High dimensional Markov Chain Monte Carlo Methods: theory, methods and application

Alain Durmus

Université Paris-Saclay
École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : Telecom ParisTech
Établissement d'accueil : *Laboratoire d'accueil :* Laboratoire traitement et communication de l'information, UMR 5141 CNRS

THÈSE DE DOCTORAT ÈS MATHÉMATIQUES

Spécialité : Mathématiques appliquées

Alain Durmus

High dimensional Markov Chain Monte Carlo Methods: theory, methods and applications
Méthodes de Monte Carlo par chaîne de Markov en grandes dimensions : théorie, méthodes et applications

Date de soutenance : 2 Décembre 2016

Après avis des rapporteurs :	Andreas Eberle (University of Bonn)
	Arnaud Guillin (Université Blaise Pascal)

Jury de soutenance :	Francis Bach	(INRIA & ENS Paris) Examinateur
	Arnak Dalalyan	(ENSAE) Examinateur
	Andreas Eberle	(University of Bonn) Rapporteur
	Gersende Fort	(Telecom ParisTech) Codirectrice de thèse
	Arnaud Guillin	(Université Blaise Pascal) Rapporteur
	Éric Moulines	(École polytechnique) Directeur de thèse
	Jean-Christophe Pesquet	(CentraleSupelec) Examinateur
	Eero Saksman	(University of Helsinki) Examinateur

NNT :

"When you knocked upon my door
And I said hello Satan
I believe it's time to go"

Me and the Devil,
Robert Johnson

"Fils de pécore et de minus
Ris pas de la pauvre Vénus"

La complainte des filles de joie,
Georges Brassens

Acknowledgements
Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse Éric Moulines. Merci pour tout le savoir qu'il m'a transmis et le temps qu'il m'a accordé pour m'apprendre le métier de chercheur. Je lui suis aussi reconnaissant des sujets qu'il m'a proposés. Ils m'ont chacun passionné. Enfin je le remercie aussi pour toutes les discussions mathématiques ou non, ses conseils, sa bonne humeur..., cela a toujours été un réel plaisir de travailler avec lui. Je sais que même ses rares remises à l'ordre n'était que pour mon bien et cela ne m'a que plus motivé !

J'adresse toute ma reconnaissance à ma co-directrice de thèse Gersende Fort. Son aide fut très précieuse lors du commencement de ma thèse et mes recherches. Je pense que le travail que j'ai pu effectuer pendant ces trois ans lui doit beaucoup.

I am thankful to Arnaud Guillin and Andreas Eberle to have accepted to review this manuscript. It is a real honour. I thank them also for the nice and productive visits at Bonn.

Je remercie aussi les autres membres du jury, Francis Bach, Arnak Dalalyan et Jean-Michel Pesquet d'avoir accepté de participer à la soutenance de cette thèse. J'espère que M. Bach et M. Dalalyan apprécieront mon travail comme j'ai apprécié leur cours au Master MVA (bien que M. Dalalyan semble persuader que j'étais un fantôme !).

I thank Eero Saskman to have accepted to attend to my defense. We have met each other several times and it was always a real pleasure to talk classical music, mathematics and life with him.

I am very grateful to the professor Gareth Roberts to have introduced me to the subject of the optimal scaling of Metropolis-Hastings algorithms. I also thank him for a five months visit at the university of Warwick.

I thank my co-authors, Sylvain Le Corff, Gilles Vilmart and Kostantinos Zygalakis. It was a great pleasure to work with them.

I thank Umut Simsekli for the very good team we make ! I really hope that we will be able to visit the country together in better times.

Je remercie François Roueff pour m'avoir permis d'enseigner au cours de ma thèse. J'ai énormément apprécié travailler avec lui et l'activité d'enseignant. Je remercie aussi les autres chercheurs du laboratoire STA pour lesquels je suis intervenu dans leur cours : Anne Sabourin, Olivier Fercoq et Stephan Clémençon.

Merci Randal pour les différentes pauses clopes que nous avons faites ensemble et ses vidéos hilarantes !

I thank the members of the cryptography team at ENS Paris, who first introduce me to research, David Pointcheval, Vadim Lyubashevsky, Léo Ducas and Tancrède Lepoint.

Je remercie les différentes personnes du laboratoire du LTCI que j'ai pu côtoyer au cours de ces trois années, qui sont trop nombreuses pour être toutes listées ici !

Je remercie mes professeurs de classes préparatoires, M. Peronnet et M. Dora, pour m'avoir donné le goût des mathématiques. C'est grâce aux bases qu'ils m'ont enseignées que j'en suis arrivé là.

Je remercie Quentin, Paul et Romain pour leur amitié vieille de plus de vingt ans qui n'a pas pris une ride. J'ai de la chance de vous avoir les gars. Je remercie aussi tous les autres bien sûr !

Je remercie ma mère pour l'éducation qu'elle m'a donnée et toutes les attentions qu'elle a pu m'accorder. Merci aussi pour le support qu'elle me donne encore aujourd'hui.

Je remercie Hélène de supporter mes nombreuses absences. Merci pour les joies et les peines quotidiennes qu'elle m'apporte. Merci pour son affection et sa tendresse.

Contents

1 Introduction en français **1**
- 1.1 Motivations, un petit détour par les statistiques bayésiennes . . . 1
- 1.2 Algorithmes de Monte Carlo et Monte Carlo par chaînes de Markov . . . 2
- 1.3 Les échantillonneurs de type Metropolis-Hastings . . . 3
- 1.4 La diffusion de Langevin . . . 5
- 1.5 Convergence des méthodes MCMC . . . 8
- 1.6 Échelonnage optimal d'algorithmes de type Metropolis-Hastings . . . 14
- 1.7 Plan du manuscrit et contributions . . . 18

2 Introduction **31**
- 2.1 Motivation: a short digression on Bayesian statistics . . . 31
- 2.2 Monte Carlo algorithms and Markov Chain Monte Carlo . . . 32
- 2.3 Metropolis-Hastings type samplers . . . 33
- 2.4 The overdamped Langevin diffusion . . . 35
- 2.5 Convergence of MCMC methods . . . 38
- 2.6 Optimal scaling for Metropolis-Hastings type algorithms . . . 44
- 2.7 Outlines and contributions . . . 48

I Convergence of Markov chains and applications to MCMC **57**

3 Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis Adjusted Langevin Algorithm **59**
- 3.1 Quantitative bounds in Wasserstein distance . . . 61
- 3.2 Application to the EI-MALA algorithm . . . 67
- 3.3 Simulation . . . 71
- 3.4 Proofs . . . 75

