboration avec Tony Lelièvre, et soumis à *Archive for Rational Mechanics and Analysis* [Lelièvre 2010a].

Sommaire

6.1 Introduction 101
6.1.1 The adaptive biasing force method 103
6.1.2 Existing convergence results, and the multiple channel scenario 104
6.2 The bi-channel model and statement of the main results . . 105
6.2.1 The bi-channel model 106
6.2.2 A partial differential equation formulation 107
6.2.3 Entropy and Fisher information 111
6.2.4 The free energy as a bias in each channel 114
6.2.5 Main result 115
6.3 Proof of main result 116
6.3.1 Preliminary computations on the total entropy E 116
6.3.2 Controlling E_M 122
6.3.3 Controlling E_m 122
6.3.4 Controlling E_c 125
6.3.5 Completing the proof 130

6.1 Introduction

We consider a system of N particles with positions $q \in \mathcal{D} \subset \mathbb{R}^{3N}$. In statistical physics, one is interested in calculating averages with respect to the Boltzmann-Gibbs measure

$$d\nu(q) = Z^{-1}\exp(-\beta V(q))dq, \tag{6.1}$$

with $V : \mathcal{D} \to \mathbb{R}$ the potential energy function, $Z = \int_{\mathcal{D}} \exp(-\beta V(q))\, dq$ the partition function, $\mathcal{D} = \{q, V(q) < \infty\}$ the configuration space, and $\beta = 1/(k_B T)$, where k_B is the Boltzmann constant and T is the temperature. The function V is the

energy associated with the positions of the particles and is assumed to be such that $Z < \infty$. The probability measure ν represents the equilibrium measure sampled by the particles in the canonical ensemble. A typical dynamics that can be used to sample this measure through trajectorial averages is the overdamped Langevin dynamics

$$dQ_t = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t, \tag{6.2}$$

where $(B_t)_{t\geq 0}$ is a $3N$-dimensional standard Brownian motion. Indeed, under loose assumptions on V, one has the ergodic property: for any smooth test function φ,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(Q_t)\,dt = \int_{\mathcal{D}} \varphi\,d\nu.$$

The efficiency of this sampling procedure, which can be shown to be related to the convergence rate to equilibrium of the above dynamics is often hindered by *metastabilities in the potential function* V, namely regions of low energy are separated by high energy barriers. Equivalently, in terms of the probability measure, ν is typically a *multimodal measure*, with regions of high probability separated by regions of low probability. To circumvent this issue, a one-dimensional collective variable (or reaction coordinate) $\xi : \mathcal{D} \to \mathcal{M}$ is introduced, which will be used to define a biasing potential for (6.2). In the following, we will assume that $|\nabla\xi| > 0$ on $\mathcal{D}$, and that $\mathcal{M} = \mathbb{T}$, where $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is the one-dimensional torus (which typically corresponds to the case where the reaction coordinate represents a dihedral angle, for example to characterize the conformation of a molecule). Before defining more precisely the biased dynamics in the next section, we need to introduce a few notation.

The image of the measure ν in ξ is given by

$$d\nu^\xi(z) = Z^{-1}\exp(-\beta A(z))\,dz,$$

where A is the so-called *free energy*, defined by

$$A(z) = -\beta^{-1}\ln(Z_z) \tag{6.3}$$

where

$$Z_z = \int_{\Sigma_z} \exp(-\beta V(q))\,\delta_{\xi(q)-z}(dq) \tag{6.4}$$

is the partition function on the submanifold $\Sigma_z = \{q \in \mathcal{D} \mid \xi(q) = z\}$. The measure $\delta_{\xi(q)-z}(dq)$ is defined through the conditioning formula: for any smooth test function $\varphi : \mathcal{D} \to \mathbb{R}$,

$$\int_{\mathcal{D}} \varphi(q)\,dq = \int_{\mathcal{M}}\int_{\Sigma_z} \varphi(q)\delta_{\xi(q)-z}(dq)\,dz.$$

Using the co-area formula (see [Lelièvre 2008, Appendix A]), one can also identify this measure as $\delta_{\xi(q)-z}(dq) = |\nabla\xi|^{-1}d\sigma_{\Sigma_z}$, where σ_{Σ_z} is the Lebesgue measure on Σ_z. We assume in the following that ξ and V are such that $Z_z < \infty$, for all $z \in \mathcal{M}$.

Practitioners are typically interested in free energy differences $A(z) - A(z_0)$, which can be obtained by computing (and integrating) the derivative $A'(z)$, the so-called *mean force*

$$A'(z) = Z_z^{-1} \int_{\Sigma_z} f(q) \exp(-\beta V(q)) \, \delta_{\xi(q)-z}(dq) \tag{6.5}$$

where f is the *local mean force* defined by

$$f = \frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right). \tag{6.6}$$

The function f can be understood as the negative force projected onto the reaction coordinate, plus some correction term related to the curvature of the submanifolds Σ_z. Notice that the mean force (6.5) is in fact a conditional expectation

$$A'(z) = \mathbb{E}_\nu[f(Q)|\xi(Q) = z] = \int_{\Sigma_z} f d\nu_{|z}, \tag{6.7}$$

where

$$d\nu_{|z} = \frac{\exp(-\beta V(q)) \delta_{\xi(q)-z}(dq)}{\exp(-\beta(A(z)))}$$

denotes the probability measure ν conditioned to a fixed value z of $\xi(q)$. This measure is supported on the submanifold Σ_z. For the derivation of (6.5)–(6.6)–(6.7) which is again based on the co-area formula, the reader is referred to [Ciccotti 2008, Sprik 1998, den Otter 1998].

6.1.1 The adaptive biasing force method

The adaptive biasing force (ABF) method [Darve 2001, Hénin 2004] uses an estimate of the mean force A' to bias the standard overdamped Langevin dynamics (6.2) in order to overcome metastabilities in ξ. The bottom line of the approach is that it should be easier to sample the probability measure with density proportional to $\exp(-\beta(V - A \circ \xi))$ than the original Boltzmann-Gibbs measure ν, since the marginal probability of the former in ξ is a uniform probability measure on $\mathbb{T}$, while the marginal of the latter (namely $\exp(-\beta A(z))\, dz$) is typically multimodal.

The ABF dynamics is given by

$$\begin{cases} dX_t = -\nabla(V - A_t \circ \xi)(X_t)dt + \sqrt{2\beta^{-1}} dB_t, \\ A_t'(z) = \mathbb{E}\left[f(X_t)|\xi(X_t) = z\right], \end{cases} \tag{6.8}$$

where A_t' is an on-the-fly estimate of the mean force, which, in view of the definition (6.7), is expected to be a good estimate of A'. The law of X_t has density $\psi(t, \cdot)$, which satisfies the non-linear Fokker-Planck equation:

$$\begin{cases} \partial_t \psi = \mathrm{div}(\nabla(V - A_t \circ \xi)\psi) + \beta^{-1}\Delta\psi, \\ A_t'(z) = \dfrac{\displaystyle\int_{\Sigma_z} f \ \psi \ \delta_{\xi(q)-z}(dq)}{\displaystyle\int_{\Sigma_z} \psi \ \delta_{\xi(q)-z}(dq)}. \end{cases} \tag{6.9}$$

Roughly speaking, the biasing force $\nabla(A_t \circ \xi)$ "flattens the free-energy barriers in ξ". To support this claim, let us simply indicate that if $|\nabla\xi|$ is constant, the marginal density in ξ satisfies a simple heat equation, with zero bias, see [Lelièvre 2007b, Lelièvre 2008] and also Equation (6.16) below. Existence and uniqueness of solutions to (6.8) are studied in [Jourdain 2009] and a study of the longtime convergence of (6.9) can be found in [Lelièvre 2008], the results of which are briefly discussed below.

6.1.2 Existing convergence results, and the multiple channel scenario

It has been shown in [Lelièvre 2008] that, under appropriate assumptions, the biasing force A'_t in (6.8) (actually for a slightly different dynamics which reduces to (6.8) if $|\nabla\xi|$ is constant for example) converges to the mean force A' exponentially fast in the longtime limit. The rate of convergence was estimated as the minimum of (i) the rate at which the law of $\xi(X_t)$ converges to equilibrium, and (ii) the smallest logarithmic Sobolev inequality constant (LSI constant, discussed in Section 6.2.3) of the conditional probability measures $\nu_{|z}$, for $z \in \mathcal{M}$. Thanks to the bias in the direction of the reaction coordinate, it can be shown that the rate of convergence of the marginal in ξ is actually not the limiting rate in practice since it satisfies a simple diffusion equation. The real limitation is thus the metastable features (*i.e.* the multimodality) of the family of laws $\nu_{|z}$, which is quantified through the LSI constants associated to these measures: roughly speaking, the smaller the constant the more multimodal the probability measure. These constants may be in some cases extremely small, at least for some values of $z \in \mathcal{M}$. The question we address in this paper is the *optimality* of this theoretical rate of convergence for the ABF method.

The generic situation is indeed that the LSI constants for the measures $\nu_{|z}$ are not small uniformly in z. Typically, there exists some values of z for which these measures are multimodal. This situation is often encountered in practice due to two reasons. First, finding a suitable reaction coordinate, namely in our context one that ensures that there are no metastabilities associated to the equilibrium measures $\nu_{|z}$, is not trivial for a large-dimensional system. Secondly, a low-dimensional reaction coordinate may simply not be sufficient to describe all metastabilities of the system. In such cases, the results of [Lelièvre 2008] predict a very small rate of convergence for the ABF dynamics (6.8), and thus, the inefficiency of this biasing procedure.

As a typical case for which such difficulties appear, we will consider in the following the so-called 'multiple-channel situation' (see Figure 6.1 on the left, for a bi-channel case). In such a situation, starting from a metastable state and as the system evolves in the direction of increasing values of the reaction coordinate ξ, it can follow different channels, which are separated (in the 'orthogonal direction to ξ') by arbitrary high energy barriers. In other words, the energy landscape features parallel valleys which are orthogonal to the isocontours of the reaction coordinate. In such a prototypical situation, the conditional probability measures $\nu_{|z}$ are indeed

multimodal, for the values of z corresponding to the system being in one of these channels.

However, recent numerical experiments [Minoukadeh 2010] (based on a discretization of the ABF dynamics (6.8) by multiple walkers simulated in parallel) suggest that in fact, high energy barriers in Σ_z do not always hinder the convergence of the ABF method. The multiple walkers are made to follow similar dynamics (6.8), but driven by independent Brownian motions. The chemical system considered in [Minoukadeh 2010] is the compact states of the deca-alanine peptide (the reaction coordinate ξ being thus the end-to-end distance of the peptide). Due to some 'buckling effects', this is a typical multiple channel situation, since the molecule can shrink to various compact states (see [Hénin 2010]). In [Minoukadeh 2010], numerical results show that the ABF approach indeed yields reliable results in such a situation. We interpret this as follows. When encountered with a fork in the channel, each walker is likely to travel down a different channel. Thus, it is indeed almost impossible for a given walker to switch from one channel to another, once it has entered one of them, but this appears not to be necessary to obtain reliable results. It suggests that the theoretical rate of convergence obtained in [Lelièvre 2008] is actually not optimal.

Inspired by these numerical results, we present herein an improved theoretical rate of convergence of the ABF method. The rate will be shown to depend on the LSI constants of the family of equilibrium measures conditioned to being in Σ_z *and* a channel. By doing so, we show that high energy barriers separating the channels do not in fact affect the rate of convergence of the method. The crucial assumptions needed to show our result are: (i) exchange between the two channels is possible for *some* values of ξ (see [H1] below) and (ii) the free energy is a 'good bias' in each channel (see [H4] below). This is formalized in the main result of this paper, namely Theorem 5 below.

For some technical reasons, we were actually unable to prove this result on the original ABF dynamics (6.8). We will thus consider a slightly different system (that we call *the bi-channel model*) which retains the most important features of the dynamics (6.8) when applied to a potential exhibiting two parallel channels in the direction of ξ, separated by a high energy barrier.

The paper is organized as follows. In Section 6.2 we give details of the bi-channel model, define some probability measures and recall some entropy definitions before presenting the main result. Finally, the proof of the main result is given in Section 6.3.

6.2 The bi-channel model and statement of the main results

In this section, we present a model to describe the bi-channel scenario. In the following, we treat the case $d = 2$ (so that the position is $q = (x, y)$), $\mathcal{D} = \mathbb{T} \times \mathbb{R}$, $\mathcal{M} = \mathbb{T}$ and $\xi : \mathcal{D} \to \mathbb{T}$, where $\xi(x, y) = x$. We further assume without loss

of generality that $\beta = 1$, which can be done by a change of variables $\tilde{t} = \beta^{-1}t$, $\tilde{\psi}(\tilde{t}, x) = \psi(t, x)$, $\tilde{V}(x, y) = \beta V(x, y)$. With these assumptions, some notation may be simplified: $|\nabla\xi| = 1$, $\Sigma_z = \mathbb{R}$ and $\delta_{\xi(q)-z}(dq) = dy$. Furthermore, the definition of the local mean force in (6.6) simplifies to

$$f = \partial_x V$$

and the free energy and its derivative are given by

$$A(x) = -\ln \int_{\mathbb{R}} \exp(-V(x, y))\, dy \quad \text{and} \quad A'(x) = \frac{\displaystyle\int_{\mathbb{R}} \partial_x V(x, y) \exp(-V(x, y))\, dy}{\displaystyle\int_{\mathbb{R}} \exp(-V(x, y))\, dy}. \tag{6.10}$$

We would like to emphasize that the choice of the domain $\mathcal{M} = \mathbb{T}$ and reaction coordinate $\xi(x, y) = x$ is merely to reduce technicalities, see [Lelièvre 2008] for appropriate tools to treat general ξ. In particular, the results can be straightforwardly generalized to the case $\mathcal{D} = \mathbb{T} \times \mathbb{R}^{d-1}$ and $\xi(q_1, \ldots, q_d) = q_1$. Likewise, the generalization to a situation with multiple channels (more than two) is straightforward (see Remark 7 below for another generalization).

6.2.1 The bi-channel model

The bi-channel situation is characterized by the existence of two channels joining an initial and final state on a potential energy surface V, separated from each other by a region of high energy, as depicted in the left of Figure 6.1. As explained above, we were not able to analyze the original ABF dynamics (6.8) in this situation, because of some technical difficulties in expressing the probability density flux from one channel to the other.

We therefore analyze the convergence of the ABF method for a slightly different model, which is schematically represented in right of Figure 6.1. Each channel is described by a potential energy function $V_i : \mathcal{D} \to \mathbb{R}$, where $i \in \{0, 1\}$ denotes the channel index. The stochastic process we consider now is actually a couple (X_t, I_t), where the position vector X_t lives at time t on the potential V_{I_t}, $I_t \in \{0, 1\}$ being the index of the visited channel at time t. The channel index I_t is allowed to switch to $1 - I_t$ only if $\xi(X_t)$ lies in some designated regions (typically at the two ends of the two channels).

The dynamics for the ABF dynamics in the bi-channel model is then

$$\begin{cases} dX_t = -\nabla(V_{I_t} - A_t \circ \xi)(X_t)dt + \sqrt{2}\, dB_t, \\ A'_t(x) = \mathbb{E}\left[\partial_x V_{I_t}(X_t) | \xi(X_t) = x\right], \\ I_t \in \{0, 1\} \text{ is a jump process with generator} \\ \qquad L\varphi(x, y, i) = -\lambda(x)(\varphi(x, y, i) - \varphi(x, y, 1 - i)). \end{cases} \tag{6.11}$$

In terms of the stochastic process I_t, switching between the two potentials (namely

change of I_t to $1 - I_t$) occurs at times

$$\tau_{n+1} = \inf \left\{ t > \tau_n \;\middle|\; \int_{\tau_n}^{t} \lambda(\xi(X_s))ds > T_n \right\},$$

where $\tau_0 = 0$ and T_n are i.i.d. random variables drawn from the exponential distribution with parameter 1. In this way $\lambda(x)$ denotes the rate at which the trajectories jump from potential V_i to potential V_{1-i}. Note that this rate depends only on the position, x, in the reaction coordinate. The bi-channel feature of the model is related to the fact that we assume that the rate λ is zero (there is no switching) in some region of the reaction coordinate. Outside of this region, the rate is supposed to be constant, and the potential functions are assumed to be identical (the particles live in the same potential). Let us state this as a formal assumption,

$$\textbf{[H1]} \quad \exists\, \mathcal{E} \subset \mathbb{T},\ \ \lambda(x) = \lambda \mathbf{1}_{\mathbb{T}\setminus\mathcal{E}}(x) \text{ and } \forall x \in \mathbb{T}\setminus\mathcal{E}, V_0(x,\cdot) = V_1(x,\cdot).$$

The region $\mathcal{E} \subset \mathbb{T}$ in the above hypothesis represents the region where the two channels are separated by high energy barriers, see Figure 6.1. It is assumed to have a Lebesgue measure different from 0 and 1.