4 Subgeometric rates of convergence in Wasserstein distance for Markov chains **83**
- 4.1 Introduction . . . 83
- 4.2 Main results . . . 86

4.3 Application . . . 89
4.4 Proofs of Section 4.2 . . . 95
4.5 Proofs of Section 4.3.3 . . . 105
4.6 Wasserstein distance: some useful properties . . . 110
4.7 Subgeometric functions and sequences . . . 111

II Study of ULA **113**

5 Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm **115**
5.1 Introduction . . . 115
5.2 General conditions for the convergence of ULA . . . 117
5.3 Practical conditions for geometric ergodicity of the Langevin diffusion and their consequences for ULA . . . 123
5.4 Proofs . . . 133
5.5 Quantitative convergence bounds in total variation for diffusions . . . 144

6 Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm **151**
6.1 Introduction . . . 152
6.2 Non-asymptotic bounds in Wasserstein distance of order 2 for ULA . . . 154
6.3 Mean square error and concentration for Lipschitz functions . . . 159
6.4 Quantitative bounds in total variation distance . . . 162
6.5 Mean square error and concentration for bounded measurable functions . 165
6.6 Convergence in Wasserstein distance of order p . . . 167
6.7 Numerical experiments . . . 168
6.8 Proofs . . . 171
6.9 Contraction results in total variation for some functional autoregressive models . . . 194

III Optimal scaling of Metropolis-Hastings type algorithms **201**

7 Optimal scaling of the Random Walk Metropolis algorithm under L^p mean differentiability **203**
7.1 Introduction . . . 203
7.2 Positive Target density on the real line . . . 205
7.3 Target density supported on an interval . . . 209
7.4 Proofs of Section 7.2 . . . 214
7.5 Proofs of Section 7.3 . . . 231

8 Fast Langevin based algorithm for MCMC in high dimensions **245**

8.1 Introduction 245
8.2 Preliminaries 248
8.3 Main scaling results 253
8.4 Geometric ergodicity results for high order Langevin schemes 255
8.5 Numerical illustration of the improved efficiency 265
8.6 Proof of Theorems 8.4 and 8.5 270
8.7 Postponed proofs 278

A Markov processes **III**

A.1 Markov chains III
A.2 Markov processes in continuous time VII
A.3 Results on diffusions VIII

B Supplement to "Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm" **IX**

B.1 Discussion of Theorem 6.4 IX
B.2 Discussion of Theorem 6.7 XII
B.3 Generalization of Theorem 6.4 XIII
B.4 Explicit bounds on the MSE XXI
B.5 Explicit bounds based on Theorem 6.4 XXI

C Appendix of Chapter 8 **XXIX**

Bibliography **i**

List of Figures **xiii**

List of Tables **xv**

Chapter 1

Introduction en français

1.1 Motivations, un petit détour par les statistiques bayésiennes

L'échantillonnage d'une loi de probabilité est la principale motivation des travaux de cette thèse. Ce problème provient de diverses applications. En particulier, l'inférence bayesienne repose sur l'exploration d'une loi a posteriori.

La statistique bayésienne suppose un modèle probabiliste sur des données d'observation w: w est supposé être une réalisation d'une variable aléatoire W à valeurs dans un espace mesurable $(\mathsf{W}, \mathcal{W})$ et on considère un modèle paramétrique pour W, $(\mathsf{W}, \mathcal{W}, \mathcal{P}_\Theta)$ où $\mathcal{P}_\Theta$ est un ensemble de mesures de probabilité défini par:

$$\mathcal{P}_\Theta = \{K(\vartheta, \cdot) \mid \vartheta \in \Theta\} \ .$$

K est un noyau de transition sur $(\mathsf{W}, \mathcal{W})$ et $(\Theta, \mathcal{F})$ est un ensemble mesurable. Dans la plupart des applications, Θ est soit un espace discret, une partie de $\mathbb{R}^d$, $d \geq 1$, ou un espace fonctionnel. On suppose que le modèle est dominé par une mesure μ sur $(\mathsf{W}, \mathcal{W})$, i.e. K admet une densité de transition par rapport à μ: il existe une fonction mesurable $\mathcal{L} : \Theta \times \mathsf{W} \to \mathbb{R}_+$ telle que pour tout $\vartheta \in \Theta$ et $w \in \mathsf{W}$,

$$\frac{\mathrm{d}K(\vartheta, \cdot)}{\mathrm{d}\mu}(w) = \mathcal{L}(w|\vartheta) \ .$$

La fonction $\vartheta \mapsto \mathcal{L}(W|\vartheta)$ s'appelle la vraisemblance du modèle. Alors que en statistique fréquentiste, on infèrerait le paramètre ϑ en maximisant la vraisemblance, les statistiques bayesiennes considèrent le paramètre ϑ comme lui-même une réalisation d'une variable aléatoire θ, et posent une loi de probabilité a priori sur cette dernière, que l'on notera ν_θ. Le théorème de Bayes [Sch95, Théorème 1.31] donne alors l'expression de la loi conditionnelle de θ sachant W, appelée loi a posteriori, en fonction de la vraisemblance et de la loi a priori ν_θ. Cette loi conditionnelle admet une densité de transition par rapport à ν_θ qui est donnée pour ν_θ-presque tout ϑ et μ-presque tout w par

$$\mathrm{p}_{\theta|W}(w, \vartheta) = \frac{\mathcal{L}(w|\vartheta)}{\mathrm{p}_W(w)} \ , \tag{1.1}$$

où p_W est la densité marginale de W donnée pour tout $w \in \mathsf{W}$ par

$$\mathrm{p}_W(w) = \int_\Theta \mathcal{L}(w|\vartheta)\nu_\theta(\mathrm{d}\vartheta) . \tag{1.2}$$

Si w sont des données d'observation, la mesure de probabilité associée à la densité $\mathrm{p}_{\theta|W}(\mathrm{w}, \cdot)$ est appelée la distribution a posteriori.

1.2 Algorithmes de Monte Carlo et Monte Carlo par chaînes de Markov

Soit maintenant une mesure de probabilité π sur un espace d'état mesurable $(\mathsf{E}, \mathcal{E})$ et $f : \mathsf{E} \to \mathbb{R}$ une fonction intégrable par rapport à π. La mesure de probabilité π sera appelée distribution cible. On s'intéresse à l'estimation de la quantité $\int_\mathsf{E} f(x)\pi(\mathrm{d}x)$. Gardons à l'esprit que la plupart du temps E est soit discret ou une partie de $\mathbb{R}^d$. Les méthodes de Monte Carlo classiques sont basées sur la loi forte de grands nombres pour les suites indépendantes et identiquement distribuées (i.i.d.) de loi π : soient $(Y_i)_{i\in\mathbb{N}}$ une suite i.i.d. de loi π alors $\int_\mathsf{E} f(x)\pi(\mathrm{d}x)$ est approchée par la suite d'estimateurs définie pour tout $N \in \mathbb{N}^*$ par

$$\hat{f}_N = \frac{1}{N}\sum_{i=0}^{N-1} f(Y_i) . \tag{1.3}$$

Par la loi forte des grands nombres $\hat{f}_N$ converge presque sûrement vers $\int_\mathsf{E} f(x)\pi(\mathrm{d}x)$ lorsque N tend vers l'infini, et si f^2 est intégrable par rapport à π, alors le théorème central limite donne une expression de l'erreur asymptotique. Bien que cette procédure soit très simple, elle requiert des échantillons i.i.d. de loi π. Pour cela, sauf cas triviaux, les méthodes les plus connues sont la méthode de la transformée inverse et la méthode de rejet, voir [RC10, section 2.1.2] et [RC10, section 2.3].

Cependant lorsque la dimension de l'espace d'état devient grand, ces méthodes deviennent vite inefficaces. De plus dans le cadre de l'inférence bayésienne, π admet une densité par rapport à la loi a priori (1.1) donnée avec les notations de la Section 1.1 par:

$$x \mapsto \frac{\mathcal{L}(\mathrm{w}|x)}{\mathrm{p}_W(\mathrm{w})} ,$$

où w sont les données d'observation. On peut observer que cette densité n'est connue qu'à une constante multiplicative près. Sauf dans le cas de lois dites conjuguées [Rob06, section 3.3], la densité marginale p_W définie par (1.2) n'a pas de forme analytique. Cette contrainte rend d'autant plus difficile l'application des deux méthodes mentionnées ci-dessus.

Une autre classe de méthode de Monte Carlo sont les méthodes de Monte Carlo par chaînes de Markov. La suite de variables aléatoires $(Y_i)_{i\in\mathbb{N}}$ peut ne plus être i.i.d. et être corrélée. Plus précisément, un estimateur de $\int_\mathsf{E} f(x)\pi(\mathrm{d}x)$ est toujours défini par (1.3), mais $(Y_i)_{i\in\mathbb{N}}$ est une chaîne de Markov de noyau P et de loi invariante π. Tout comme dans le cas i.i.d., sous de bonnes hypothèses sur π et le noyau P, une loi forte des grands nombres et un théorème central limite peuvent être établis [MT09, chapitre 17], ce qui justifie ces méthodes. En effet, [MT09, Theorem 17.1.7] montre que si P est Harris récurrent[1] et admet π comme loi invariante, alors pour tout $x \in \mathsf{E}$

$$N^{-1} \sum_{i=0}^{N-1} f(Y_i) = \int_{\mathsf{E}} f(y)\pi(\mathrm{d}y) \;, \qquad \mathbb{P}_x\text{-presque sûrement} \;,$$

où $\mathbb{P}_x$ est la loi induite par le noyau P et la distribution initiale δ_x sur $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ et $(Y_i)_{i \in \mathbb{N}}$ est la chaîne canonique sur $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \mathbb{P}_x)$. Un premier problème quant à ces méthodes est de trouver un noyau de transition admettant π comme loi invariante. Les algorithmes de type Metropolis-Hastings sont des méthodes permettant de construire de tel noyau de transition.

1.3 Les échantillonneurs de type Metropolis-Hastings

La première version de l'algorithme de Metropolis-Hastings a été proposée dans [Met+53], et généralisée dans [Has70] et [Tie98]. Considérons un noyau de transition P sur $(\mathsf{E}, \mathcal{E})$ sous la forme: pour tout $x \in \mathsf{E}$ et $\mathtt{A} \in \mathcal{E}$,

$$P(x, \mathtt{A}) = \int_{\mathtt{A}} \alpha(x,y) Q(x, \mathrm{d}y) + \delta_x(\mathtt{A}) \int_{\mathsf{E}} (1 - \alpha(x,y)) Q(x, \mathrm{d}y) \;, \tag{1.4}$$

où Q est un noyau sur $(\mathsf{E}, \mathcal{E})$, que l'on appelera noyau de proposition, et $\alpha : \mathsf{E} \times \mathsf{E} \to [0,1]$ une fonction mesurable, que l'on appelera ratio d'acceptation. On peut facilement simuler une chaîne de Markov de noyau P si c'est le cas pour le noyau Q. Soit $x \in \mathsf{E}$ et W de loi $Q(x, \cdot)$. Définissons la variable aléatoire Y par

$$Y = \begin{cases} W & \text{avec probabilité } \alpha(x, W) \\ x & \text{sinon} \;. \end{cases}$$

Alors, on déduit aisément que Y a pour loi $P(x, \cdot)$. On peut observer que si P est sous la forme (1.4) alors par définition il est réversible par rapport à π si et seulement si pour toute fonction mesurable bornée $f : \mathsf{E}^2 \to \mathbb{R}$,

$$\int_{\mathsf{E}^2} f(x,y) \pi(x) P(\mathrm{d}x, \mathrm{d}y) = \int_{\mathsf{E}^2} f(x,y) \pi(\mathrm{d}y) P(y, \mathrm{d}x) \;,$$

ce qui est équivalent à

$$\int_{\mathsf{E}^2} f(x,y) \alpha(x,y) \pi(\mathrm{d}x) Q(x, \mathrm{d}y) = \int_{\mathsf{E}^2} f(x,y) \alpha(y,x) \pi(\mathrm{d}y) Q(y, \mathrm{d}x) \;.$$

Définissons les deux mesures sur $(\mathsf{E}^2, \mathcal{E}^{\otimes 2})$, μ et μ^{T} pour tout $\mathtt{A} \in \mathcal{E}^{\otimes 2}$ par[2]:

$$\mu(\mathtt{A}) = \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\mathtt{A}}(y,z) \pi(\mathrm{d}y) Q(y, \mathrm{d}z)$$
$$\mu^{\mathrm{T}}(\mathtt{A}) = \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\mathtt{A}}(z,y) \pi(\mathrm{d}y) Q(y, \mathrm{d}z) \;.$$

[1] voir Definition A.14

[2] voir Definition-Proposition A.1 pour une définition rigoureuse

Alors P est réversible par rapport à π si il existe un ensemble mesurable $\mathsf{S} \in \mathcal{E}^{\otimes 2}$, qui soit symétrique, tel que α est nul sur S^c, $\mu(\cdot \cap \mathsf{S})$ soit absolument continue par rapport à $\mu^{\mathrm{T}}(\cdot \cap \mathsf{S})$ de densité m et μ-presque partout

$$\alpha(x,y)m(x,y) = \alpha(y,x) \ . \tag{1.5}$$

On dira que le noyau P est de type Metropolis-Hastings si il est sous la forme (1.4) et vérifie (1.5). Deux exemples importants que l'on traitera dans ce manuscrit sont les suivants.