The main qualitative difference between the bi-channel model we study, and the original ABF dynamics (6.8) is that the switching only depends on the x-position and not on the y-position. However, the ABF dynamics (6.11) conserves the main difficulty of the original one, namely the metastability of the dynamics in terms of visited channels for some values of ξ. At times t such that $\xi(X_t) \in \mathcal{E}$, I_t cannot switch to $1 - I_t$. In particular, it can be checked that the proof of [Lelièvre 2008] applied to (6.11) in the case $\mathcal{E} = \emptyset$ leads to an estimated rate of convergence limited by λ and is thus eventually zero if λ goes to zero (see Remark 9 below). The aim of this work is to study the case $\mathcal{E} \neq \emptyset$ used to obtain exponential rate of convergence even if $\lambda = 0$ in some region (see Theorem 5 below).

Remark 7 *On Figure 6.1, we represent the bi-channel case with* two *metastable states linked by two different channels. The region where the stochastic process can jump from one channel to another has two connected components. We would like to emphasize that our result also applies to the case where this region is simply connected, namely a situation where two channels start from a metastable state (along the reaction coordinate value) but do not end in another metastable state. This situation is of course less favorable in terms of speed of convergence to equilibrium, which would be reflected in our theoretical result through the parameter θ (see [H4] below). This is actually typically the situation of the numerical experiments in [Minoukadeh 2010] mentioned above since the various compact states obtained do not belong to the same metastable basin.*

6.2.2 A partial differential equation formulation

Let us introduce the time marginal of the process (X_t, I_t) in (6.11):

$$d\mu_t = \sum_{i=0}^{1} \psi(t,x,y,i)\delta_i \, dx \, dy,$$

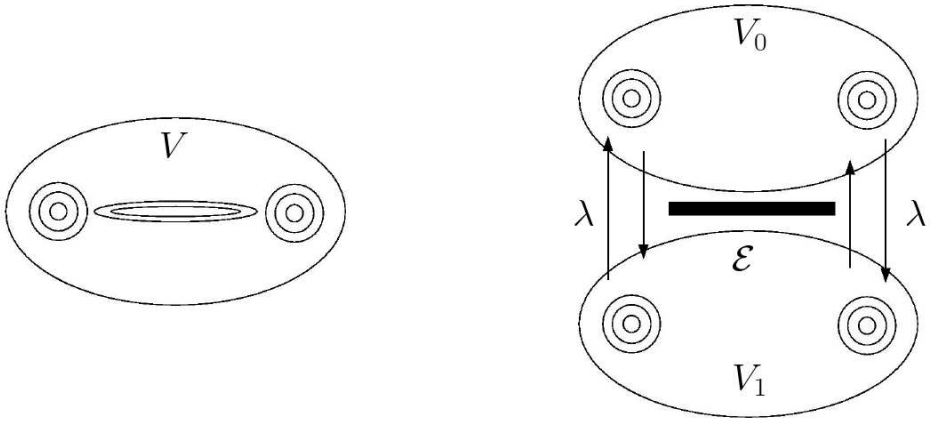


Figure 6.1: Left: Contour plot of a 2-dimensional potential energy surface demonstrating the bi-channel scenario. Right: In the bi-channel model the two channels are described by two potential functions $V_i : \mathcal{D} \to \mathbb{R}$, $i \in \{0, 1\}$. Exchange between the two channels is allowed only in regions $\mathbb{T}\backslash\mathcal{E}$ at a rate $\lambda > 0$.

where δ_i is the Dirac measure on the singleton $\{i\}$. When necessary, we shall denote the i-dependency of the density by a subscript ψ_i, so that ψ_{1-i} denotes $\psi(t, x, y, 1-i)$ for example. The evolution of the densities are described by a system of non-linear partial differential equations: $\forall i \in \{0, 1\}$,

$$\partial_t \psi = \mathrm{div}(\nabla(V_i - A_t \circ \xi)\psi) + \Delta\psi - \lambda \circ \xi \, [\psi - \psi_{1-i}] \text{ on } \mathcal{D}, \tag{6.12}$$

where, we recall, the last term is zero for $\xi(x, y) \in \mathcal{E}$. The non-linearity is due to the definition of the mean force estimate, given by

$$A'_t(x) = \frac{\displaystyle\sum_{i=0}^{1} \int_{\mathbb{R}} \partial_x V_i(x, y) \; \psi(t, x, y, i) \; dy}{\displaystyle\sum_{i=0}^{1} \int_{\mathbb{R}} \psi(t, x, y, i) \; dy}. \tag{6.13}$$

Using hypothesis [H1], it can be checked that if ψ_∞ is defined as a probability density proportional to $e^{-(V_i - A_\infty \circ \xi)}$ where A_∞ is a given long-time limit for A_t, then ψ_∞ is a stationary solution to (6.12). Then, by replacing ψ by ψ_∞ in (6.13) and comparing with the definition of the mean force (6.10), it is clear that by choosing $A'_\infty = A'$, one obtains a stationary solution of the system (6.12)–(6.13) written as:

$$\psi_\infty(x, y, i) = \frac{e^{-(V_i(x,y) - A(x))}}{\displaystyle\sum_{i=0}^{1} \int_{\mathcal{D}} e^{-(V_i - A \circ \xi)} \; dx \; dy}. \tag{6.14}$$

Measure	Definition	Description of measure
$d\mu_t$	$\sum_{i=0}^{1} \psi(t,x,y,i)\delta_i \; dx \; dy$	Probability measure in domain $\mathcal{D} \times \{0,1\}$
$d\mu_t^{\xi,I}$	$\sum_{i=0}^{1} \psi^{\xi,I}(t,x,i)\delta_i \, dx$	Marginal measure in reaction coordinate and channel
$d\mu_t^{\xi}$	$\psi^{\xi}(t,x) \, dx$	Marginal measure in reaction coordinate
$d\mu_{t\|x,i}$	$\frac{\psi(t,x,y,i) \; dy}{\psi^{\xi,I}(t,x,i)}$	Measure conditioned to being at x and i
$d\mu_{t\|x}$	$\frac{\sum_{i=0}^{1} \psi(t,x,y,i)\delta_i \; dy}{\psi^{\xi}(t,x)}$	Measure conditioned to being at x
$d\mu_{t\|x}^{I}$	$\frac{\sum_{i=0}^{1} \psi^{\xi,I}(t,x,i)\delta_i}{\psi^{\xi}(t,x)}$	Marginal in I, conditioned to being at x

Table 6.1: A list of probability measures.

The associated equilibrium measure for the process (X_t, I_t) writes:

$$d\mu_\infty = \sum_{i=0}^{1} \psi_\infty(x,y,i)\delta_i \; dx \; dy.$$

We will need further notation for marginal and conditional laws associated to μ_t and μ_∞. The next two sections give precise definitions for these measures, and Table 6.1 summarizes the notation.

6.2.2.1 Marginal laws

We are now in the position to define marginal probability densities. The image of the probability measure μ_t in ξ and I is denoted by

$$d\mu_t^{\xi,I} = \sum_{i=0}^{1} \psi^{\xi,I}(t,x,i)\delta_i \, dx \quad \text{where} \quad \psi^{\xi,I}(t,x,i) = \int_{\mathbb{R}} \psi(t,x,y,i) \; dy.$$

The evolution of the density $\psi^{\xi,I}$ is described by the (non-closed) partial differential equation

$$\partial_t \psi^{\xi,I} = \int_{\mathbb{R}} \partial_x(\partial_x(V_i - A_t \circ \xi)\psi) \; dy + \partial_{xx}\psi^{\xi,I} - \lambda(x)(\psi^{\xi,I} - \psi_{1-i}^{\xi,I}), \tag{6.15}$$

obtained by integrating (6.12) in y. The associated equilibrium measure is

$$d\mu_\infty^{\xi,I} = \sum_{i=0}^{1} \psi_\infty^{\xi,I}(x,i)\delta_i \, dx \quad \text{where} \quad \psi_\infty^{\xi,I}(x,i) = \int_{\mathbb{R}} \psi_\infty(x,y,i) \, dy.$$

The marginal measure in ξ only is denoted by

$$d\mu_t^\xi = \psi^\xi(t,x) \, dx \quad \text{where} \quad \psi^\xi(t,x) = \sum_{i=0}^{1} \psi^{\xi,I}(t,x,i) = \sum_{i=0}^{1} \int_{\mathbb{R}} \psi(t,x,y,i) \, dy.$$

By summing (6.15) over i and using the definition (6.13) of A_t', it is easy to check that ψ^ξ satisfies a closed, very simple partial differential equation (this is similar to the original ABF dynamics, see [Lelièvre 2007b, Lelièvre 2008]):

$$\partial_t \psi^\xi = \partial_{xx} \psi^\xi \quad \text{on } \mathbb{T}. \tag{6.16}$$

Thanks to the adaptive bias, along ξ, the barriers have been flattened. The long-time limit of ψ^ξ is given by

$$\psi_\infty^\xi(x) = \sum_{i=0}^{1} \int_{\mathbb{R}} \psi_\infty(x,y,i) \, dy = 1,$$

which corresponds to the uniform probability measure on the torus $\mathbb{T}$.

6.2.2.2 Conditional laws

Let us introduce the measure $\mu_{t|x}$ of (X_t, I_t) conditioned to being at a specified point x in the reaction coordinate:

$$d\mu_{t|x} = \frac{\displaystyle\sum_{i=0}^{1} \psi(t,x,y,i)\delta_i \, dy}{\psi^\xi(t,x)}. \tag{6.17}$$

Its long-time limit is

$$d\mu_{\infty|x} = \frac{\displaystyle\sum_{i=0}^{1} \psi_\infty(x,y,i)\delta_i \, dy}{\psi_\infty^\xi(x)}. \tag{6.18}$$

To study the bi-channel model, we will also need to introduce the measures conditioned to being at fixed ξ *and* a particular channel $i \in \{0,1\}$. This measure and its long-time limit are defined respectively by

$$d\mu_{t|x,i} = \frac{\psi(t,x,y,i) \, dy}{\psi^{\xi,I}(t,x,i)} \quad \text{and} \quad d\mu_{\infty|x,i} = \frac{\psi_\infty(x,y,i) \, dy}{\psi_\infty^{\xi,I}(x,i)}. \tag{6.19}$$

Finally, we will also need the marginals in I of the probability measures $\mu_{t|x}$ and $\mu_{\infty|x}$. These Bernoulli probability measures $\mu_{t|x}^I$ and $\mu_{\infty|x}^I$ represent the proportion

of the marginal distribution $\psi^{\xi}(t,\cdot)$ and ψ^{ξ} in each channel. They are formally defined as

$$d\mu^I_{t|x} = \frac{\sum_{i=0}^{1} \psi^{\xi,I}(t,x,i)\delta_i}{\psi^{\xi}(t,x)} \quad \text{and} \quad d\mu^I_{\infty|x} = \frac{\sum_{i=0}^{1} \psi^{\xi,I}_{\infty}(x,i)\delta_i}{\psi^{\xi}_{\infty}(x)}. \tag{6.20}$$

6.2.3 Entropy and Fisher information

In this section we recall some well-known results for defining relative entropy between two probability measures, which can be seen as a measure of the 'distance' between those. A general introduction to this topic can be found in [Ané 2000, Villani 2003] and applications to study the longtime behavior of Fokker-Planck type equations are presented in [Arnold 2001].

Definition 6 ***(Entropy).*** *For any two probability measures μ and ν such that μ is absolutely continuous with respect to ν (denoted as $\mu \ll \nu$), the relative entropy is defined as*

$$H(\mu|\nu) = \int \ln\left(\frac{d\mu}{d\nu}\right) d\mu.$$

The positivity of the relative entropy can be shown using the inequality $\ln(1/x) \geq 1 - x$. Furthermore, $H(\mu|\nu) = 0$ if and only if $\mu = \nu$.

Definition 7 ***(Csiszar-Kullback inequality).*** *For measures μ and ν which have densities with respect to the Lebesgue measure, the following holds*

$$\|\mu - \nu\|_{L^1} \leq \sqrt{2H(\mu|\nu)}.$$

This allows us to control the L^1-norm of the difference of two probability measures by their relative entropy.

Definition 8 ***(Fisher information).*** *For any probability measure μ absolutely continuous with respect to ν, the Fisher information is given by*

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

Definition 9 ***(Logarithmic Sobolev inequality).*** *The probability measure ν is said to satisfy a logarithmic Sobolev inequality with constant $\rho > 0$ (in short: LSI(ρ)) if for all probability measures μ such that $\mu \ll \nu$,*

$$H(\mu|\nu) \leq \frac{1}{2\rho} F(\mu|\nu).$$

Such an inequality holds for Gaussian measures [Gross 1975] for example, and more generally [Bakry 1984] for any measure ν with density proportional to e^{-V}, where V is α-convex (in which case the LSI constant ρ is equal to α). Besides, there exists a perturbation result [Holley 1987]: if $\tilde{\nu}$ is a probability measure such that $d\tilde{\nu}/d\nu = e^U$, where ν satisfies a LSI(ρ) and U is a bounded function, then $\tilde{\nu}$ satisfies a LSI with constant $\tilde{\rho} = \rho\, e^{-\mathrm{osc}(U)}$, where $\mathrm{osc}(U) = \sup(U) - \inf(U)$. Thus, a very large class of probability measures satisfy a LSI. An important feature of the LSI constant is that it degenerates to zero in the case where the probability measure is multimodal. For example, if $d\nu = Z^{-1}\exp(-\beta W(x))\,dx$ and $W(x) = x^4/4 - x^2/2$ is the double-well potential in dimension 1, then the LSI constant scales as $\exp(-\beta\Delta W)$ where $\Delta W = W(0) - W(1) > 0$ is the height of the barrier. Such inequalities thus hold under rather loose assumptions, but the constant ρ is very small for a multimodal measure. For example, in typical situations encountered in molecular dynamics, the LSI constant for the measure ν defined in (6.1) is extremely small, which is related to the fact that it is difficult to sample directly this measure.

To analyze the convergence of the ABF dynamics for the bi-channel model, we make use of LSIs for the conditional measures $\mu_{\infty|x,i}$ (see assumption [H3] below), which are the equilibrium canonical measures, conditioned to being at a fixed value of ξ *and in a channel*. In [Lelièvre 2008], log-Sobolev inequalities for the equilibrium canonical measures conditioned only to being at a fixed value of ξ were considered, but in the bi-channel case, those are typically very small due to the presence of high energy barriers 'orthogonal' to the isocontours of the reaction coordinate (see again Figure 6.1).

Let us now define the Wasserstein distance between two probability measures.

Definition 10 ***(Wasserstein distance).*** *The Wasserstein distance with linear cost between probability measures μ and ν is defined as*

$$W(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{D}\times\mathcal{D}} |y - y'|\ \pi(dy, dy'),$$

where $\Pi(\mu,\nu)$ denotes the set of coupling probability measures on $\mathcal{D}\times\mathcal{D}$, with marginals μ and ν.

Definition 11 ***(Talagrand inequality).*** *The probability measure ν is said to satisfy a Talagrand inequality with constant $\rho > 0$ (or $T(\rho)$) if for all probability measures μ such that $\mu \ll \nu$,*

$$W(\mu,\nu) \leq \sqrt{\frac{2}{\rho} H(\mu|\nu)}. \tag{6.21}$$

Logarithmic Sobolev inequalities and Talagrand inequalities are related (see [Otto 2000]):

Lemma 3 *If ν satisfies LSI(ρ), then ν satisfies $T(\rho)$.*

Below, we present entropies that prove useful in obtaining convergence results of the bi-channel ABF model. In this paper, we are primarily interested in the convergence to a stationary state of the Fokker-Planck equation (6.12)–(6.13), and thus of

the associated partial differential equations (6.15) and (6.16). Relative entropies will therefore be defined for some probability measures with respect to their long-time limits.