(A) Supposons qu'il existe un mesure ν sur $(\mathsf{E}, \mathcal{E})$ qui domine π et Q, i.e. π admet une densité encore notée π par rapport à ν et Q admet une densité de transition q par rapport à ν. Alors si on définit

$$\mathsf{S} = \left\{ (x,y) \in \mathsf{E}^2 \,\middle|\, \pi(x)q(x,y) > 0 \ , \pi(y)q(y,x) > 0 \right\} \ ,$$

et

$$\alpha(x,y) = \begin{cases} \min\left(1, \dfrac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) & \text{pour } (x,y) \in \mathsf{S} \\ 0 & \text{sinon} \ , \end{cases} \tag{1.6}$$

les conditions pour que P soit réversible sont satisfaites avec

$$m(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) \ ,$$

défini sur S.

Dans ce cadre un choix simple pour le noyau de proposition Q est la mesure de domination ν si celle-ci est une mesure de probabilité. Alors les variables aléatoires proposées au cours de l'algorithme constituent une suite i.i.d. de loi commune ν. Observons que dans ce cas pour tout $(x,y) \in \mathsf{S}$,

$$\alpha(x,y) = \min\left(1, \pi(y)/\pi(x)\right) \ . \tag{1.7}$$

Cet algorithme est appelé l'algorithme de Metropolis-Hastings indépendant. Une autre possibilité pour définir un noyau de Metropolis-Hastings, lorsque la structure de E le permet et que ν est invariante par translation, est de considérer une marche aléatoire associée à une mesure symétrique sur $(\mathsf{E}, \mathcal{E})$ qui admet une densité $\tilde{q}$ par rapport à ν. Aussi, les variables aléatoires proposées au cours de l'algorithme sont de la forme

$$W_{k+1} = Y_k + Z_{k+1} \ ,$$

où Y_k est l'état courant à l'itération k et $(Z_i)_{i \geq 1}$ est une suite i.i.d. de variables aléatoires de densité $\tilde{q}$. On peut observer que dans ce cas le noyau de transition est donné pour tout $x, y \in \mathsf{E}$ par $q(x,y) = \tilde{q}(y-x)$ et comme $\tilde{q}$ est supposé symétrique α est encore une fois sous la forme (1.7). Cette méthode est appelée l'algorithme de Metropolis à marche aléatoire symétrique. Un exemple de ce type de méthode est l'algorithme de Metropolis à marche aléatoire symétrique sur $\mathbb{R}^d$ muni de la mesure de Lebesgue et où

$$q(x,y) = (2\pi\varsigma^2)^{-d/2} \exp\left(- \|y-x\|^2 / (2\varsigma^2)\right) \ .$$

(B) Si π admet une densité positive, encore notée π, par rapport à une mesure ν, et le noyau Q est réversible par rapport à ν, alors en définissant $\mathsf{S} = \mathsf{E}$ et pour tout $(x, y) \in \mathsf{E}^2$,

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right) ,$$

(1.4) définit toujours un noyau de Metropolis-Hastings réversible par rapport à π. Notons que dans ce cadre contrairement à (A), il n'est pas nécessaire que Q soit dominé par rapport à ν.

1.4 La diffusion de Langevin

1.4.1 Dynamique markovienne continue

Au lieu de considérer des dynamiques discrètes, on peut aussi penser à utiliser des processus markoviens en temps continu associés à un semi-groupe $(\mathbf{P}_t)_{t\geq 0}$ sur $(\mathsf{E}, \mathcal{E})$, pour lequel π est invariante. Pour cela, on suppose que E est un espace polonais localement compact et que $(\mathbf{P}_t)_{t\geq 0}$ est fellerien [3]. On associe à $(\mathbf{P}_t)_{t\geq 0}$, un opérateur appelé son générateur et noté $\mathscr{A}$. Notons $C_0(\mathsf{E})$ l'ensemble des fonctions continues nulles à l'infini[4]. L'ensemble de définition $\mathscr{D}(\mathscr{A})$ de $\mathscr{A}$ est l'ensemble des fonctions $h \in C_0(\mathsf{E})$ pour lesquelles il existe une fonction $g_h \in C_0(\mathsf{E})$ telle que pour tout $x \in \mathsf{E}$,

$$g_h(x) = \lim_{t\to 0} t^{-1} \left\{\mathbf{P}_t h(x) - h(x)\right\} .$$

Alors on définit pour tout $h \in \mathscr{D}(\mathscr{A})$, $\mathscr{A}h = g_h$. L'étude du générateur associé à un semi-groupe markovien permet de déduire de nombreuses informations sur celui-ci. Soit $\mathbb{A}$ une algèbre incluse dans $\mathscr{D}(\mathscr{A})$ et dense dans $C_0(\mathsf{E})$. Notamment si $\mathsf{E} = \mathbb{R}^d$, alors on peut considérer pour $\mathbb{A}$, l'ensemble des fonctions k-fois différentiable et à support compact, noté $C_c^k(\mathbb{R}^d)$, pour $k \in \mathbb{N} \cup \{\infty\}$. Par [RY99, Proposition 1.5, chapitre VII] et [EK86, Théorème 9.17, chapitre 3], si pour tout $h \in \mathbb{A}$, $\int_{\mathsf{E}} \mathscr{A}h(x)\mathrm{d}\pi(x) = 0$, alors π est invariante pour $(\mathbf{P}_t)_{t\geq 0}$. Si π est une distribution invariante pour $(\mathbf{P}_t)_{t\geq 0}$ et que $(\mathbf{P}_t)_{t\geq 0}$ est Harris récurrent[5] alors on a toujours une loi forte des grands nombres [RY99, Théorème 3.12, chapitre X] : pour tout $x \in \mathsf{E}$ et toute fonction $f \in \mathrm{L}^1(\pi)$

$$\lim_{T\to+\infty} T^{-1} \int_0^T f(\mathbf{Y}_s)\mathrm{d}s = \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x) , \qquad \mathbb{P}_x\text{-presque sûrement} ,$$

où $(\mathbf{Y}_t)_{t\geq 0}$ est le processus markovien canonique associé à $(\mathbf{P}_t)_{t\geq 0}$ et $\mathbb{P}_x$ l'unique probabilité induite par le semi-groupe et la distribution initiale δ_x[6]. Nous nous intéresserons dans la suite aux semi-groupes de Markov associés aux équations différentielles stochastiques (EDS) homogènes sur $\mathbb{R}^d$.

[3] voir Definition A.25
[4] voir Definition A.24
[5] voir Definition A.30
[6] voir Theorem A.26

Soit $b : \mathbb{R}^d \to \mathbb{R}$ et σ une fonction de $\mathbb{R}^d$ dans $M_{d,m}(\mathbb{R})$, l'ensemble des matrices de dimension $d \times m$, telle que pour tout $x \in \mathbb{R}^d$, $\sigma(x)\sigma(x)^{\mathrm{T}}$ est définie positive. On suppose que ces deux fonctions sont localement Lipschitziennes et on considère l'EDS associée:

$$\mathrm{d}\mathbf{Y}_t = b(\mathbf{Y}_t)\mathrm{d}t + \sigma(\mathbf{Y}_t)\mathrm{d}B_t^m ,$$

où $(B_t^m)_{t\geq 0}$ est un mouvement brownien m-dimensionnel. Par [IW89, Théorème 2.3, Théorème 3.1, chapitre 4], pour toute condition initiale $\mathbf{Y}_0 = x \in \mathbb{R}^d$, cette EDS admet une unique solution $(\mathbf{Y}_t)_{t\in 0}$ sur un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ muni d'un mouvement brownien $(B_t^m)_{t\geq 0}$, jusqu'à un temps d'explosion ξ qui est le temps d'arrêt défini par

$$\xi = \inf\{t \geq 0 \,|\, \mathbf{Y}_t = \infty\} .$$

On suppose que presque sûrement $\xi = +\infty$. Alors par [SV79, Corollaire 10.1.5] et [IW89, Théorème 6.1, chapitre 4], la loi de $(\mathbf{Y}_t)_{t\in 0}$ définit un semi-groupe de Markov fellerien $(\mathbf{P}_t)_{t\geq 0}$ donné pour tout $\mathtt{A} \in \mathcal{B}(\mathbb{R}^d)$ et $x \in \mathbb{R}^d$ par $\mathbf{P}_t(x, \mathtt{A}) = \mathbb{P}_x[\mathbf{Y}_t \in \mathtt{A}]$. De plus [Bha78, Lemme 2.4] montre que ce semi-groupe est irréductible[7] pour la mesure de Lebesgue. Enfin [Bha78, Théorème 3.3] donne une critère pour que la diffusion soit Harris récurrente.

Dans le cas où $\mathsf{E} = \mathbb{R}^d$ et π admet une densité positive par rapport à la mesure de Lebesgue de la forme

$$\pi(x) = \mathrm{e}^{-U(x)} \bigg/ \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)}\mathrm{d}x \quad , \tag{1.8}$$

où $U : \mathbb{R}^d \to \mathbb{R}$ est une fonction continûment différentiable, la diffusion de Langevin définie par

$$\mathrm{d}\mathbf{Y}_t^{\mathrm{L}} = -\nabla U(\mathbf{Y}_t^{\mathrm{L}}) + \sqrt{2}\mathrm{d}B_t^d , \tag{1.9}$$

admet π comme distribution invariante. Son générateur $\mathscr{A}^{\mathrm{L}}$ est donné pour toute fonction $h \in C_c^2(\mathbb{R}^d)$ et $x \in \mathbb{R}^d$ par

$$\mathscr{A}^{\mathrm{L}}h(x) = -\langle \nabla U(x), \nabla h(x)\rangle + \Delta h(x) .$$

Pour que le processus soit non explosif, on suppose que pour tout $x \in \mathbb{R}^d$ [MT93b, Théorème 2.1],

$$\langle \nabla U(x), x\rangle \geq -a_1 \|x\|^2 - a_2 ,$$

pour des constantes $a_1, a_2 \in \mathbb{R}_+$. Par un simple changement de variable, on montre que pour toute fonction $h \in C_c^2(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \left\{\mathscr{A}^{\mathrm{L}}h\right\}(x)\pi(x)\mathrm{d}x = 0 ,$$

ce qui montre que π est invariante pour le semi-groupe de Markov $(\mathbf{P}_t^{\mathrm{L}})_{t\geq 0}$ associé à (1.9). Donc par [MT09, Proposition 10.1.1] et Corollaire A.34, $(\mathbf{P}_t^{\mathrm{L}})_{t\geq 0}$ est Harris récurrent.

À de rares exceptions, il n'existe pas de façon simple de simuler des trajectoires solutions de l'EDS. Bien que des simulations exactes aient été proposées (voir [**beskos:roberts:2005**]), leur mise en oeuvre en temps long est très coûteuse.

[7]voir Definition A.28

1.4.2 Deux algorithmes MCMC basés sur l'equation de Langevin: l'ULA et le MALA

On considère dans cette thèse la discrétisation de Euler-Maruyama associée à (1.9) et définie pour une condition initiale donnée par : pour tout $k \geq 0$,

$$Y_{k+1} = Y_k - \gamma \nabla U(Y_k) + \sqrt{2\gamma} Z_{k+1} \ , \tag{1.10}$$

où $\gamma > 0$ est le pas de discrétisation et $(Z_k)_{k \geq 1}$ est une suite i.i.d. de variables aléatoires gaussiennes centrées réduites d dimensionnelle. La discrétisation $(Y_n)_{n \in \mathbb{N}}$ peut alors être considérée comme une trajectoire approchée de $(\mathbf{Y}_t)_{t \geq 0}$ et utilisée pour échantillonner suivant π. Cet algorithme a été tout d'abord proposé par [Erm75] et [Par81] pour des applications en dynamique moléculaire. Il a été ensuite rendu populaire dans la communauté de l'apprentissage statistique par [Gre83], [GM94] et en statistiques computationnelles par [Nea93] et [RT96a]. Comme dans [RT96a], cet algorithme sera appelé dans ce manuscrit l'algorithme de Langevin non-ajusté (ULA pour Unadjusted Langevin Algorithm).