The *total entropy* will be denoted by

$$E(t) = H(\mu_t|\mu_\infty) = \sum_{i=0}^{1} \int_{\mathcal{D}} \ln\left(\frac{\psi}{\psi_\infty}\right) \psi \, dx \, dy.$$

The so-called *macroscopic entropy* is defined as

$$E_M(t) = H(\mu_t^\xi|\mu_\infty^\xi) = \int_{\mathbb{T}} \ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right) \psi^\xi \, dx,$$

the *local entropy* at a fixed value x in the reaction coordinate by

$$e_m(t,x) = H(\mu_{t|x}|\mu_{\infty|x}) = \sum_{i=0}^{1} \int_{\mathbb{R}} \ln\left(\frac{\psi}{\psi^\xi} \Big/ \frac{\psi_\infty}{\psi_\infty^\xi}\right) \frac{\psi}{\psi^\xi} \, dy,$$

and the *microscopic entropy* by

$$E_m(t) = \int_{\mathbb{T}} e_m(t,x) \; \psi^\xi(t,x) \, dx.$$

With the above definitions, it is easy to show that

$$E(t) = E_M(t) + E_m(t). \tag{6.22}$$

In order to treat the bi-channel case, we define a *channel-local entropy*, defined by

$$e_{\mathrm{cl}}(t,x,i) = H(\mu_{t|x,i}|\mu_{\infty|x,i}) = \int_{\mathbb{R}} \ln\left(\frac{\psi}{\psi^{\xi,I}} \Big/ \frac{\psi_\infty}{\psi_\infty^{\xi,I}}\right) \frac{\psi}{\psi^{\xi,I}} \, dy.$$

Two hypothesis that will be essential to the results presented below are: a so-called 'bounded coupling' assumption on the cross derivative $\partial_{x,y} V_i$ (see [Lelièvre 2008, Lelièvre 2009]), and an assumption on the logarithmic Sobolev inequality constants for the probability measures $\mu_{\infty|x,i}$. The hypotheses read

[H2] $\left\{ \begin{array}{c} \forall i \in \{0,1\}$, V_i and ξ are sufficiently differentiable functions such that $\exists \, C, M > 0$ \\ $\|\partial_{x,y} V_i\|_{L^\infty(\mathbb{T}\times\mathbb{R})} \leq M < \infty$ and $\left\| \frac{\int_{\mathbb{R}} \partial_x V_i e^{-V_i} \, dy}{\int_{\mathbb{R}} e^{-V_i} \, dy} \right\|_{L^\infty(\mathbb{T})} \leq C < \infty$, \end{array} \right.$

[H3] $\left\{ \begin{array}{c} \forall i \in \{0,1\}$, V_i and ξ are such that $\exists \rho > 0$, $\forall x \in \mathbb{T}$, $\forall i \in \{0,1\}$, \\ the conditional measure $\mu_{\infty|x,i}$ satisfies LSI(ρ). \end{array} \right.$

The hypothesis [H3] gives us that $\forall x \in \mathbb{T}, \forall i \in \{0,1\}$

$$H(\mu_{t|x,i}|\mu_{\infty|x,i}) \quad \leq \quad \frac{1}{2\rho} F(\mu_{t|x,i}|\mu_{\infty|x,i}),$$

or equivalently

$$\int_{\mathbb{R}} \ln\left(\frac{\psi}{\psi^{\xi,I}}\Big/\frac{\psi_\infty}{\psi_\infty^{\xi,I}}\right)\frac{\psi}{\psi^{\xi,I}}\,dy \le \frac{1}{2\rho}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\frac{\psi}{\psi^{\xi,I}}\,dy. \tag{6.23}$$

Finally, an entropy that appears in later calculations, is that of the Bernoulli measure $\mu^I_{t|x}$ with respect to its long-time limit

$$e_c(t,x) = H(\mu^I_{t|x}|\mu^I_{\infty|x}) = \sum_{i=0}^{1}\ln\left(\frac{\psi^{\xi,I}}{\psi^\xi}\Big/\frac{\psi_\infty^{\xi,I}}{\psi_\infty^\xi}\right)\frac{\psi^{\xi,I}}{\psi^\xi}.$$

The so-called *channel entropy* is then given by

$$\begin{aligned} E_c(t) &= \int_{\mathbb{T}} e_c(t,x)\psi^\xi\,dx \\ &= \sum_{i=0}^{1}\int_{\mathbb{T}}\ln\left(\frac{\psi^{\xi,I}}{\psi^\xi}\Big/\frac{\psi_\infty^{\xi,I}}{\psi_\infty^\xi}\right)\psi^{\xi,I}. \end{aligned} \tag{6.24}$$

6.2.4 The free energy as a bias in each channel

As well as the hypothesis [H3] on the conditional measures $\mu_{\infty|x,i}$, an assumption will also be necessary to ensure that the free energy is a 'good bias' *in each channel*. More precisely, once the bias has converged, the marginals along ξ in each channel must converge sufficiently quickly to their long-time limit. Roughly speaking, the channels should not be too 'asymmetrical'. The aim of this section is to state this more formally, see assumption [H4] below.

Consider that the system is already nearly at equilibrium, in the sense that

$$A'_t = A'_\infty = A',$$

and

$$\forall i \in \{0,1\},\quad \int_{\mathbb{R}}\partial_x V_i\,d\mu_{t|x,i} = \int_{\mathbb{R}}\partial_x V_i\,d\mu_{\infty|x,i}.$$

Then the marginal density $\psi^{\xi,I}$ can be shown to satisfy (see (6.15) and (6.40) below)

$$\partial_t\psi^{\xi,I} = -\mathcal{L}_i\psi^{\xi,I} = \partial_x\left(\psi_\infty^{\xi,I}\partial_x(\psi^{\xi,I}/\psi_\infty^{\xi,I})\right) - \lambda(x)(\psi^{\xi,I} - \psi^{\xi,I}_{1-i}). \tag{6.25}$$

It can be shown that the operator $\mathcal{L} = (\mathcal{L}_0, \mathcal{L}_1)$ is symmetric and positive definite with respect to the inner product $\langle f, g\rangle = \sum_{i=0}^{1}\int_{\mathbb{T}} f_i(x)g_i(x)\,\frac{1}{\psi_\infty^{\xi,I}(x,i)}\,dx$ and has a spectral gap $\theta > 0$ (see Lemma 14 below). In other words, $\mathcal{L}$ is such that for all functions $f : \mathbb{T}\times\{0,1\} \to \mathbb{R}$, with $f_i \in H^1\left(\frac{1}{\psi_\infty^{\xi,I}(x,i)}dx\right)$, $f \neq 0$ and $\sum_{i=0}^{1}\int_{\mathbb{T}} f_i(x)\,dx = 0$, we have

$$\langle f, f\rangle \le \frac{1}{\theta}\langle f, \mathcal{L}f\rangle. \tag{6.26}$$

As will be apparent in the proof, it will be necessary for the spectral gap θ to be sufficiently large.

[H4] $\left\{ \begin{array}{c} \theta > \theta_{\min} \text{ with } \theta_{\min} = \dfrac{8(C + M\rho^{-1/2})^2 \tilde{M}}{c} \\ \text{where } 0 < c, \tilde{M} < \infty \text{ are such that } \inf\limits_{x,i} \psi_\infty^{\xi,I} = c \text{ and } \sup_x \psi^\xi(0,x) = \tilde{M}. \end{array} \right.$

We recall that the constants C, M and ρ have been introduced in assumptions [H2] and [H3] above. The fact that $c > 0$ is not restrictive since $\forall i \in \{0,1\}$, V_i is a continuous function and $\psi_\infty^{\xi,I} \propto \int_{\mathbb{R}} e^{-V_i}\, dy > 0$ is continuous and defined on the compact space $\mathbb{T}$. Similarly, since ψ^ξ satisfies the heat equation (6.16), the assumption $\tilde{M} < \infty$ is not restrictive. If, for example, the initial condition has a Dirac mass marginal in ξ, for any positive time $t_0 > 0$, $\psi^\xi(t_0, \cdot)$ is a bounded function, and one has simply to consider the dynamics on $[t_0, \infty)$.

Remark 8 *In the hypothesis [H2], the assumption on the cross derivative $\partial_{x,y} V_i$ could in fact be replaced by $\|\partial_x V_i\|_{L^\infty(\mathbb{T}\times\mathbb{R})} \leq M < \infty$, in which case, in [H4], the minimum value for θ would be changed to $\theta_{\min} = \frac{8M^2\tilde{M}}{c}$ and the Talagrand inequalities in Lemmas 8 and 13 would be replaced by Csiszar-Kullback inequalities.*

6.2.5 Main result

We are now in position to present the main result of the paper.

Theorem 5 *Assume hypotheses [H1]-[H4]. There exists a smooth function $\Lambda : (\theta_{\min}, \infty) \to (0, \rho)$ which is increasing and such that:*

$$\Lambda(\rho + 2\theta_{\min}) = \frac{\rho}{2} \quad \textit{and} \quad \Lambda(\theta) \to \left\{ \begin{array}{ll} 0 & \textit{as} \quad \theta \to \theta_{\min} \\ \rho & \textit{as} \quad \theta \to \infty \end{array} \right.$$

for which we can prove the following convergence result: for any $\varepsilon \in (0, \Lambda(\theta))$, there exists a constant $K > 0$ such that, $\forall t > 0$,

$$E_m(t) \leq K \exp\left(-2 \min\{(\Lambda(\theta) - \varepsilon), 4\pi^2\}\, t\right). \tag{6.27}$$

This implies that the total entropy E and thus $\|\psi(t,\cdot) - \psi_\infty\|_{L^1}^2$ converge exponentially fast to zero with the same rate. Furthermore, the biasing force A'_t converges to the mean force A' in the following sense: $\forall t \geq 0$,

$$\int_{\mathbb{T}} |A'_t(x) - A'(x)|^2 \psi^\xi(t,x)\, dx \leq 2(C + M\rho^{-1/2})^2\, E_m(t).$$

As a consequence, for any positive time $t_0 > 0$ and $\varepsilon \in (0, \Lambda(\theta))$, there exists a constant $\bar{K}$ such that $\forall t \geq t_0$,

$$\int_{\mathbb{T}} |A'_t(x) - A'(x)|^2\, dx \leq \bar{K} \exp\left(-2 \min\{(\Lambda(\theta) - \varepsilon), 4\pi^2\}\, t\right).$$

The term $4\pi^2$ corresponds to the exponential rate of convergence of ψ^ξ to ψ^ξ_∞ (see (6.16)), and is clearly not the bottleneck. There are actually various ways to make this rate as small as needed (see Remark 11 in [Lelièvre 2008]).

Thus, this result essentially shows that the ABF method converges at a rate which is limited by the multimodality of the equilibrium canonical measures conditioned to being at fixed value of ξ *and in a channel* (quantified by the constant ρ), if the free energy is a bias which enables a fast exploration of each channel (this is quantified by the constant θ, which should be sufficiently large for $\Lambda(\theta)$ to be indeed close to ρ). Thus, the convergence may be fast even if switching between the two channels is impossible for some values of ξ. If the spectral gap θ is sufficiently large, we thus recover a similar expression for the rate of convergence of the ABF method as the one derived in [Lelièvre 2008], with ρ being now the LSI constant of the canonical measures $\mu_{|x,i}$.

We would like to emphasize that our arguments hold under the following assumption of existence of regular solutions: We assume that we are given a process (X_t, I_t) and a function A'_t which satisfy (6.11), and such that X_t, conditionally on $I_t = i$ has a smooth density $\psi(t, x, y, i)$. We suppose that this density is sufficiently regular so that the entropy estimates below are valid. We refer for example to [Arnold 2001] for an appropriate functional framework in which such entropy estimates hold.

6.3 Proof of main result

In order to prove the exponential decay of the microscopic entropy in Theorem 5, we use the fact that the time evolution of the microscopic entropy can be expressed as a combination of the evolution of the total and the marginal entropies, from (6.22)

$$\frac{dE_m}{dt} = \frac{dE}{dt} - \frac{dE_M}{dt}.$$

In order to obtain results for the microscopic entropy, we begin by treating the time evolutions of the total entropy and the macroscopic entropy separately.

6.3.1 Preliminary computations on the total entropy E

In order to study the evolution of the total entropy, some auxiliary results will be needed and are given in the lemmas below. First, it will be useful to write the Fokker-Planck equation associated to ψ in a different form.

Lemma 4 *The Fokker-Planck equation (6.12) for ψ can be rewritten as*

$$\partial_t \psi = \mathrm{div}(\psi_\infty \nabla(\psi/\psi_\infty)) + \partial_x((A' - A'_t)\psi) - \lambda(x)(\psi - \psi_{1-i}).$$

Proof : By developing the right hand side, we obtain

$$\begin{aligned}
\partial_t \psi &= \mathrm{div}\left(\nabla\psi - \frac{\psi}{\psi_\infty}\nabla\psi_\infty\right) + \partial_x((A' - A'_t)\psi) - \lambda(x)(\psi - \psi_{1-i}) \\
&= \mathrm{div}\left(\nabla\psi + \nabla(V_i - A \circ \xi)\psi\right) + \partial_x(A'\psi) - \partial_x(A'_t\psi) - \lambda(x)(\psi - \psi_{1-i}) \\
&= \mathrm{div}\left(\nabla(V_i - A_t \circ \xi)\psi\right) + \Delta\psi - \lambda(x)(\psi - \psi_{1-i}),
\end{aligned}$$

which is indeed the Fokker-Planck equation (6.12). ♢

The above may now be used to estimate the time evolution of the total entropy.

Lemma 5 *The total entropy E satisfies*

$$\frac{dE}{dt} \leq -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\nabla \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}(A'-A'_t)\partial_x\left[\ln\left(\frac{\psi}{\psi_\infty}\right)\right]\psi. \quad (6.28)$$

Proof : First, by definition of the total entropy, we have

$$\begin{aligned}\frac{dE}{dt} &= \frac{d}{dt}\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\ln\left(\frac{\psi}{\psi_\infty}\right)\psi \\ &= \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\ln\left(\frac{\psi}{\psi_\infty}\right)\partial_t\psi, \end{aligned} \quad (6.29)$$

using the fact that ψ is a probability density. Next, we use Lemma 4 to obtain

$$\begin{aligned}\frac{dE}{dt} &= \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\mathrm{div}\left(\psi_\infty\nabla\left(\frac{\psi}{\psi_\infty}\right)\right)\ln\left(\frac{\psi}{\psi_\infty}\right) + \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\partial_x((A'-A'_t)\psi)\ln\left(\frac{\psi}{\psi_\infty}\right) \\ &\quad -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\lambda(x)\,\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i}) \quad (6.30) \\ &\leq -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\nabla\ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}(A'-A'_t)\partial_x\left[\ln\left(\frac{\psi}{\psi_\infty}\right)\right]\psi, \end{aligned}$$

where the last line is a result of integration by parts and the fact that (6.30) is non-positive, which is proved below in Lemma 6. ♢

Lemma 6 *For ψ satisfying (6.12) and ψ_∞ its long-time limit, the following holds*

$$-\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\lambda(x)\,\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i}) \leq 0.$$

Proof : First, recall that $\lambda(x) = 0, \forall x \in \mathcal{E}$. We consider therefore the left hand side of the inequality at fixed $(x, y) \in (\mathbb{T}\backslash\mathcal{E})\times\mathbb{R}$. At fixed (x, y), we consider renormalized

(Bernoulli) probabilities denoted by $\tilde{\psi}_i = \psi_i/(\psi_0+\psi_1)$, so that $\tilde{\psi}_0+\tilde{\psi}_1=1$.

$$
\begin{aligned}
&-\sum_{i=0}^{1}\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i}) \\
&\quad=(-\psi_0+\psi_1)\left[\ln\left(\frac{\psi_0}{\psi_{\infty,0}}\right)-\ln\left(\frac{\psi_1}{\psi_{\infty,1}}\right)\right] \\
&\quad=(\psi_0+\psi_1)(-\tilde{\psi}_0+\tilde{\psi}_1)\left[\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)-\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\right] \\
&\quad=(\psi_0+\psi_1)\left[-\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)\tilde{\psi}_0-\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\tilde{\psi}_1\right. \\
&\qquad\left.+\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)\tilde{\psi}_1+\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\tilde{\psi}_0\right] \\
&\quad=(\psi_0+\psi_1)\left[-2\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)\tilde{\psi}_0-2\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\tilde{\psi}_1\right. \\
&\qquad\left.+(\tilde{\psi}_0+\tilde{\psi}_1)\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)+(\tilde{\psi}_0+\tilde{\psi}_1)\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\right] \\
&\quad=(\psi_0+\psi_1)\left[-2\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_i}{\tilde{\psi}_\infty}\right)\tilde{\psi}_i+\ln\left(\frac{\tilde{\psi}_0}{\tilde{\psi}_{\infty,0}}\right)+\ln\left(\frac{\tilde{\psi}_1}{\tilde{\psi}_{\infty,1}}\right)\right].
\end{aligned}
$$