Le désavantage de cette méthode est que même si $(Y_n)_{n \in \mathbb{N}}$ possède une unique distribution invariante π_γ et est ergodique (ce qui est garanti sous des conditions très faibles sur U), cette mesure est dans la grande majorité des cas différente de π. Utiliser une telle méthode introduit alors un biais dans le calcul de $\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x)$, i.e.

$$\lim_{N \to +\infty} N^{-1} \sum_{k=0}^{N-1} f(Y_k) = \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi_\gamma(x) \neq \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x) \ , \qquad \mathbb{P}_x\text{-presque sûrement} \ .$$

Néanmoins [TT90] ont montré que sous de bonnes hypothèses sur f (régularité et croissance polynomiale), la chaîne $(Y_n)_{n \in \mathbb{N}}$ et sur la diffusion $(\mathbf{Y}_t)_{t \geq 0}$, il existe une constante C qui dépend de f et π telle que pour tout $\gamma > 0$ dans un voisinage de 0

$$\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi_\gamma(x) - \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x) = C\gamma + \mathcal{O}(\gamma^2) \ .$$

Ainsi prendre un pas de discrétisation suffisamment petit permet de réduire l'erreur dans le calcul de $\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x)$ et justifie l'intérêt de cette méthode. Pour supprimer ce biais, il a été proposé dans [RDF78] et [RT96a] d'utiliser le noyau de transition associé à la discrétisation d'Euler (1.10) comme noyau de proposition dans un algorithme de Metropolis Hastings. Suivant [RT96a], nous appellerons cet algorithme l'algorithme de Langevin-Metropolis ajusté (MALA pour Metropolis Adjusted Langevin Algorithm).

Une autre méthode pour supprimer le biais de ULA est d'utiliser des pas décroissants $(\gamma_k)_{k \geq 1}$ qui vérifient $\lim_{k \to +\infty} \gamma_k = 0$ et $\sum_{k=1}^{+\infty} \gamma_k = +\infty$. On définit la chaîne de Markov inhomogène $(Y_n)_{n \in \mathbb{N}}$ associée, avec une condition initiale donnée, pour tout $k \geq 0$ par

$$Y_{k+1} = Y_k - \gamma_{k+1} \nabla U(Y_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} \ ,$$

où $(Z_k)_{k\geq 1}$ est une suite i.i.d. de variables aléatoires gaussiennes centrées réduites d dimensionnelle. Il a été montré dans [LP02, Théorème 6] que sous de bonnes hypothèses sur f et π, pour tout $x \in \mathbb{R}^d$,

$$\lim_{N\to+\infty} \frac{\sum_{k=0}^{N-1} \gamma_{k+1} f(Y_k)}{\sum_{k=1}^{N} \gamma_k} = \int_{\mathbb{R}^d} f(x)\mathrm{d}\pi(x) , \qquad \mathbb{P}_x\text{-presque sûrement} .$$

Dans la suite, cet algorithme sera toujours appelé l'algorithme ULA.

1.5 Convergence des méthodes MCMC

Soit $(Y_k)_{k\in\mathbb{N}}$ une chaîne de Markov homogène, de distribution initiale μ_0 et de noyau P, sur un espace polonais E muni d'une distance $\mathbf{d}$ et de sa tribu borélienne toujours notée $\mathcal{E}$. Définissons la suite des lois marginales $(\mu_0 P^k)_{k\in\mathbb{N}^*}$ de la chaîne $(Y_k)_{k\in\mathbb{N}}$ par récurrence pour tout $k \in \mathbb{N}^*$ et $\mathtt{A} \in \mathsf{E}$:

$$\mu_0 P^k(\mathtt{A}) = \int_{\mathsf{E}} \mathbb{1}_{\mathtt{A}}(y)\mu_0 P^{k-1}(\mathrm{d}x)P(x, \mathrm{d}y) .$$

Nous nous intéressons dans cette section à l'existence et surtout l'unicité d'une mesure de probabilité invariante pour P. Nous donnons de plus des résultats de convergence de la suite de mesures de probabilité $(\mu_0 P^k)_{k\in\mathbb{N}^*}$ vers l'unique mesure de probabilité invariante de P lorsqu'elle existe.

Nous avons vu dans les Section 1.3 et Section 1.4.2 des méthodes permettant de construire des chaînes de Markov ayant une probabilité donnée π comme probabilité invariante. L'étude de la convergence des lois marginales est motivée par l'analyse de la convergence de l'estimateur $\hat{f}_N$, défini par (1.3) vers $\int_{\mathbb{R}^d} f(x)\mathrm{d}\pi(x)$ quand N tend vers l'infini. En effet, comme $(Y_k)_{k\in\mathbb{N}}$ n'est pas un échantillon i.i.d. de loi π, une première étape est de mesurer le biais de l'estimation défini par

$$\left|\mathbb{E}_x\left[\hat{f}_N\right] - \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x)\right| = \left|\frac{1}{N+1}\sum_{i=0}^{N}\left\{\mathbb{E}_x\left[f(Y_i)\right] - \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x)\right\}\right| .$$

La convergence des lois marginales $(\mu_0 P^k)_{k\in\mathbb{N}}$ de la chaîne a été l'objet de nombreuses études [MT09], [Num84], [HMS11]. Cette convergence est établie pour différentes distances sur l'espace des mesures de probabilité sur $(\mathsf{E}, \mathcal{E})$, notée $\mathcal{P}(\mathsf{E})$. Nous considérerons deux sortes de distances dans ce manuscrit, les distances en V-variation totale et les distances de Wasserstein.

1.5.1 Convergence des chaînes de Markov

Métriques sur l'espace des mesures de probabilité

Soit $V : \mathsf{E} \to [1, \infty)$ une fonction mesurable. Pour $h : \mathsf{E} \to \mathbb{R}$ une fonction mesurable, on définit la V-norme de h par

$$\|h\|_V = \sup_{x\in\mathsf{E}} |h(x)|/V(x) .$$

Soit μ une mesure signée bornée sur $(\mathsf{E}, \mathcal{E})$. On définit la V-variation totale de μ par

$$\|\mu\|_V = (1/2) \sup_{\|h\|_V \leq 1} \left| \int_{\mathsf{E}} h(x) \mathrm{d}\mu(x) \right| .$$

Si $V \equiv 1$, alors $\|\cdot\|_V$ est la variation totale notée $\|\cdot\|_{\mathrm{TV}}$. Pour deux mesures de probabilité $\mu, \nu \in \mathcal{P}(\mathsf{E})$, la distance en V-variation totale entre μ et ν est la V-variation totale de la mesure $\mu - \nu$. De même la distance en variation totale entre μ et ν est la variation totale de la mesure $\mu - \nu$. L'ensemble des mesures de probabilités $\{\mu \in \mathcal{P}(\mathsf{E}) \,|\, V \in \mathrm{L}^1(\mu)\}$ est un espace de Banach lorsqu'il est muni de la distance en V-variation totale, voir [DMS14, Proposition 6.16].