From hypothesis [H1], $\forall(x,y)\in(\mathbb{T}\backslash\mathcal{E})\times\mathbb{R}$, $\tilde{\psi}_{\infty,1}=\tilde{\psi}_{\infty,0}=\frac{1}{2}$, which allows the above to be written as

$$
\begin{aligned}
&-\sum_{i=0}^{1}\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i}) \\
&\quad=(\psi_0+\psi_1)\left[-2\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_i}{\tilde{\psi}_\infty}\right)\tilde{\psi}_i+2\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_i}{\tilde{\psi}_\infty}\right)\tilde{\psi}_\infty\right] \\
&\quad=-2(\psi_0+\psi_1)\left[\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_i}{\tilde{\psi}_\infty}\right)\tilde{\psi}_i+\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_\infty}{\tilde{\psi}_i}\right)\tilde{\psi}_\infty\right] \qquad (6.31) \\
&\quad\leq 0.
\end{aligned}
$$

The last line is due to the fact that the two terms between brackets are non-negative, since they are relative entropies. ♢

Remark 9 *Let us consider the case where $\mathcal{E}=\emptyset$ and thus $\forall x\in\mathbb{T}$, $\lambda(x)=\lambda>0$, which implies (see [H1]) $V_0=V_1$ and therefore $\psi_{\infty,0}=\psi_{\infty,1}$ everywhere on $\mathbb{T}\times\mathbb{R}$.*

In this case, it follows from (6.31) that

$$
\begin{aligned}
-\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\lambda(x)\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i}) & \\
&\leq -2\lambda\int_{\mathbb{T}\times\mathbb{R}}(\psi_0+\psi_1)\sum_{i=0}^{1}\ln\left(\frac{\tilde{\psi}_i}{\tilde{\psi}_\infty}\right)\tilde{\psi}_i \\
&\leq -2\lambda\int_{\mathbb{T}\times\mathbb{R}}\sum_{i=0}^{1}\ln\left(\frac{\psi}{\psi_\infty}\right)\psi+2\lambda\int_{\mathbb{T}\times\mathbb{R}}\ln\left(\frac{\psi_0+\psi_1}{\psi_{\infty,0}+\psi_{\infty,1}}\right)(\psi_0+\psi_1).
\end{aligned}
$$

Furthermore, since the marginal $\psi^{x,y}:=\psi_0+\psi_1$ satisfies in this specific case

$$
\partial_t\psi^{x,y}=\operatorname{div}(\psi_\infty^{x,y}\nabla(\psi^{x,y}/\psi_\infty^{x,y}))+\partial_x((A'-A'_t)\psi^{x,y}), \tag{6.32}
$$

one can show, using the results of [Lelièvre 2008], that $\exists C>0$, $\forall t\geq 0$,

$$
-\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\lambda(x)\ln\left(\frac{\psi}{\psi_\infty}\right)(\psi-\psi_{1-i})\leq -2\lambda E+2\lambda C\mathrm{e}^{-2\min\{\rho,4\pi^2\}t}.
$$

Using this, and the fact that $-E\leq -E_m$, one can show that E_m converges to zero exponentially fast with rate

$$
2\min\{\rho,4\pi^2,\lambda\}.
$$

The convergence rate thus depends on $\lambda>0$, the rate at which switching occurs between the two channels. This is comparable to the original result obtained for the ABF algorithm, see [Lelièvre 2008]. The aim of what follows is to obtain a result in the case where $\mathcal{E}\neq\emptyset$, namely when $\lambda=0$ in some region.

In order to estimate the last term on the right hand side of (6.28), it will be helpful to express the difference of the biasing force and the mean force in terms of densities

Lemma 7 *The difference between the biasing force A'_t and the mean force A' can be expressed in the following way*

$$
A'_t-A'=\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x\ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,dy-\partial_x\ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right).
$$

Proof : We develop the expression on the right hand side and use the fact that $\psi_\infty^\xi\equiv 1$

$$\begin{aligned}
\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x\ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,dy - \partial_x\ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)& \\
&= \sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x\ln(\psi)\frac{\psi}{\psi^\xi}\,dy - \sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x\ln(\psi_\infty)\frac{\psi}{\psi^\xi}\,dy - \partial_x\ln\left(\psi^\xi\right)\\
&= \sum_{i=0}^{1}\int_{\mathbb{R}}\frac{\partial_x\psi}{\psi^\xi}\,dy + \sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x(V_i - A\circ\xi)\frac{\psi}{\psi^\xi}\,dy - \frac{\partial_x\psi^\xi}{\psi^\xi}\\
&= \sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x V_i\frac{\psi}{\psi^\xi}\,dy - \sum_{i=0}^{1}\int_{\mathbb{R}}A'\frac{\psi}{\psi^\xi}\,dy\\
&= A'_t - A',
\end{aligned}$$

The last line is a result of the definition of A'_t in (6.13) and the fact that A is a function of x only. ♢

Another useful estimate for the difference between A'_t and A' is given in the following lemma.

Lemma 8 *The difference of the biasing force and the mean force can be bounded by the microscopic entropy as:*

$$\int_{\mathbb{T}}|A'_t - A'|^2\psi^\xi\,dx \le 2R^2E_m(t),$$

where

$$R = \left(C + M\rho^{-1/2}\right).$$

Proof : We begin by showing that $|A'_t(x) - A'(x)| \le M\sqrt{2e_m(t,x)/\rho}$. By definition, we have

$$\begin{aligned}
A'_t(x) - A'(x) &= \frac{\displaystyle\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x V_i\psi\,dy}{\displaystyle\sum_{i=0}^{1}\int_{\mathbb{R}}\psi\,dy} - \frac{\displaystyle\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x V_i\psi_\infty\,dy}{\displaystyle\sum_{i=0}^{1}\int_{\mathbb{R}}\psi_\infty\,dy}\\
&= \sum_{i=0}^{1}\int_{\mathbb{R}}\left(\left(\partial_x V_i\frac{\psi^{\xi,I}}{\psi^\xi}\right)\frac{\psi}{\psi^{\xi,I}} - \left(\partial_x V_i\frac{\psi^{\xi,I}_\infty}{\psi^\xi_\infty}\right)\frac{\psi_\infty}{\psi^{\xi,I}_\infty}\right)\,dy\\
&= \sum_{i=0}^{1}\int_{\mathbb{R}\times\mathbb{R}}\left(\partial_x V_i(x,y)\frac{\psi^{\xi,I}}{\psi^\xi} - \partial_x V_i(x,y')\frac{\psi^{\xi,I}_\infty}{\psi^\xi_\infty}\right)\pi(dy,dy')
\end{aligned}$$

where $\pi(dy, dy')$ is any coupling measure on $\mathbb{R} \times \mathbb{R}$ with marginals $\mu_{t|x,i}$ and $\mu_{\infty|x,i}$. Next, using Taylor's expansion on $\partial_x V_i(x, y)$, we have

$$\begin{aligned}
&A_t'(x) - A'(x) \\
&\quad = \sum_{i=0}^{1} \int_{\mathbb{R}\times\mathbb{R}} \left(\left(\partial_x V_i(x, y') + \partial_{x,y} V_i(x, \eta(y, y'))(y - y')\right) \frac{\psi^{\xi,I}}{\psi^{\xi}} - \partial_x V_i(x, y') \frac{\psi_{\infty}^{\xi,I}}{\psi_{\infty}^{\xi}} \right) \pi(dy, dy') \\
&\quad = \sum_{i=0}^{1} \int_{\mathbb{R}\times\mathbb{R}} \left(\partial_x V_i(x, y') \left(\frac{\psi^{\xi,I}}{\psi^{\xi}} - \frac{\psi_{\infty}^{\xi,I}}{\psi_{\infty}^{\xi}} \right) + \partial_{x,y} V_i(x, \eta(y, y'))(y - y') \frac{\psi^{\xi,I}}{\psi^{\xi}} \right) \pi(dy, dy') \\
&\quad = \sum_{i=0}^{1} \left(\frac{\psi^{\xi,I}}{\psi^{\xi}} - \frac{\psi_{\infty}^{\xi,I}}{\psi_{\infty}^{\xi}} \right) \int_{\mathbb{R}} \partial_x V_i(x, y') \frac{\psi_{\infty}}{\psi_{\infty}^{\xi,I}} \, dy' \\
&\qquad + \sum_{i=0}^{1} \frac{\psi^{\xi,I}}{\psi^{\xi}} \int_{\mathbb{R}\times\mathbb{R}} \partial_{x,y} V_i(x, \eta(y, y'))(y - y') \pi(dy, dy')
\end{aligned}$$

where $\eta(y, y') \in [y, y']$. Recall from [H2] that $\exists C, M > 0$ such that

$$\left\| \int_{\mathbb{R}} \partial_x V_i \, d\mu_{\infty|x,i} \right\|_{L^\infty(\mathbb{T})} \leq C,$$

and $\|\partial_{x,y} V_i\|_{L^\infty} \leq M$. Furthermore, with the use of the Csiszar-Kullback inequality, we have

$$\begin{aligned}
|A_t'(x) - A'(x)| &\leq C \sum_{i=0}^{1} \left| \frac{\psi^{\xi,I}}{\psi^{\xi}} - \frac{\psi_{\infty}^{\xi,I}}{\psi_{\infty}^{\xi}} \right| + M \sum_{i=0}^{1} \frac{\psi^{\xi,I}}{\psi^{\xi}} \int_{\mathbb{R}\times\mathbb{R}} |y - y'| \pi(dy, dy') \\
&\leq C \sqrt{2H(\mu_{t|x}^I, \mu_{\infty|x}^I)} + M \sum_{i=0}^{1} \frac{\psi^{\xi,I}}{\psi^{\xi}} \int_{\mathbb{R}\times\mathbb{R}} |y - y'| \pi(dy, dy').
\end{aligned}$$

Next, by the Talagrand inequality (6.21) and the concavity of the function $x \mapsto \sqrt{x}$,

$$\begin{aligned}
|A_t'(x) - A'(x)| &\leq C \sqrt{2H(\mu_{t|x}^I, \mu_{\infty|x}^I)} + M \sum_{i=0}^{1} \frac{\psi^{\xi,I}}{\psi^{\xi}} \sqrt{\frac{2}{\rho} H(\mu_{t|x,i}, \mu_{\infty|x,i})} \\
&\leq C \sqrt{2H(\mu_{t|x}, \mu_{\infty|x})} + M \sqrt{\frac{2}{\rho} \sum_{i=0}^{1} H(\mu_{t|x,i}, \mu_{\infty|x,i}) \frac{\psi^{\xi,I}}{\psi^{\xi}}} \\
&\leq C \sqrt{2H(\mu_{t|x}, \mu_{\infty|x})} + M \sqrt{\frac{2}{\rho} H(\mu_{t|x}, \mu_{\infty|x})}
\end{aligned}$$

Therefore

$$|A_t'(x) - A'(x)|^2 \leq 2 \left(C + M \rho^{-1/2} \right)^2 e_m(t, x).$$

The result follows immediately, since $E_m(t) = \int_{\mathbb{T}} e_m(t, x) \, \psi^{\xi} dx$. ♢

We are now equipped with the right tools to control dE/dt, and are left to handle the evolution of the macroscopic entropy.

6.3.2 Controlling E_M

Due to the free diffusion equation satisfied by ψ^ξ, the macroscopic entropy E_M is easily controlled.

Lemma 9 *The macroscopic entropy satisfies*

$$\frac{dE_M}{dt} = -\int_{\mathbb{T}} \left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi = -F(\psi^\xi|\psi^\xi_\infty). \tag{6.33}$$

Proof : Using (6.16) and integration by parts

$$\begin{aligned}
\frac{dE_M}{dt} &= \frac{d}{dt}\int_{\mathbb{T}} \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\psi^\xi \\
&= \int_{\mathbb{T}} \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\partial_{xx}\psi^\xi \\
&= -\int_{\mathbb{T}} \left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2 \psi^\xi.
\end{aligned}$$

Lemma 10 *The Fisher information for the marginal density ψ^ξ decays exponentially fast with rate $r = 8\pi^2$*

$$F(\psi^\xi(t,\cdot)|\psi^\xi_\infty) \leq F_0 \, \exp(-8\pi^2 t),$$

where $F_0 = F(\psi^\xi(0,\cdot)|\psi^\xi_\infty)$.

Proof : See Lemma 12 of reference [Lelièvre 2008].

♢

In light of the lemmas presented above, hypotheses [H1]-[H4] may now be used to control the evolution of the microscopic entropy.

6.3.3 Controlling E_m

The aim of this section is to obtain an estimation on the evolution of th microscopic entropy E_m, see (6.36) below. We first begin by using the fact that the channel-local conditional measures $\mu_{\infty|x,i}$ satisfy LSI(ρ) to bound the microscopic entropy.

Lemma 11 *Under the hypothesis [H3], the microscopic entropy E_m satisfies*

$$E_m \leq \frac{1}{2\rho}\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}} \left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2 \psi \; dx \; dy + E_c,$$

where E_c is defined in (6.24).
Proof : We fist consider the local entropy $e_m(t,x)$, which can be decomposed into

the sum of the channel-local entropy and the entropy of the Bernoulli measures.

$$\begin{aligned}
e_m(t,x) &= H(\mu_{t|x}|\mu_{\infty|x}) \\
&= \sum_{i=0}^{1}\int_{\mathbb{R}} \ln\left(\frac{\psi}{\psi^\xi}\Big/\frac{\psi_\infty}{\psi_\infty^\xi}\right)\frac{\psi}{\psi^\xi}\,dy \\
&= \sum_{i=0}^{1}\left[\int_{\mathbb{R}} \ln\left(\frac{\psi}{\psi^{\xi,I}}\Big/\frac{\psi_\infty}{\psi_\infty^{\xi,I}}\right)\frac{\psi}{\psi^\xi}\,dy + \int_{\mathbb{R}} \ln\left(\frac{\psi^{\xi,I}}{\psi^\xi}\Big/\frac{\psi_\infty^{\xi,I}}{\psi_\infty^\xi}\right)\frac{\psi}{\psi^\xi}\,dy\right] \\
&\le \sum_{i=0}^{1}\left[\frac{1}{2\rho}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\frac{\psi}{\psi^\xi}\,dy\right] + \sum_{i=0}^{1}\left[\ln\left(\frac{\psi^{\xi,I}}{\psi^\xi}\Big/\frac{\psi_\infty^{\xi,I}}{\psi_\infty^\xi}\right)\frac{\psi^{\xi,I}}{\psi^\xi}\right] \\
&= \frac{1}{2\rho}\sum_{i=0}^{1}\int_{\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\frac{\psi}{\psi^\xi}\,dy + e_c(t,x),
\end{aligned}$$

where the inequality is a direct result of [H3]. The microscopic entropy is then

$$\begin{aligned}
E_m(t) &= \int_{\mathbb{T}} e_m(t,x)\psi^\xi\,dx \\
&\le \frac{1}{2\rho}\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi\,dx\,dy + E_c,
\end{aligned}$$

as required. ♢

The time evolution of E_m may now be expressed using results of Lemmas 5 and 9.

$$\begin{aligned}
\frac{dE_m}{dt} &= \frac{dE}{dt} - \frac{dE_M}{dt} \\
&\le -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\nabla \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}(A' - A'_t)\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi \\
&\quad + \int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi_\infty^\xi}\right)\right|^2\psi^\xi.
\end{aligned}$$

We may now apply Lemma 7 and use integration by parts

$$\begin{aligned}
\frac{dE_m}{dt} &= -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi \\
&\quad + \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left(\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\frac{\psi}{\psi^\xi}\,dy - \partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right)\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi \\
&\quad + \int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2\psi^\xi \\
&= -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi \\
&\quad - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi + \int_{\mathbb{T}}\left(\sum_{i=0}^{1}\int_{\mathbb{R}}\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi\,dy\right)^2\frac{1}{\psi^\xi} \qquad (6.34) \\
&\quad - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi + \int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2\psi^\xi.
\end{aligned}$$

Notice that (6.34) is non-positive by the Cauchy-Schwartz inequality. We therefore have

$$\begin{aligned}
\frac{dE_m}{dt} &\leq -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi \\
&\quad - \sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\partial_x \ln\left(\frac{\psi}{\psi_\infty}\right)\psi + \int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2\psi^\xi \\
&= -\sum_{i=0}^{1}\int_{\mathbb{T}\times\mathbb{R}}\left|\partial_y \ln\left(\frac{\psi}{\psi_\infty}\right)\right|^2\psi - \int_{\mathbb{T}}\left[\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\psi^\xi\right](A'_t - A') \qquad (6.35) \\
&\leq -2\rho E_m + 2\rho E_c + \sqrt{\int_{\mathbb{T}}|A'_t - A'|^2\,\psi^\xi}\sqrt{\int_{\mathbb{T}}\left|\partial_x \ln\left(\frac{\psi^\xi}{\psi^\xi_\infty}\right)\right|^2\psi^\xi}.
\end{aligned}$$

Line (6.35) is a result of Lemma 7 and the last inequality is due to Lemma 11 and a further application of the Cauchy-Schwartz inequality. Now, using Lemmas 8 and 10, we obtain

$$\frac{dE_m}{dt} \leq -2\rho E_m + 2\rho E_c + R\sqrt{2E_m}\sqrt{F_0 \mathrm{e}^{-8\pi^2 t}},.$$

where we recall $F_0 = F(\psi^\xi(0,\cdot)|\psi^\xi_\infty)$. Finally, using Young's inequality: $\forall \varepsilon > 0$, $\forall a, b \in \mathbb{R}$, $ab < \varepsilon a^2 + \frac{1}{4\varepsilon}b^2$, we obtain

$$\frac{dE_m}{dt} \leq -2\left(\rho - R^2\varepsilon\right)E_m + 2\rho E_c + \frac{1}{4\varepsilon}F_0 \mathrm{e}^{-8\pi^2 t}, \qquad (6.36)$$

where $\varepsilon > 0$ will be chosen optimally later in the proof. We are left to control the channel entropy term E_c in order to conclude.