Une autre distance que l'on considère est la distance de Wasserstein associée à la distance $\mathbf{d}$ sur E. Soit $h : \mathsf{E} \to \mathbb{R}$ une fonction lipschitzienne i.e. il existe $C \geq 0$ tel que pour tout $x, y \in \mathsf{E}$, $|h(x) - h(y)| \leq C\mathbf{d}(x, y)$. Pour $h : \mathsf{E} \to \mathbb{R}$ lipschitzienne, on définit

$$\|h\|_{\mathrm{Lip}} = \sup_{x,y \in \mathsf{E}} \left\{ \frac{|h(x) - h(y)|}{\mathbf{d}(x, y)} \right\}$$

Définissons l'ensemble des mesures de probabilité $\mathcal{P}_1(\mathsf{E})$ par

$$\mathcal{P}_1(\mathsf{E}) = \{\mu \in \mathcal{P}(\mathsf{E}) \,|\, \int_{\mathsf{E}} \mathbf{d}(x, x_0) \mathrm{d}\mu(x) < +\infty\} ,$$

pour un élément $x_0 \in \mathsf{E}$ fixé. La distance de Wasserstein est définie par pour tout $\mu, \nu \in \mathcal{P}_1(\mathsf{E})$,

$$W_{\mathbf{d}}(\mu, \nu) = \sup_{\|h\|_{\mathrm{Lip}} \leq 1} \left| \int_{\mathsf{E}} h(x) \mathrm{d}\mu(x) - \int_{\mathsf{E}} h(x) \mathrm{d}\nu(x) \right| .$$

D'après [Vil09, Theorems 6.8 and 6.16], $\mathcal{P}_1(\mathsf{E})$ muni de $W_{\mathbf{d}}$ est un espace polonais.

La distance de Wasserstein et la distance en variation totale ne sont la plupart du temps pas comparables, sauf lorsque $\mathbf{d}$ est bornée et dans ce cas pour tout $\mu, \nu \in \mathcal{P}(\mathsf{E})$,

$$W_{\mathbf{d}}(\mu, \nu) \leq \sup_{x,y \in \mathsf{E}^2} \{\mathbf{d}(x, y)\} \|\mu - \nu\|_{\mathrm{TV}} .$$

Il est aisé de voir que la convergence en l'une de ces deux distances implique la convergence faible.

D'après le théorème de Monge-Kantorovich [Vil09, Theorem 5.10], la distance de Wasserstein et la distance en variation totale entre deux mesures de probabilité μ et ν sur E, possèdent des formes duales qui font intervenir l'ensemble des couplages entre μ et ν. Une mesure de probabilité ζ sur $(\mathsf{E} \times \mathsf{E}, \mathcal{E} \otimes \mathcal{E})$ est un plan de transport entre μ et ν si sa première marginale $\zeta(\cdot \times \mathsf{E})$ est égale à μ et sa seconde $\zeta(\mathsf{E} \times \cdot)$ est égale à ν. L'ensemble des plans de transport entre μ et ν sera noté $\Pi(\mu, \nu)$. Il est d'usage d'appeler couplage de μ et ν tout couple de variables aléatoires (X, Y) de loi $\zeta \in \Pi(\mu, \nu)$. L'ensemble des couplages de μ et ν sera noté $\tilde{\Pi}(\mu, \nu)$. La distance en variation totale peut s'écrire sous la forme : pour tout $\mu, \nu \in \mathcal{P}(\mathsf{E})$

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf_{\zeta \in \Pi(\mu,\nu)} \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\Delta_{\mathsf{E}}}(x, y) \zeta(\mathrm{d}x, \mathrm{d}y) = \inf_{(\mathrm{X},\mathrm{Y}) \in \tilde{\Pi}(\mu,\nu)} \mathbb{P}(\mathrm{X} \neq \mathrm{Y}) ,$$

où $\Delta_{\mathsf{E}} = \{(x, y) \in \mathsf{E} \times \mathsf{E} \,|\, x = y\}$. De même la distance de Wasserstein s'écrit sous la forme : pour tout $\mu, \nu \in \mathcal{P}(\mathsf{E})$

$$W_{\mathbf{d}}(\mu, \nu) = \inf_{\zeta \in \Pi(\mu,\nu)} \int_{\mathsf{E}\times\mathsf{E}} \mathbf{d}(x, y)\zeta(\mathrm{d}x, \mathrm{d}y) = \inf_{(\mathrm{X},\mathrm{Y})\in\tilde{\Pi}(\mu,\nu)} \mathbb{E}[\mathbf{d}(\mathrm{X}, \mathrm{Y})] \ .$$

De plus, l'infimum est atteint pour ces deux distances.

Convergence des chaînes en variation totale et V-variation totale

L'analyse des chaînes est grandement facilitée lorsque l'on suppose que la chaîne possède un atome accessible. Un ensemble mesurable $\boldsymbol{\alpha} \in \mathcal{E}$ est un atome pour le noyau P sur $(\mathsf{E}, \mathcal{E})$ si il existe une mesure de probabilité ν sur $(\mathsf{E}, \mathcal{E})$ telle que pour tout $x \in \boldsymbol{\alpha}$, $P(x, \cdot) = \nu$. En effet, dans ce cas l'existence d'une mesure invariante et l'analyse de la convergence en V-variation totale se ramène à l'étude du temps de retour à cet atome [MT09, section 10.2, 13.2, 14.1, 15.1].

Soit $(Y_k)_{k\in\mathbb{N}}$ et $(\mathcal{F}_k)_{k\in\mathbb{N}}$ la chaîne et la filtration canonique associées à P. Pour tout $\mathtt{A} \in \mathcal{E}$, on définit le temps de retour à $\mathtt{A}$ par:

$$\boldsymbol{\sigma}_{\mathtt{A}} = \inf \{ k \in \mathbb{N}^* \,|\, Y_k \in \mathtt{A} \} \ .$$

On définit ensuite les temps de retour successifs $(\boldsymbol{\sigma}_{\mathtt{A}}^{(m)})_{m\in\mathbb{N}^*}$ de P à $\mathtt{A}$ par récurrence. Pour $m = 1$, $\boldsymbol{\sigma}_{\mathtt{A}}^{(m)} = \boldsymbol{\sigma}_{\mathtt{A}}$ et pour $m \geq 2$, on pose

$$\boldsymbol{\sigma}_{\mathtt{A}}^{(m)} = \inf \{ k \in \mathbb{N}^* \,|\, Y_{\boldsymbol{\sigma}_{\mathtt{A}}+k} \in \mathtt{A} \} \ .$$

Observons que pour tout $\mathtt{A} \in \mathcal{E}$ et $m \in \mathbb{N}^*$, $\boldsymbol{\sigma}_{\mathtt{A}}^{(m)}$ est un $(\mathcal{F}_k)_{k\in\mathbb{N}}$-temps d'arrêt.