6.3.4 Controlling E_c

The aim of this section is to obtain a control on the evolution of E_c, the relative entropy of $\mu^I_{t|x}$ and $\mu^I_{\infty|x}$, and more precisely of a quantity P which is the weighted χ^2-distance between two measures, see (6.38) and (6.49) below. Recall that $E_c = \int_{\mathbb{T}} H(\mu^I_{t|x}|\mu^I_{\infty|x})\psi^\xi \, dx$, where the integrand is the relative entropy of a Bernoulli measure. Poincaré and logarithmic Sobolev inequalities have been studied for Bernoulli measures [Ané 2000]. In order to obtain an exponentially decaying relative entropy, however, a suitable semi-group and its infinitesimal generator is needed for the measure. In the case of the bi-channel model (in particular [H1]), we face problems due to the region $\mathcal{E} \subset \mathbb{T}$, where no exchange is permitted between the two channels: in this region, the speed at which the measure $\mu^I_{t|x}$ reaches equilibrium cannot directly be controlled. To circumvent this issue, we consider the spectral gap of an adequate operator and resort to the Poincaré inequality.
By the definition of E_c and using the inequality $\forall x > 0$, $x\ln(x) \leq x(x-1)$ and the fact that $\sum_{i=0}^1 \int_{\mathbb{T}} \psi^{\xi,I}(t,x,i)\,dx = 1$, we obtain

$$\begin{aligned} E_c &= \sum_{i=0}^{1} \int_{\mathbb{T}} \ln\left(\frac{\psi^{\xi,I}}{\psi^\xi} \Big/ \frac{\psi^{\xi,I}_\infty}{\psi^\xi_\infty} \right) \psi^{\xi,I} \\ &= \sum_{i=0}^{1} \int_{\mathbb{T}} \ln\left(\psi^{\xi,I}/\psi^{\xi,I}_\infty \right) \psi^{\xi,I} - E_M \\ &\leq \sum_{i=0}^{1} \int_{\mathbb{T}} \left(\frac{\psi^{\xi,I}}{\psi^{\xi,I}_\infty} - 1 \right)^2 \psi^{\xi,I}_\infty \, dx - E_M \end{aligned} \tag{6.37}$$

We therefore have

$$E_c \leq P, \tag{6.38}$$

where

$$P = \sum_{i=0}^{1} \int_{\mathbb{T}} \left(\frac{\psi^{\xi,I}}{\psi^{\xi,I}_\infty} - 1 \right)^2 \psi^{\xi,I}_\infty \, dx. \tag{6.39}$$

In order to proceed and consider the time derivative of P, we will need some further results to express the evolution of the marginal density $\psi^{\xi,I}$. The idea is to compare the evolution of $\psi^{\xi,I}$ with the dynamics of this density if A'_t and $\int_{\mathbb{T}} \partial_x V d\mu_{t|x,i}$ were already at equilibrium (see Section 6.2.4).

Lemma 12 *The Fokker-Planck equation (6.15) for $\psi^{\xi,I}$ can be rewritten as*

$$\begin{aligned} \partial_t \psi^{\xi,I} &= \partial_x \left(\psi^{\xi,I}_\infty \partial_x \left(\frac{\psi^{\xi,I}}{\psi^{\xi,I}_\infty} \right) \right) + \partial_x \left(\left(\frac{\int_{\mathbb{R}} \partial_x V_i \psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}} \partial_x V_i \psi_\infty}{\psi^{\xi,I}_\infty} \right) \psi^{\xi,I} \right) \\ &\quad + \partial_x((A' - A'_t)\psi^{\xi,I}) - \lambda(x)(\psi^{\xi,I} - \psi^{\xi,I}_{1-i}) \end{aligned} \tag{6.40}$$

Proof : First we will show that

$$\partial_t \psi^{\xi,I} = \partial_x \left(\int_{\mathbb{R}} \psi_\infty \partial_x \left(\frac{\psi}{\psi_\infty} \right) \, dy \right) + \partial_x((A' - A'_t)\psi^{\xi,I}) - \lambda(x)(\psi^{\xi,I} - \psi^{\xi,I}_{1-i}). \tag{6.41}$$

By developing the right hand side, we have

$$\begin{aligned}\partial_t\psi^{\xi,I} &= \partial_x\left(\int_{\mathbb{R}}\partial_x\psi - \frac{\psi}{\psi_\infty}\partial_x\psi_\infty\right) + \partial_x\int_{\mathbb{R}}(A'-A_t')\psi - \lambda(x)(\psi^{\xi,I}-\psi_{1-i}^{\xi,I}) \\ &= \partial_x\left(\int_{\mathbb{R}}\partial_x(V_i - A\circ\xi)\psi\right) + \partial_{xx}\psi^{\xi,I} + \partial_x\int_{\mathbb{R}}(A'-A_t')\psi - \lambda(x)(\psi^{\xi,I}-\psi_{1-i}^{\xi,I}) \\ &= \int_{\mathbb{R}}\partial_x\left(\partial_x(V_i - A_t\circ\xi)\psi\right) + \partial_{xx}\psi^{\xi,I} - \lambda(x)(\psi^{\xi,I}-\psi_{1-i}^{\xi,I})\end{aligned}$$

which is indeed the Fokker-Planck equation (6.15) associated to $\psi^{\xi,I}$.

Next, we show that $\forall x\in\mathbb{T}$, $\forall i\in\{0,1\}$

$$\int_{\mathbb{R}}\psi_\infty\partial_x\left(\frac{\psi}{\psi_\infty}\right)\,dy = \psi_\infty^{\xi,I}\partial_x\left(\frac{\psi^{\xi,I}}{\psi_\infty^{\xi,I}}\right) + \left(\frac{\int_{\mathbb{R}}\partial_x V_i\psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}}\partial_x V_i\psi_\infty}{\psi_\infty^{\xi,I}}\right)\psi^{\xi,I}. \quad (6.42)$$

To prove the above, notice that

$$\begin{aligned}\int_{\mathbb{R}}\psi_\infty\partial_x\left(\frac{\psi}{\psi_\infty}\right)\,dy &- \psi_\infty^{\xi,I}\partial_x\left(\frac{\psi^{\xi,I}}{\psi_\infty^{\xi,I}}\right) \\ &= \int_{\mathbb{R}}\partial_x\psi\;dy - \int_{\mathbb{R}}\frac{\psi}{\psi_\infty}\partial_x\psi_\infty - \partial_x\psi^{\xi,I} + \frac{\psi^{\xi,I}}{\psi_\infty^{\xi,I}}\partial_x\psi_\infty^{\xi,I} \\ &= \int_{\mathbb{R}}(\partial_x(V_i - A\circ\xi))\psi\;dy - \frac{\int_{\mathbb{R}}\partial_x(V_i - A\circ\xi)e^{-(V_i-A\circ\xi)}dy}{\int_{\mathbb{R}}e^{-(V_i-A\circ\xi)}\,dy}\psi^{\xi,I} \\ &= \int_{\mathbb{R}}(\partial_x V_i)\psi\;dy - A'(x)\psi^{\xi,I} - \frac{\int_{\mathbb{R}}\partial_x V_i e^{-(V_i-A\circ\xi)}\,dy}{\int_{\mathbb{R}}e^{-(V_i-A\circ\xi)}\,dy}\psi^{\xi,I} + A'(x)\psi^{\xi,I}\end{aligned}$$

Finally, by using the fact that the free energy A is independent of y, we obtain

$$\int_{\mathbb{R}}\psi_\infty\partial_x(\psi/\psi_\infty)\,dy - \psi_\infty^{\xi,I}\partial_x(\psi^{\xi,I}/\psi_\infty^{\xi,I}) = \left(\frac{\int_{\mathbb{R}}\partial_x V_i\psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}}\partial_x V_i\psi_\infty}{\psi_\infty^{\xi,I}}\right)\psi^{\xi,I}, \quad (6.43)$$

as required. The final result (6.40) is obtained by substituting (6.42) into (6.41). ♢

Notice that (6.40) is comparable to (6.25), only with additional terms due to the fact that A_t and $\int_{\mathbb{T}}\partial_x V d\mu_{t|x,i}$ have not yet converged. The difference of the biasing force and mean force, $A_t' - A'$, was already estimated in Lemmas 7 and 8. We are therefore left to control the remaining term.

Lemma 13 $\forall x\in\mathbb{T}$, $\forall i\in\{0,1\}$,

$$\left|\frac{\int_{\mathbb{R}}\partial_x V_i\psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}}\partial_x V_i\psi_\infty}{\psi_\infty^{\xi,I}}\right| \le M\sqrt{\frac{2}{\rho}H\left(\mu_{t|x,i}\,|\mu_{\infty|x,i}\right)}.$$

As a consequence,

$$\sum_{i=0}^{1} \int_{\mathbb{T}} \left| \frac{\int_{\mathbb{R}} \partial_x V_i \psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}} \partial_x V_i \psi_\infty}{\psi_\infty^{\xi,I}} \right|^2 \psi^{\xi,I}\, dx \leq \frac{2M^2}{\rho} E_m. \tag{6.44}$$

Proof : Let $\Pi(\mu_{t|x,i}, \mu_{\infty|x,i})$ be the set of coupling measures on $\mathbb{R}\times\mathbb{R}$ with marginals $\mu_{t|x,i}$ and $\mu_{\infty|x,i}$ respectively and let $\pi \in \Pi$. Then

$$\begin{aligned}
\left| \int_{\mathbb{R}} \left(\partial_x V_i \frac{\psi}{\psi^{\xi,I}} - \partial_x V_i \frac{\psi_\infty}{\psi_\infty^{\xi,I}} \right) \right| &= \left| \int_{\mathbb{R}\times\mathbb{R}} \left(\partial_x V_i(x,y) - \partial_x V_i(x,y') \right)\, \pi(dy, dy') \right| \\
&\leq \|\partial_{x,y} V_i\|_{L^\infty} \int_{\mathbb{R}\times\mathbb{R}} |y - y'|\, \pi(dy, dy') \\
&\leq M \sqrt{\frac{2}{\rho} H(\mu_{t|x,i} | \mu_{\infty|x,i})}
\end{aligned}$$

where we have used Lemma 3 since $\mu_{\infty|x,i}$ satisfies LSI(ρ). Equation (6.44) follows immediately from the fact that $\sum_{i=0}^{1} \int_{\mathbb{T}} H(\mu_{t|x,i}|\mu_{\infty|x,i}) \psi^{\xi,I}\, dx \leq E_m$. ♢

One result that is now needed to derive estimates on the evolution of P is the existence of a spectral gap of the operator describing the dynamics once A'_t and $\int_{\mathbb{R}} \partial_x V_i d\mu_{t|x,i}$ have converged (see (6.25)). We now justify the existence of such a spectral gap.

In order to do so, let us define the vector spaces

$$\mathcal{V}_l = \left\{ v : \mathbb{T} \times \{0,1\} \to \mathbb{R} \;\middle|\; \forall i \in \{0,1\},\ \frac{v_i}{\psi_\infty^{\xi,I}(x,i)} \in L^2\left(\mathbb{T}, \psi_\infty^{\xi,I}(x,i)\, dx\right),\ \sum_{i=0}^{1} \int_{\mathbb{T}} v_i(x)\, dx = l \right\}$$

and

$$\mathcal{W}_l = \left\{ w \in \mathcal{V}_l \;\middle|\; \forall i \in \{0,1\},\ \frac{w_i}{\psi_\infty^{\xi,I}(x,i)} \in H^1\left(\mathbb{T}, \psi_\infty^{\xi,I}(x,i)\, dx\right),\ \sum_{i=0}^{1} \int_{\mathbb{T}} w_i(x)\, dx = l \right\}.$$

A function ϕ in $\mathcal{V}_l$ (or in $\mathcal{W}_l$) will also be considered as a vector valued function as $\phi : \begin{cases} \mathbb{T} \to \mathbb{R}^2 \\ x \mapsto (\phi_0(x), \phi_1(x)) \end{cases}$. Notice that $\phi \in \mathcal{W}_1$ if and only if $f := \phi - \psi_\infty^{\xi,I} \in \mathcal{W}_0$.

Lemma 14 *Recall the operator $\mathcal{L} = (\mathcal{L}_0, \mathcal{L}_1)$, with $\mathcal{L}_i$ defined as in (6.25),*

$$\mathcal{L}_i \phi = -\left[\partial_x \left(\psi_{\infty,i}^{\xi,I} \partial_x \left(\phi_i / \psi_{\infty,i}^{\xi,I} \right) \right) - \lambda(x)(\phi_i - \phi_{1-i}) \right].$$

Then

i) The operator $\mathcal{L}$ is symmetric and positive definite with respect to the inner product

$$\langle f, g \rangle = \sum_{i=0}^{1} \int_{\mathbb{T}} f_i(x) g_i(x)\, \frac{1}{\psi_\infty^{\xi,I}(x,i)}\, dx.$$

ii) $\mathcal{L}$ has a spectral gap $\theta > 0$ in the sense that

$$\inf_{f\in\mathcal{W}_0, f\neq 0} \frac{\langle f, \mathcal{L}f\rangle}{\langle f, f\rangle} = \theta > 0. \tag{6.45}$$

Proof : i) To show symmetry of the operator $\mathcal{L}$, consider functions $\varphi, \phi \in \mathcal{W}_0$. Now, using the fact that $\forall x \in \mathbb{T}\setminus\mathcal{E}$, $\psi^{\xi,I}_{\infty,0}(x) = \psi^{\xi,I}_{\infty,1}(x)$,

$$\begin{aligned}
&\sum_{i=0}^{1}\int_{\mathbb{T}} \varphi_i \mathcal{L}_i\phi \frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx \\
&\quad= -\sum_{i=0}^{1}\int_{\mathbb{T}} \varphi_i \partial_x\left(\psi^{\xi,I}_{\infty,i}\partial_x\left(\frac{\phi_i}{\psi^{\xi,I}_{\infty,i}}\right)\right)\frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx + \sum_{i=0}^{1}\int_{\mathbb{T}}\lambda(x)\varphi_i(\phi_i-\phi_{1-i})\frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx \\
&\quad= \sum_{i=0}^{1}\int_{\mathbb{T}} \partial_x\left(\frac{\varphi_i}{\psi^{\xi,I}_{\infty,i}}\right)\partial_x\left(\frac{\phi_i}{\psi^{\xi,I}_{\infty,i}}\right)\psi^{\xi,I}_{\infty,i}\,dx + \int_{\mathbb{T}}\lambda(x)(\varphi_0-\varphi_1)(\phi_0-\phi_1)\frac{1}{\psi^{\xi,I}_{\infty,0}}\,dx \\
&\quad= -\sum_{i=0}^{1}\int_{\mathbb{T}} \phi_i \partial_x\left(\psi^{\xi,I}_{\infty,i}\partial_x\left(\frac{\varphi_i}{\phi_{\infty,i}}\right)\right)\frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx + \int_{\mathbb{T}}\lambda(x)(\phi_0-\phi_1)(\varphi_0-\varphi_1)\frac{1}{\psi^{\xi,I}_{\infty,0}}\,dx \\
&\quad= \sum_{i=0}^{1}\int_{\mathbb{T}} \phi_i \mathcal{L}_i\varphi \frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx.
\end{aligned} \tag{6.46}$$

From (6.46), we conclude positive definiteness of $\mathcal{L}$

$$\sum_{i=0}^{1}\int_{\mathbb{T}} \phi_i \mathcal{L}_i\phi \frac{1}{\psi^{\xi,I}_{\infty,i}}\,dx = \sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\phi_i}{\psi^{\xi,I}_{\infty,i}}\right)\right|^2\psi^{\xi,I}_{\infty,i}\,dx + \int_{\mathbb{T}}\lambda(x)(\phi_0-\phi_1)^2\frac{1}{\psi^{\xi,I}_{\infty,0}}\,dx > 0. \tag{6.47}$$

Notice that the above is strictly positive for any $\phi\in\mathcal{W}_0$, since $\langle\phi,\mathcal{L}\phi\rangle = 0$ if and only if $\phi_0 = \phi_1 = 0$.

ii) In fact, one can check that $\exists\kappa > 0$, such that $\forall\phi\in\mathcal{W}_0$, $\phi\neq 0$,

$$\sum_{i=0}^{1}\int_{\mathbb{T}} \phi_i \mathcal{L}_i\phi \frac{1}{\psi^{\xi,I}_{\infty}}\,dx \geq \kappa\sum_{i=0}^{1}\int_{\mathbb{T}}\left(|\phi_i|^2 + |\nabla\phi_i|^2\right)\frac{1}{\psi^{\xi,I}_{\infty}}\,dx.$$

Therefore, by the Lax-Milgram theorem, $\mathcal{L}^{-1}$ is well defined from $\mathcal{V}_0$ to $\mathcal{W}_0$ and thus compact from $\mathcal{V}_0$ to $\mathcal{V}_0$. From the symmetry and positive definiteness of $\mathcal{L}$, and the fact that its inverse is a compact operator from $\mathcal{V}_0$ to $\mathcal{V}_0$, it has a strictly positive and discrete spectrum. There exists a set of eigenvectors $(v_n)_{n\geq 1}$, orthonormal with respect to the inner product $\langle\cdot,\cdot\rangle$, forming a basis of $\mathcal{V}_0$ and $\mathcal{W}_0$, and associated to an increasing sequence of eigenvalues $(\sigma_n)_{n\geq 1}$, such that $\lim_{n\to\infty}\sigma_n = \infty$. In particular, there exists a spectral gap: $\theta = \sigma_1 > 0$. ♢

Remark 10 *In the case where a function $\phi\in\mathcal{W}_1$ satisfies $\partial_t\phi = \mathcal{L}\phi$, it holds that*

$$\forall t\geq 0,\quad \|\phi - \psi^{\xi,I}_{\infty}\|^2 \leq Ke^{-2\theta t}, \tag{6.48}$$

where $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$ *and* $K = \sum_{n\geq 1}\langle\phi(0,\cdot)-\psi_\infty^{\xi,I}, v_n\rangle^2$. *This is easily obtained by noticing that* $\phi - \psi_\infty^{\xi,I} \in \mathcal{W}_0$ *and therefore can be expressed in terms of the orthonormal eigenvectors* $(v_n)_{n\geq 1}$

$$\phi - \psi_\infty^{\xi,I} = \sum_{n\geq 1}\langle\phi(0,\cdot) - \psi_\infty^{\xi,I}, v_n\rangle v_n e^{-\sigma_n t}.$$

The result (6.48) *follows immediately since* $\forall n \geq 2$, $\sigma_n \geq \sigma_1 = \theta$.

With these tools at hand, let us consider the time evolution of the functional P defined in (6.39)

$$\begin{aligned}
\frac{1}{2}\frac{dP}{dt} &= \sum_{i=0}^{1}\int_{\mathbb{T}}\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}} - 1\right)\partial_t\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\psi_\infty^{\xi,I}\,dx \\
&= \sum_{i=0}^{1}\int_{\mathbb{T}}\psi^{\xi,I}\partial_t\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\,dx - \sum_{i=0}^{1}\int_{\mathbb{T}}\partial_t\psi^{\xi,I}\,dx \\
&= \sum_{i=0}^{1}\int_{\mathbb{T}}\psi^{\xi,I}\partial_t\psi^{\xi,I}\frac{1}{\psi_\infty^{\xi,I}}\,dx.
\end{aligned}$$

Using equation (6.40) of Lemma 12, we get

$$\begin{aligned}
\frac{1}{2}\frac{dP}{dt} &= \sum_{i=0}^{1}\int_{\mathbb{T}}\psi^{\xi,I}\partial_t\psi^{\xi,I}\frac{1}{\psi_\infty^{\xi,I}}\,dx \\
&= -\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx - \sum_{i=0}^{1}\int_{\mathbb{T}}\lambda(x)\psi_i^{\xi,I}(\psi_i^{\xi,I} - \psi_{1-i}^{\xi,I})\frac{1}{\psi_{\infty,i}^{\xi,I}}\,dx \\
&\quad -\sum_{i=0}^{1}\int_{\mathbb{T}}\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)(A' - A'_t)\psi^{\xi,I}\,dx \\
&\quad -\sum_{i=0}^{1}\int_{\mathbb{T}}\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\left(\frac{\int_{\mathbb{R}}\partial_x V\psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}}\partial_x V\psi_\infty}{\psi_\infty^{\xi,I}}\right)\psi^{\xi,I}\,dx.
\end{aligned}$$

Notice that, by developing the sum and using the fact that $\psi_{\infty,0}^{\xi,I} = \psi_{\infty,1}^{\xi,I}$ for $\lambda(x) \neq 0$, the second term may be replaced by $\displaystyle\int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I} - \psi_1^{\xi,I}\right|^2\frac{1}{\psi_{\infty,0}^{\xi,I}}\,dx$. Finally by using Young's inequality on the last two terms, we obtain for $\alpha > 0$

$$\begin{aligned}
\frac{1}{2}\frac{dP}{dt} &= -\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx - \int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I} - \psi_1^{\xi,I}\right|^2\frac{1}{\psi_{\infty,0}^{\xi,I}}\,dx \\
&\quad + \frac{1}{4\alpha}\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi^{\xi,I}\,dx + \alpha\int_{\mathbb{T}}|A' - A'_t|^2\psi^\xi\,dx \\
&\quad + \frac{1}{4\alpha}\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi^{\xi,I}\,dx + \alpha\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\frac{\int_{\mathbb{R}}\partial_x V\psi}{\psi^{\xi,I}} - \frac{\int_{\mathbb{R}}\partial_x V\psi_\infty}{\psi_\infty^{\xi,I}}\right|^2\psi^{\xi,I}\,dx.
\end{aligned}$$

Next, by Lemmas 8 and 13,

$$\begin{aligned}\frac{1}{2}\frac{dP}{dt} &= -\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx-\int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I}-\psi_1^{\xi,I}\right|^2\frac{1}{\psi_\infty^{\xi,I}}\,dx\\ &\quad+\frac{1}{2\alpha}\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi^{\xi,I}\,dx+2\alpha\frac{M^2}{\rho}E_m+2\alpha R^2E_m.\end{aligned}$$

Notice that, in the third term, $\psi^{\xi,I}\le\psi^\xi\le\tilde{M}$ for $\tilde{M}=\left\|\psi^\xi(0,\cdot)\right\|_{L^\infty}$ and $1\le\psi_\infty^{\xi,I}/c$ for $c=\min\limits_{x,i}\psi_\infty^{\xi,I}$. This gives

$$\begin{aligned}\frac{1}{2}\frac{dP}{dt} &\le -\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx-\int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I}-\psi_1^{\xi,I}\right|^2\frac{1}{\psi_\infty^{\xi,I}}\,dx\\ &\quad+\frac{\tilde{M}}{2\alpha c}\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx+4\alpha R^2E_m.\end{aligned}$$

Finally, by grouping terms together and using the fact that α may be chosen such that $\tilde{M}/2\alpha c<1$ (an appropriate choice for α is given later in the proof), we have

$$\begin{aligned}\frac{1}{2}\frac{dP}{dt} &\le -\left(1-\frac{\tilde{M}}{2\alpha c}\right)\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx-\int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I}-\psi_1^{\xi,I}\right|^2\frac{1}{\psi_\infty^{\xi,I}}\,dx+4\alpha R^2E_m\\ &\le -\left(1-\frac{\tilde{M}}{2\alpha c}\right)\left[\sum_{i=0}^{1}\int_{\mathbb{T}}\left|\partial_x\left(\frac{\psi^\xi}{\psi_\infty^{\xi,I}}\right)\right|^2\psi_\infty^{\xi,I}\,dx+\int_{\mathbb{T}}\lambda(x)\left|\psi_0^{\xi,I}-\psi_1^{\xi,I}\right|^2\frac{1}{\psi_\infty^{\xi,I}}\,dx\right]+4\alpha R^2E_m\\ &\le -\left(1-\frac{\tilde{M}}{2\alpha c}\right)\sum_{i=0}^{1}\int_{\mathbb{T}}(\psi^{\xi,I}\mathcal{L}_i\psi^{\xi,I})\frac{1}{\psi_\infty^{\xi,I}}\,dx+4\alpha R^2E_m\\ &\le -\left(1-\frac{\tilde{M}}{2\alpha c}\right)\theta P+4\alpha R^2E_m,\end{aligned}\tag{6.49}$$

where the last line is a result of (6.45), with $f:=\psi^{\xi,I}-\psi_\infty^{\xi,I}$. Notice that $f\in\mathcal{W}_0$ since the normalization for $\psi^{\xi,I}$ is $\sum\limits_{i=0}^{1}\int_{\mathbb{T}}\psi^{\xi,I}dx=1$.

To complete the proof of Theorem 5, we now need to study the system of inequalities (6.36) and (6.49).

6.3.5 Completing the proof

To show that E_m decays exponentially fast, we study the system of two inequalities (6.36) and (6.49). Since, from (6.38), $E_c\le P$, the system to be studied is

$$\begin{cases}\dfrac{dE_m}{dt} &\le -2\left(\rho-R^2\varepsilon\right)E_m+2\rho P+\dfrac{1}{4\varepsilon}F_0\,\mathrm{e}^{-8\pi^2t},\\ \dfrac{dP}{dt} &\le 8\alpha R^2E_m-2\left(1-\dfrac{\tilde{M}}{2\alpha c}\right)\theta P.\end{cases}\tag{6.50}$$

The positive parameters α and ε remain to be chosen, in order to obtain an exponential convergence, with the best possible rate. To fix α, let us first study the eigenvalues of the matrix of coefficients, neglecting the terms in ε.

Lemma 15 *Let us assume [H4]. The matrix*

$$A = \begin{pmatrix} -\rho & \rho \\ 4\alpha R^2 & -\left(1 - \frac{\tilde{M}}{2\alpha c}\right)\theta \end{pmatrix}$$

is negative definite and α may be chosen so that the eigenvalues $-\lambda_\pm$ of A are such that

$$-\lambda_- \leq -\lambda_+ = -\Lambda(\theta) < 0$$

where $\Lambda : (\theta_{\min}, \infty) \to (0, \rho)$ is a positive, increasing function. We recall that $\theta_{\min} = \dfrac{8\tilde{M}R^2}{c}$, where $R = C + M\rho^{-1/2}$. The function Λ is such that $\Lambda \to 0$ as $\theta \to \theta_{\min}$ and $\Lambda \to \rho$ as $\theta \to \infty$. Moreover, $\Lambda(\rho + 2\theta_{\min}) = \dfrac{\rho}{2}$.

Proof : In order to prove the negative definiteness of the matrix A, we show that for certain values of $\alpha > 0$, $\mathrm{tr}(A) < 0$ and $\det(A) > 0$. In the following, we only consider positive values of α (which is imposed by the previous computations). We have

$$\mathrm{tr}(A) = -\rho - \left(1 - \frac{\tilde{M}}{2\alpha c}\right)\theta < 0 \ \text{ iff } \ \alpha > \frac{\tilde{M}\theta}{2c(\rho + \theta)} \tag{6.51}$$

and

$$\det(A) = \theta\rho\left(1 - \frac{\tilde{M}}{2\alpha c}\right) - 4\alpha R^2\rho > 0 \text{ iff } \alpha \in (\alpha_-, \alpha_+),\ \alpha_\pm = \frac{\theta c \pm \sqrt{\theta^2 c^2 - 8\tilde{M}\theta R^2 c}}{8R^2 c}. \tag{6.52}$$

The interval (α_-, α_+) is indeed well defined and included in $[0, \infty)$ since $\theta > \theta_{\min} = 8\tilde{M}R^2/c$ (hypothesis [H4]). We seek an optimal α that minimizes eigenvalue $-\lambda_+$ and satisfies (6.51) and (6.52). An analytical solution cannot be easily obtained. We choose

$$\alpha = \alpha^* := \frac{\tilde{M}}{c},$$

which appears to be very close to the optimal choice, from numerical computations. Notice that α^* satisfies (6.51) and (6.52) since, for $\alpha = \alpha^*$, $\mathrm{tr}(A) = -\rho - \theta/2 < 0$ and $\det(A) = \theta\rho/2 - 4R^2\tilde{M}\rho/c > 0$. The eigenvalues of the matrix are now given by

$$-\lambda_\pm = \frac{1}{2}\left(-\left(\rho + \frac{\theta}{2}\right) \pm \sqrt{\left(\rho - \frac{\theta}{2}\right)^2 + \frac{16R^2\tilde{M}\rho}{c}}\right) \quad < 0.$$

The rate of convergence of the system is given by the largest of the two eigenvalues $-\lambda_+$. Let us introduce the function

$$\Lambda(\theta) = -\frac{1}{2}\left(-\left(\rho + \frac{\theta}{2}\right) + \sqrt{\left(\rho - \frac{\theta}{2}\right)^2 + \frac{16R^2\tilde{M}\rho}{c}}\right)$$

such that $\lambda_+ = \Lambda(\theta)$. It is easily shown that Λ is an increasing function of θ with

$$\Lambda(\theta) \to \left\{ \begin{array}{ll} 0 & \text{as } \theta \to \frac{8\tilde{M}R^2}{c}, \\ \rho & \text{as } \theta \to \infty. \end{array} \right.$$

Moreover, it is easy to check that $\Lambda(\rho + 2\theta_{\min}) = \dfrac{\rho}{2}$, which concludes the proof. $\diamondsuit$

We are now in position to complete the proof of Theorem 5. Let $Y(t) = (E_m(t), P(t))$ be the vector of solutions of (6.50). Using the fact that $E_m \leq \|Y\|_2$ (where $\|Y\|_2$ denotes the Euclidean norm of the two-dimensional vector Y), we obtain

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\|Y\|_2^2 &= \frac{1}{2}\frac{d}{dt}(E_m^2 + P^2) \\ &\leq 2Y^T A Y + 2R^2\varepsilon E_m^2 + \frac{1}{4\varepsilon}F_0\,\mathrm{e}^{-8\pi^2 t}E_m \\ &\leq -2\Lambda(\theta)\,\|Y\|_2^2 + 2R^2\varepsilon\|Y\|_2^2 + \frac{1}{4\varepsilon}F_0\,\mathrm{e}^{-8\pi^2 t}\|Y\|_2, \end{aligned}$$

and as a result,

$$\frac{d\|Y\|_2}{dt} \leq -2(\Lambda(\theta) - R^2\varepsilon)\|Y\|_2 + \frac{1}{4\varepsilon}F_0\,\mathrm{e}^{-8\pi^2 t}. \tag{6.53}$$

For arbitrary small $\varepsilon > 0$, let us consider $\lambda_\varepsilon = \Lambda(\theta) - R^2\varepsilon < \Lambda(\theta)$. We may assume without loss of generality that $\lambda_\varepsilon \neq 4\pi^2$. Then, from (6.53), one gets:

$$E_m \leq \|Y\|_2 \leq K_\varepsilon\,\mathrm{e}^{-2\min\{\lambda_\varepsilon, 4\pi^2\}t}, \tag{6.54}$$

where

$$K_\varepsilon = 2\,\max\left\{\sqrt{E_m^2(0) + P^2(0)}, \frac{F_0}{8\varepsilon|\lambda_\varepsilon - 4\pi^2|}\right\},$$

which concludes the proof of (6.27).

The exponential convergence of the total entropy E results from the relation $E = E_M + E_m$, (6.54) and Lemmas 9 and 10. The Csiszar-Kullback inequality implies the same for $\|\psi(t,\cdot) - \psi_\infty\|_{L^1}^2$.

Finally, the convergence results on A_t' are easily obtained from Lemma 8 and the fact that ψ^ξ is bounded from below by a positive constant for times larger than an arbitrary small positive time, see the beginning of Section 3.3.2 in [Lelièvre 2008] for more details.

Bibliography

[Ackland 2004] G.J. Ackland, M.I. Mendelev, D.J. Srolovitz, S. Han et A.V. Barashev. *Development of an interatomic potential for phosphorus impurities in α-iron.* Journal of Physics: Condensed Matter 16, page S2629, 2004. 10, 43

[Ané 2000] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto et G. Scheffer. Sur les inégalités de Sobolev logarithmiques. SMF, 2000. 73, 74, 75, 111, 125

[Arnold 2001] A. Arnold, P. Markowich, G. Toscani et A. Unterreiter. *On logarithmic Sobolev inequalities and the rate of convergence to equilibrium for Fokker-Planck type equations.* Comm. Partial Differential Equations, vol. 26, pages 35–43, 2001. 111, 116

[Arrhenius 1889] S. Arrhenius. *Über die Reaktionngeschwindigkeit bei des Inversion von Rohrzucker durch Säuren.* J. Phys. Chem., vol. 4, pages 226–248, 1889. 18

[Assaraf 2000] R. Assaraf, M. Caffarel et A. Khelif. *Diffusion Monte Carlo methods with a fixed number of walkers.* Phys. Rev. E, vol. 61, page 4566, 2000. 87

[Ayala 1997] P.Y. Ayala et H.B. Schlegel. *A combined method for determining reaction paths, minima and transition state geometries.* J. Chem. Phys., vol. 107, no. 2, pages 375–384, 1997. 39

[Bakry 1984] D. Bakry et M. Emery. *Hypercontractivité de semi-groupes de diffusion.* C. R. Acad. Sci Paris Sér. I, vol. 299, pages 775–778, 1984. 74, 112

[Banerjee 1985] A. Banerjee, N. Adams et J. Simons. *Search for Saddle points on Surfaces.* J. Phys. Chem, vol. 89, pages 52–57, 1985. 33

[Barkema 1996a] G.T. Barkema et N. Mousseau. *Event-Based Relaxation of Continuous Disordered Systems.* Phys. Rev. Lett., vol. 77, pages 4358–4361, 1996. 18, 34, 35

[Barkema 1996b] G.T. Barkema et N. Mousseau. *Event-based relaxation of continuous disordered systems.* Phys. Rev. Lett., vol. 77, no. 21, pages 4358–4361, 1996. 39

[Barkema 1998] G.T. Barkema et N. Mousseau. *Identification of relaxation and diffusion mechanisms in amorphous silicon.* Phys. Rev. Lett., vol. 81, no. 9, pages 1865–1868, 1998. 39

[Barkema 1999] G.T. Barkema et N. Mousseau. *Exploring structural mechanisms in disordered materials using the activation-relaxation technique.* Comp. Phys. Commun., vol. 121/122, pages 206–209, 1999. 35, 39

[Barkema 2001] G.T. Barkema et N. Mousseau. *The activation-relaxation technique: an efficient algorithm for sampling energy landscapes.* Comp. Mater. Sci., vol. 20, pages 285–292, 2001. 39

[Baskes 1992] M. I. Baskes. *Modified embedded-atom potentials for cubic materials and impurities.* Phys. Rev. B, vol. 46, no. 5, pages 2727–2742, 1992. 10

[Bennett 1976] C. H. Bennett. *Efficient estimation of free energy differences from Monte-Carlo data.* J. Comput. Phys., vol. 22, pages 245–268, 1976. 67, 81

[Bhandarkar 2003] M. Bhandarkar, R. Brunner, C. Chipot, A. Dalke, S. Dixit, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, J. Hénin, W. Humphrey, D. Hurwitz, N. Krawetz, S. Kumar, M. Nelson, J. Phillips, A. Shinozaki, G. Zheng et F. Zhu. *NAMD User's guide, version 2.5.* Theoretical biophysics group, University of Illinois and Beckman Institute, 405 North Mathews, Urbana, Illinois 61801, 2003. 23, 82, 90, 94

[Bobkov 1999] S. G. Bobkov et F. G otze. *Exponential Integrability and Transportation Cost Related to Logarithmic Sobolev Inequalities.* J. Funct. Anal., vol. 163, no. 1, pages 1–28, 1999. 74

[Bolhuis 2002a] P. G. Bolhuis, C. Dellago, P. L. Geissler et D. Chandler. *Transition Path Sampling: throwing ropes over dark mountain passes.* Ann. Rev. Phys. Chem., vol. 54, page 20, 2002. 60

[Bolhuis 2002b] P.G. Bolhuis, D. Chandler, C. Dellago et P.L. Geissler. *Transition path sampling: Throwing ropes over rough mountain passes, in the dark.* Annual review of physical chemistry 53, pages 291–318, 2002. 39

[Bortz 1975] A. B. Bortz, M. H. Kalos et J. L. Lebowitz. *A new algorithm for Monte Carlo simulation of Ising spin systems.* J. Comp. Phys., vol. 17, no. 1, pages 10–18, 1975. 19, 30

[Bowers 2006] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossváry, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Yibing Shan, et D. E. Shaw. *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters.* Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), 2006. 24

[Brooks 1983] B.R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan et M. Karplus. *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, vol. 4, pages 187–217, 1983. 9

[Brown 1828] R. Brown. *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic*

and inorganic bodies. Philosophical Magazine, vol. 4, pages 161–173, 1828. 13

[Bussi 2006] G. Bussi, A. Laio et M. Parrinello. *Equilibrium free energies from nonequilibrium metadynamics.* Phys. Rev. Lett., vol. 96, page 090601, 2006. 70

[C. Dellago 1999] P.G. Bolhuis C. Dellago et D. Chandler. *On the calculation of reaction rate constants in the transition path ensemble.* J. Chem. Phys., vol. 110, no. 14, pages 6617–6625, 1999. 39

[Cancès 2003] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris et Y. Maday. *Computational quantum chemistry: A primer.* Handbook of Numerical Analysis, vol. 10, pages 3–270, 2003. 6

[Cancès 2009] E. Cancès, F. Legoll, M.-C. Marinica, K. Minoukadeh et F. Willaime. *An improvement to the Activation-Relaxation Technique.* J. Chem. Phys., vol. 130, page 114711, 2009. 37, 39

[Carpenter 1999] J. Carpenter, P. Clifford et P. Fearnhead. *Building robust simulation based filters for evolving data sets.* Technical Report, Department of Statistics, University of Oxford, Oxford, UK, 1999. 88

[Carter 1989] E. A. Carter, G. Ciccotti, J. T. Hynes et R. Kapral. *Constrained reaction coordinate dynamics for the simulation of rare events.* Chem. Phys. Lett., vol. 156, pages 472–477, 1989. 63, 84

[Cerjan 1981] C.J. Cerjan et W.H. Miller. *On finding transition states.* J. Chem. Phys., vol. 75, pages 2800–2806, 1981. 18, 31, 33, 39, 41

[Chipot 2005] C. Chipot et J. Hénin. *Exploring the free-energy landscape of a short peptide using an average force.* J. Chem. Phys., vol. 123, page 244906, 2005. 94

[Chipot 2007] C. Chipot et A. Pohorille. Free energy calculations: Theory and applications in chemistry and biology. Springer, 2007. 81, 86

[Chopin 2010] N. Chopin, T. Lelièvre et G. Stoltz. *Free energy methods for efficient exploration of mixture posterior densities.* 2010. 66

[Ciccotti 2005] G. Ciccotti, E. Vanden-Eijnden et R. Kapral. *Blue Moon Sampling, Vectorial Reaction Coordinates and Unbiased Constrained Dynamics.* ChemPhysChem, vol. 200, page 1809, 2005. 63

[Ciccotti 2008] G. Ciccotti, T. Lelièvre et E. Vanden-Eijnden. *Projection of diffusions on submanifolds: Application to mean force computation.* Comm. Pure Appl. Math., vol. 61, page 3, 2008. 62, 63, 79, 84, 103

[Cornell 1995] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell et P. A. Kollman. *A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules.* J. Am. Chem. Soc., vol. 117, pages 5179–5197, 1995. 9

[Crehuet 2003] R. Crehuet et M.J. Field. *A temperature-dependent nudged-elastic-band algorithm.* J. Chem. Phys., vol. 118, no. 21, pages 9563–9571, 2003. 39

[Darden 1999] T. Darden, L. Perera, L. Li et L. Pedersen. *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations.* Structure, vol. 7, pages R55–R60, 1999. 9

[Darve 2001] E. Darve et A. Pohorille. *Calculating free energies using average forces.* J. Chem. Phys., vol. 115, no. 20, pages 9196–9183, 2001. 69, 71, 81, 82, 84, 85, 103

[Darve 2006] E. Darve. *Numerical Methods for Calculating the Potential of Mean Force.* New Algorithms for Macromolecular Simulation, LNCSE, pages 213–249, 2006. 62

[David E. Shaw 1998] Ron O. Dror Jeffrey S. Kuskin Richard H. Larson John K. Salmon Cliff Young Brannon Batson Kevin J. Bowers Jack C. Chao Michael P. Eastwood Joseph Gagliardo J.P. Grossman C. Richard Ho Douglas J. Ierardi István Kolossváry John L. Klepeis Timothy Layman Christine McLeavey Mark A. Moraes Rolf Mueller Edward C. Priest Yibing Shan Jochen Spengler Michael Theobald Brian Towles David E. Shaw Martin M. Deneroff et Stanley C. Wang. *Anton, A Special-Purpose Machine for Molecular Dynamics Simulation.* Communications of the ACM, vol. 51, no. 7, pages 91–97, 1998. 23

[Daw 1983] Murray S. Daw et M. I. Baskes. *Semiempirical, Quantum Mechanical Calculation of Hydrogen Embrittlement in Metals.* Phys. Rev. Lett., vol. 50, no. 17, pages 1285–1288, Apr 1983. 10

[Daw 1984] Murray S. Daw et M. I. Baskes. *Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals.* Phys. Rev. B, vol. 29, no. 12, pages 6443–6453, Jun 1984. 10

[Del Moral 2004] P. Del Moral. Feynman-kac formulae genealogical and interacting particle systems with applications. Springer, 2004. 88

[Dellago 1998] C. Dellago, P.G. Bolhuis, F.S. Csajka et D. Chandler. *Transition path sampling and the calculation of rare events.* J. Chem. Phys., vol. 108, pages 1964–1977, 1998. 39

[Dellago 2002] C. Dellago, P.G. Bolhuis et P.L. Geissler. *Transition path sampling.* Adv. Chem. Phys., vol. 123, pages 1–78, 2002. 39, 60

[den Otter 1998] W. K. den Otter et W. J. Briels. *The calculation of free energy differences by constrained molecular dynamics simulations.* J. Chem. Phys., vol. 109, page 4139, 1998. 62, 84, 103

[Dickson 2010] B. M. Dickson, F. Legoll, T. Lelièvre, G. Stoltz et P. Fleurat-Lessard. *Free Energy Calculations: An Efficient Adaptive Biasing Potential Method.* J. Phys. Chem. B, vol. 114, no. 17, pages 5823–5830, 2010. 71

[Douc 2005] R. Douc, O. Cappe et E. Moulines. *Comparison of resampling schemes for particle filtering.* Image and Signal Processing and Analysis, 2005. ISPA 2005, pages 64–69, 2005. 88

[Doucet 2001] A. Doucet, N. de Freitas et N. J. Gordon. Sequential Monte Carlo methods in practice. Series Statistics for Engineering and Information Science. Springer, 2001. 87

[E 2002] W. E, W. Ren et E. Vanden-Eijnden. *String method for the study of rare events.* Phys. Rev. B, vol. 66, page 052301, 2002. 18, 28, 39

[E 2003] W. E, W. Ren et E. Vanden-Eijnden. *Energy landscape and thermally activated switching of submicron-sized ferromagnetic elements.* J. Appl. Phys., vol. 93, page 2275, 2003. 28

[E 2004] W. E et E. Vanden-Eijnden. *Metastability, conformation dynamics, and transition pathways in complex systems.* Multiscale Modelling and Simulation, LNCSE, pages 35–68, 2004. 58

[E 2005a] W. E, W. Ren et E. Vanden-Eijnden. *Finite temperature String method for the study of rare events.* J. Phys. Chem. B, vol. 109, no. 14, pages 6688–6693, 2005. 19, 30, 39

[E 2005b] W. E, W. Ren et E. Vanden-Eijnden. *Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes.* Chem. Phy. Lett., vol. 413, pages 242–247, 2005. 29, 58

[E 2007] W. E, W. Ren et E. Vanden-Eijnden. *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events.* J. Chem. Phys, vol. 126, page 164103, 2007. 26, 28

[Elber 1987] R. Elber et M. Karplus. *A method for determining reaction paths in large molecules - application to myoglobin.* Chem. Phys. Lett., vol. 139, no. 5, pages 375–380, 1987. 39

[Evans 1935] M. Evans et M. Polanyi. *Some applications of the transition state method to the calculation of reaction velocities, especially in solution.* Trans. Faraday Soc., vol. 31, page 875, 1935. 19

[Ewald 1921] P. Ewald. *Die Berechnung optischer und elektrostatischer Gitterpotentiale.* Ann. Phys., vol. 369, pages 253–287, 1921. 9

[Eyring 1935] H. Eyring. *The Activated Complex in Chemical Reactions.* J. Chem. Phys., vol. 3, page 107, 1935. 19

[F. Montalenti 2001] M.R. Sorensen F. Montalenti et A.F. Voter. *Closing the gap between experiment and theory: Crystal growth by temperature accelerated dynamics.* Phys. Rev. Lett., vol. 87, page 126101, 2001. 40

[Freidlin 1998] M. I. Freidlin et A. D. Wentzell. Random perturbations of dynamical systems. 1998. 18, 24

[Frenkel 1996] D. Frenkel et B. Smit. Understanding molecular simulation. Academic Press, 1996. 84

[Greengard 1987] L. Greengard et V. Rokhlin. *A Fast Algorithm for Particle Simulations.* J. Comput. Phys., vol. 73, page 325, 1987. 9

[Gross 1975] L. Gross. *Logarithmic Sobolev Inequalities.* American Journal of Mathematics, vol. 97, pages 1061–1083, 1975. 74, 112

[Grubmüller 1995] H. Grubmüller. *Predicting slow structural transitions in macromolecular systems: confoirmational flooding.* Phys. Rev. E, vol. 52, no. 3, pages 2893–2906, 1995. 39

[H. Jonsson 1998] K.W. Jacobsen H. Jonsson G. Mills. Nudged elastic band method for finding minimum energy paths of transitions in classical and quantum dynamics in condensed phase simulations, volume 4. 1998. 18, 25, 26, 39

[Hairer 2003] E. Hairer, C. Lubich et G. Wanner. *Geometric numerical integration illustrated by the Störmer-Verlet method.* Acta Numerica, vol. 12, pages 399–450, 2003. 13

[Hastings 1970] W. K. Hastings. *Monte Carlo sampling methods using Markov chains and their applications.* Biometrika, vol. 57, pages 97–109, 1970. 10

[Hénin 2004] J. Hénin et C. Chipot. *Overcoming free energy barriers using unconstrained molecular dynamics simulations.* J. Chem. Phys., vol. 121, pages 2904–2914, 2004. 71, 81, 82, 94, 95, 103

[Hénin 2010] J. Hénin, G. Fiorin, C. Chipot et M. L. Klein. *Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables.* J. Chem. Theory Comput., vol. 6, pages 35–47, 2010. 82, 95, 105

[Henkelman 1999] G. Henkelman et H. Jonsson. *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives.* J. Chem. Phys., vol. 111, page 7010, 1999. 18, 31, 32, 33, 39

[Henkelman 2000a] G. Henkelman et H. Jonsson. *Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points.* J. Chem. Phys., vol. 113, page 9978, 2000. 26, 39

[Henkelman 2000b] G. Henkelman, B.P. Uberuaga et H. Jónsson. *A climbing image nudged elastic band method for finding saddle points and minimum energy paths.* J. Chem. Phys., vol. 113, page 9901, 2000. 30

[Henkelman 2001] G. Henkelman et H. Jónsson. *Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table.* J. Chem. Phys., vol. 115, no. 21, page 9657, 2001. 32

[Heyden 2005] A. Heyden, A. T. Bell et F. J. Keil. *Efficient methods for finding transition states in chemical reactions: Comparison of improved dimer method and partitioned rational function optimization method.* J. Chem. Phys, vol. 123, page 224101, 2005. 18, 33

[Holley 1987] R. Holley et D. Stroock. *Logarithmic Sobolev inequalities and stochastic Ising models.* J. Stat. Phys., vol. 46, pages 1159–1194, 1987. 74, 112

[Iannuzzi 2003] M. Iannuzzi, A. Laio et M. Parrinello. *Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics.* Phys. Rev. Lett., vol. 90, no. 23, page 238302, 2003. 69, 81

[Jarzynski 1997] C. Jarzynski. *Nonequilibrium equality for free energy differences.* Phys. Rev. Lett., vol. 78, no. 14, pages 2690–2693, 1997. 67, 81

[Jourdain 2009] B. Jourdain, T. Lelièvre et R. Roux. *Existence, uniqueness and convergence of a particle approximation for the Adaptive Biasing Force process.* In process, 2009. 104

[Kale 1999] L. Kale, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. C. Phillips, A. Shinozaki, K. Varadarajan, et K. Schulten. *NAMD2: Greater scalability for parallel molecular dynamics.* J. Comp. Phys., vol. 151, pages 283–312, 1999. 23, 82, 94

[Kirkwood 1935] J. G. Kirkwood. *Statistical Mechanics of Fluid Mixtures.* J. Chem. Phys., vol. 3, no. 5, pages 300–313, 1935. 60, 81

[Kitagawa 1996] G. Kitagawa. *Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models.* J. Comput. Graph. Statist., vol. 5, no. 1, pages 1–25, 1996. 88

[Kohn 1965] W. Kohn et L. J. Sham. *Self-Consistent Equations Including Exchange and Correlation Effects.* Phys. Rev., vol. 140, no. 4A, pages A1133–A1138, 1965. 6

[Kumar 1992] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman et J. M. Rosenberg. *The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* J. Comput. Chem., vol. 13, no. 8, pages 1011–1021, 1992. 81

[Laio 2002] A. Laio et M. Parrinello. *Escaping free-energy minima.* Proc. Natl. Acad. Sci. U.S.A, vol. 99, page 12562, 2002. 70

[Langevin 1908] P. Langevin. *Sur la théorie du mouvement brownien.* Comptes-Rendus de l'Académie des Sciences, vol. 146, pages 530–532, 1908. 13, 83

[Larson 2009] R. H. Larson, J. K. Salmon, R. O. Dror, M. M. Deneroff, C. Young, J.P. Grossman, J. L. Klepeis Y. Shan et D. E. Shaw. *High-Throughput Pairwise Point Interactions in Anton, a Specialized Machine for Molecular Dynamics Simulation.* Proceedings of the 14th Annual International Symposium on High-Performance Computer Architecture (HPCA '08), 2009. 23

[Lelièvre 2007a] T. Lelièvre, M. Rousset et G. Stoltz. *Computation of free energy differences through nonequilibrium stochastic dynamics: The reaction coordinate case.* J. Comput. Phys., vol. 222, no. 2, pages 624–643, 2007. 82

[Lelièvre 2007b] T. Lelièvre, M. Rousset et G. Stoltz. *Computation of free energy profiles with parallel adaptive dynamics.* J. Chem. Phys., vol. 126, page 134111, 2007. 73, 85, 86, 87, 88, 104, 110

[Lelièvre 2008] T. Lelièvre, M. Rousset et G. Stoltz. *Long-time convergence of an adaptive biasing force method.* Nonlinearity, vol. 21, no. 6, pages 1155–1181, 2008. 69, 77, 79, 85, 102, 104, 105, 106, 107, 110, 112, 113, 116, 119, 122, 132

[Lelièvre 2009] T. Lelièvre. *A general two-scale criteria for logarithmic Sobolev inequalities.* J.Func. Anal., vol. 256, no. 7, pages 2211–2221, 2009. 113

[Lelièvre 2010a] T. Lelièvre et K. Minoukadeh. *Long-time convergence of the ABF method: the bi-channel case.* Submitted to Archive for Rational and Mechanical Analysis, 2010. 101

[Lelièvre 2010b] T. Lelièvre, M. Rousset et G. Stoltz. Free energy computations : A mathematical perspective. Imperial College Press, 2010. 61, 67, 70, 81

[Lelièvre 2010c] T. Lelièvre, M. Rousset et G. Stoltz. *Langevin dynamics with constraints and computation of free energy differences.* 2010. 63

[Lennard-Jones 1931] J. E. Lennard-Jones. *Cohesion.* Proceedings of the Physical Society, vol. 43, pages 461–482, 1931. 8

[MacKerel Jr. 1998] A.D. MacKerel Jr., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won et M. Karplus. Charmm: The energy function and its parameterization with an overview of the program, volume 1 of *The Encyclopedia of*

Computational Chemistry, pages 271–277. John Wiley & Sons: Chichester, 1998. 9

[MacKerell 1998] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin et M. Karplus. *All-atom empirical potential for molecular modeling and dynamics studies of protein.* J. Phys. Chem. B, vol. 102, pages 3586–3616, 1998. 94

[Maeda 2003] S. Maeda et K. Ohno. *A new method for constructing multidimensional potential energy surfaces by a polar coordinate interpolation technique.* Chem. Phys. Lett., vol. 381, no. 1-2, pages 177–186, 2003. 40

[Maeda 2005] S. Maeda, Y. Watanabe et K. Ohno. *A scaled hypersphere interpolation technique for efficient construction of multidimensional potential energy surfaces.* Chem. Phys. Lett., vol. 414, no. 4-6, pages 265–270, 2005. 40

[Malek 2000] R. Malek et N. Mousseau. *Dynamics of Lennard-Jones clusters: A characterization of the activation-relaxation technique.* Phys. Rev. E, vol. 62, pages 7723–7728, 2000. 18, 31, 34, 35, 40, 44

[Marinica 2008] M.-C. Marinica, F. Williame et N. Mousseau. *Small Interstitial clusters in iron BCC structure.* 2008. 37, 38, 43, 44, 45, 46

[Marsili 2006] S. Marsili, A. Barducci, R. Chelli, P. Procacci et V. Schettino. *Self-healing umbrella sampling: A non-equilibrium approach for quantitative free energy calculations.* J. Phys. Chem. B, vol. 110, no. 29, pages 14011–14013, 2006. 71

[Mendelev 2003] M. Mendelev, S. Han, D.J. Sroloviz, G. Ackland, D. Sun et M. Asta. *Development of new interatomic potentials appropriate for crystalline and liquid iron.* Phil. Mag, vol. 83, page 3977, 2003. 10, 43

[Mengersen 1996] K. L. Mengersen et R. L. Tweedie. *Rates of convergence of the Hastings and Metropolis algorithms.* The Annals of Statistics, vol. 24, no. 1, pages 101–121, 1996. 11

[Metropolis 1953] N. Metropolis, A. W. Rosenbluth, M.N. Rosenbluth, A. H. Teller et E. Teller. *Equations of state calculations by fast computing machines.* J. Chem. Phys., vol. 21, no. 6, pages 1087–1091, 1953. 10

[Metzner 2007] P. Metzner. *Transition Path Theory for Markov Processes.* PhD Thesis, 2007. 60

[Meyn 1993] S.P. Meyn et R.L. Tweedie. Markov chains and stochastic stability. Springer-Verlag, London, 1993. 11

[Mills 1994] G. Mills et H. Jonsson. *Quantum and thermal effects in H_2 dissociative adsorption: evaluation of free energy barriers in multidimensional quantum systems.* Phys. Rev. Lett., vol. 72, no. 7, pages 1124–1128, 1994. 39

[Minoukadeh 2010] K. Minoukadeh, C. Chipot et T. Lelièvre. *Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach.* J. Chem. Theory Comput., vol. 6, pages 1008–1017, 2010. 73, 81, 105, 107

[Miron 2004] R.A. Miron et K.A. Fichthorn. *Multiple-time scale accelerated molecular dynamics: Addressing the small-barrier problem.* Phys. Rev. Lett., vol. 93, no. 12, page 128301, 2004. 40

[Montalenti 2002] F. Montalenti et A.F. Voter. *Exploiting past visits or minimum-barrier knowledge to gain further boost in the temperature-accelerated dynamics.* J. Chem. Phys., vol. 116, no. 12, pages 4819–4828, 2002. 40

[Mousseau 1998] N. Mousseau et G.T. Barkema. *Traveling through potential energy surfaces of disordered materials: the activation-relaxation technique.* Phys. Rev. E, vol. 57, page 2419, 1998. 18, 34, 35, 39

[Mousseau 1999] N. Mousseau et G.T. Barkema. *Exploring high-dimensional energy landscapes.* Comput. Sci. Eng., vol. 1, pages 74–80, 1999. 39

[Mousseau 2000a] N. Mousseau. *Cooperative motion in Lennard-Jones binary mixtures below the glass transition.* 2000. 35

[Mousseau 2000b] N. Mousseau et G.T. Barkema. *Activated mechanisms in amorphous silicon: An activation-relaxation-technique study.* Phys. Rev. B, vol. 61, pages 1898–1906, 2000. 18, 31, 34, 39, 40, 44

[Munro 1999] L.J. Munro et D.J. Wales. *Defect migration in crystalline silicon.* Phys. Rev. B, vol. 59, pages 3969–3980, 1999. 33, 39

[Murrell 1968] J. N. Murrell et K. J. Laidler. Trans. Faraday Soc., vol. 64, page 371, 1968. 25

[Nocedal 2000] J. Nocedal et S. J. Wright. Numerical optimization. 2000. 43

[Ohno 2004] K. Ohno et S. Maeda. *A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface.* Chem. Phys. Lett., vol. 384, no. 4-6, pages 277–282, 2004. 40

[Otto 2000] F. Otto et C. Villani. *Generalization of an inequality by Talagrand, viewed as a consequence of the logarithmic Sobolev inequality.* J. Funct. Anal., vol. 173, no. 2, pages 361–400, 2000. 76, 112

[Park 2003] S. Park, F. Khalili-Araghi, E. Tajkhorshid et K. Schulten. *Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality.* J. Chem. Phys., vol. 119, page 3559, 2003. 94, 95

[Peters 2004] B. Peters, A. Heyden, A.T. Bell et A. Chakraborty. *A growing string method for determining transition states: comparison to the nudged elastic band and string methods.* J. Chem. Phys., vol. 120, page 7877, 2004. 39

[Phillips 2005] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale et Klaus Schulten. *Scalable Molecular Dynamics with NAMD.* J. Comp. Chem., vol. 26, no. 16, pages 1781–1802, 2005. 23, 82, 90, 94

[Roberts 1998] G.O. Roberts et J.S. Rosenthal. *Optimal scaling of discrete approximations to Langevin diffusions.* J. Roy. Stat. Soc. B, vol. 60, pages 255–268, 1998. 11

[Rodrigues 2008] C. I. Rodrigues, D. J. Hardy, J. E. Stone, K. Schulten et W. W. Hwu. *GPU acceleration of cutoff pair potentials for molecular modeling applications.* In CF '08: Proceedings of the 5th conference on Computing frontiers, pages 273–282. ACM, 2008. 24

[Shan 2005] Y Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror et D. E. Shaw. *Gaussian split Ewald: A fast Ewald mesh method for molecular simulation.* J. Chem. Phys., vol. 122, no. 5, page 54101, 2005. 23

[Shaw 2008] David E. Shaw, Martin M. Deneroff, Ron O. Dror, Jeffrey S. Kuskin, Richard H. Larson, John K. Salmon, Cliff Young, Brannon Batson, Kevin J. Bowers, Jack C. Chao, Michael P. Eastwood, Joseph Gagliardo, J. P. Grossman, C. Richard Ho, Douglas J. Ierardi, István Kolossváry, John L. Klepeis, Timothy Layman, Christine McLeavey, Mark A. Moraes, Rolf Mueller, Edward C. Priest, Yibing Shan, Jochen Spengler, Michael Theobald, Brian Towles et Stanley C. Wang. *Anton, a special-purpose machine for molecular dynamics simulation.* Commun. ACM, vol. 51, no. 7, pages 91–97, 2008. 23

[Shirts 2000] M. R. Shirts et V. S. Pande. *Screen Savers of the World, Unite!* Science, vol. 290, pages 1903–1904, 2000. 23

[Shirts 2001] M. R. Shirts et V. S. Pande. *Mathematical Analysis of Coupled Parallel Simulations.* Phys. Rev. Lett., vol. 86, no. 22, pages 4983–4987, 2001. 23

[Shirts 2008] M. R. Shirts et J. D. Chodera. *Statistically optimal analysis of samples from multiple equilibrium states.* J. Chem. Phys., vol. 124, no. 12, page 124105, 2008. 67, 81

[Sorensen 2000] M.R. Sorensen et A.F. Voter. *Temperature accelerated dynamics for simulation of infrequent events.* J. Chem. Phys., vol. 112, page 9599, 2000. 40

[Sprik 1998] M. Sprik et G. Cicotti. *Free energy from constrained molecular dynamics.* J. Chem. Phys., vol. 109, pages 7737–7744, 1998. 62, 84, 103

[Stone 2007] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco et K. Schulten. *Accelerating molecular modeling applications with graphics processors.* J. Comp. Chem., vol. 28, pages 2618–2640, 2007. 24

[Terentyev 2007] D.A. Terentyev, L. Malerba et M. Hou. *Dimensionality of interstitial cluster motion in bcc-Fe.* Phys. Rev B, vol. 75, page 104108, 2007. 43, 44

[Terentyev 2008] D.A. Terentyev, T.P.C. Klaver, P. Olsson, M.-C. Marinica, F. Willaime, C. Domain et L. Malerba1. *Self-Trapped Interstitial-Type Defects in Iron.* Phys. Rev. Lett., vol. 100, page 145503, 2008. 43

[Vanden-Eijnden 2006] E. Vanden-Eijnden. *Transition Path Theory.* Lect. Notes Phys., vol. 703, pages 439–478, 2006. 60

[Vanden-Eijnden 2009a] E. Vanden-Eijnden et M. Venturoli. *Markovian milestoning with Voronoi tessellations.* J. Chem. Phys., vol. 130, page 194101, 2009. 30

[Vanden-Eijnden 2009b] E. Vanden-Eijnden et M. Venturoli. *Revisiting the finite temperature string method for the calculation of reaction tubes and free energies.* J. Chem. Phys., vol. 130, page 194103, 2009. 30

[Verlet 1967] L. Verlet. *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.* Phys. Rev., vol. 159, no. 1, page 98, 1967. 13

[Villani 2003] C. Villani. Topics in Optimal Transportation. American Mathematical Society, 2003. 73, 111

[Vocks 2005] H. Vocks, M.V. Chubynsky, G.T. Barkema et N. Mousseau. *Activated sampling in complex materials at finite temperature: the properly-obeying-probability activation-relaxation technique.* J. Chem. Phys., vol. 123, no. 24, page 244707, 2005. 40

[Voter 1997a] A.F. Voter. *Hyperdynamics: accelerated molecular dynamics of infrequent events.* Phys. Rev. Lett., vol. 78, page 3908, 1997. 40

[Voter 1997b] A.F. Voter. *A method for accelerating the molecular dynamics simulation of infrequent events.* J. Chem. Phys., vol. 106, page 4665, 1997. 40

[Voter 1998] A.F. Voter. *Parallel replica method for dynamics of infrequent events.* Phys. Rev. B, vol. 57, page R13985, 1998. 40

[Wales 2002] D.J. Wales. *Discrete path sampling.* Mol. Phys., vol. 100, no. 2, pages 3285–3305, 2002. 39

[Wales 2003] D.J. Wales. Energy landscapes. 2003. 33, 39

[Wales 2005] D. J. Wales. *A Simplified Eigenvector-Following Technique for Locating Transition Points in an Energy Landscape.* J. Phys. Chem., vol. 109, pages 9578–9583, 2005. 33

[Wang 2001a] F. Wang et D. P. Landau. *Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram.* Phys. Rev. E, vol. 64, page 056101, 2001. 70

[Wang 2001b] F. Wang et D. P. Landau. *Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States.* Phys. Rev. Lett., vol. 86, no. 10, pages 2050–2053, Mar 2001. 69, 70, 81, 84

[Young 1966] W. M. Young et E. W. Elcock. *Monte Carlo studies of vacancy migration in binary ordered alloys: I.* Proc. Phys. Soc., vol. 89, page 735, 1966. 19, 30

[Zwanzig 1954] R. W. Zwanzig. *High-temperature equation of state by a perturbation method I. Nonpolar gases.* J. Chem. Phys., vol. 22, no. 8, pages 1420–1426, 1954. 64, 81

...