Sauf dans le cas où l'espace d'état est discret. L'existence d'un atome est une condition très forte qui n'est que très rarement vérifiée. Une condition moins forte est l'existence d'un ensemble small pour P. Soit $\mathtt{n} \in \mathbb{N}^*$. L'ensemble $\mathtt{C} \in \mathcal{E}$ est un ensemble small ou $\mathtt{n}$-small pour P si il existe une mesure non triviale σ-finie ν sur $(\mathsf{E}, \mathcal{E})$ telle que pour tout $x \in \mathtt{C}$, $P^{\mathtt{n}}(x, \cdot) \geq \nu(\cdot)$. Lorsque P admet un ensemble 1-small $\mathtt{C}$, alors la technique de scission (splitting) due à [Num78] permet de construire un noyau de transition $\check{P}$ sur l'espace étendu $(\mathsf{E} \times \{0, 1\}, \mathcal{E} \otimes \mathcal{B}(\{0, 1\}))$, qui admet $\mathtt{C} \times \{1\}$ comme atome et tel que pour tout $x \in \mathsf{E}$, $\mathsf{b} \in \{0, 1\}$ et $\mathtt{A} \in \mathcal{E}$,

$$\check{P}((x, \mathsf{b}), \mathtt{A} \times \{0, 1\}) = P(x, \mathtt{A}) \ .$$

Aussi le temps de retour à l'atome $\mathtt{C} \times \{1\}$ de $\check{P}$ est intimement lié aux temps de retour successifs de P à $\mathtt{C}$ et l'analyse du noyau P se fait à travers le processus de renouvellement défini par ces temps d'arrêts. Si $\mathtt{C}$ est un ensemble petite[8], ce dernier est tout de même un ensemble 1-small pour un noyau échantillonné[9] associé à P et la technique de scission s'applique au noyau échantillonné. Ainsi moralement, l'étude des chaînes irréductibles

[8] cf Definition A.8
[9] cf Definition A.5

qui ont un ensemble petite, se ramène à l'étude des temps de retour à cet ensemble et à utiliser les résultats pour les chaînes possédant un atome.

Des contrôles sur les moments des temps de retour à un ensemble $\mathsf{C} \in \mathcal{E}$ peuvent être obtenus à partir de conditions de dérive (drift) de Foster-Lyapunov.

(1) La première condition de dérive que nous présentons est la suivante : il existe une fonction mesurable propre $V : \mathsf{E} \to \bar{\mathbb{R}}_+$ et $a \in \mathbb{R}_+$ tels que

$$PV \leq V - 1 + a\,\mathbb{1}_{\mathsf{C}} \;, \qquad \sup_{x \in \mathsf{C}} V(x) < +\infty \;. \tag{1.11}$$

D'après le théorème de Dynkin[10], pour tout $x \in \mathsf{E}$,

$$\mathbb{E}_x\left[\boldsymbol{\sigma}_{\mathsf{C}}\right] \leq V(x) + a\mathbb{1}_{\mathsf{C}}(x) \;, \qquad \sup_{x \in \mathsf{C}} \mathbb{E}_x\left[\boldsymbol{\sigma}_{\mathsf{C}}\right] < +\infty \;. \tag{1.12}$$

[MT09, p. 13.0.1] montre que si C est petite, P est irréductible, apériodique et Harris récurrente, alors (1.12) est équivalent à l'existence d'une unique mesure de probabilité invariante π pour P et pour tout $x \in \mathbb{R}^d$,

$$\lim_{n \to +\infty} \|\delta_x P^n - \pi\|_{\mathrm{TV}} = 0 \;.$$

Si P n'est pas Harris récurrent mais seulement irréductible, apériodique et récurrent, et (1.12) est satisfait, alors d'après [MT09, Theorem 9.1.5][11], il existe une unique mesure de probabilité invariante π pour P et un ensemble $\mathsf{N} \in \mathcal{E}$ tels que $\pi(\mathsf{N}) = 0$ et pour tout $x \in \mathsf{N}^{\mathrm{c}}$,

$$\lim_{n \to +\infty} \|\delta_x P^n - \pi\|_{\mathrm{TV}} = 0 \;.$$

(2) La seconde inégalité de dérive que nous présentons permet d'avoir une convergence en des distances plus “fortes” que la distance en variation totale. Supposons qu'il existe une fonction mesurable propre $V : \mathsf{E} \to \bar{\mathbb{R}}_+$, une fonction mesurable $\mathrm{f} : \mathsf{E} \to [1, +\infty)$ et $a \in \mathbb{R}_+$ tels que

$$PV \leq V - \mathrm{f} + a\,\mathbb{1}_{\mathsf{C}} \;, \qquad \sup_{x \in \mathsf{C}} V(x) < +\infty \;.$$

Toujours d'après le théorème de Dynkin, nous avons

$$\mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \mathrm{f}(Y_k)\right] \leq V(x) + a\mathbb{1}_{\mathsf{C}}(x) \;, \qquad \sup_{x \in \mathsf{C}} \mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \mathrm{f}(Y_k)\right] < +\infty \;. \tag{1.13}$$

[MT09, Theorem 14.0.1] montre que si P est irréductible apériodique et C est un ensemble petite, (1.13) implique que P admet une unique distribution invariante π et pour tout $x \in \{V < +\infty\}$,

$$\lim_{n \to +\infty} \|\delta_x P^n - \pi\|_{\mathrm{f}} = 0 \;.$$

[10] cf Corollary A.20

[11] cf Theorem A.15

(3) Nous présentons finalement une condition impliquant un taux de convergence de $(\delta_x P^n)_{n\in\mathbb{N}}$ vers π en V-variation totale. On suppose qu'il existe une fonction mesurable propre $V : \mathsf{E} \to [1,+\infty]$, $\lambda \in [0,1)$ et $a \in \mathbb{R}_+$ tels que

$$PV \leq \lambda V + a\,\mathbb{1}_{\mathsf{C}}\ , \qquad \sup_{x\in\mathsf{C}} V(x) < +\infty\ . \tag{1.14}$$

Le théorème de Dynkin implique que pour tout $\kappa \in (1, \lambda^{-1})$, il existe $C \geq 0$ tel que

$$\mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \kappa^k V(Y_k)\right] \leq C(V(x) + a\mathbb{1}_{\mathsf{C}}(x))\ , \qquad \sup_{x\in\mathsf{C}} \mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \kappa^k V(Y_k)\right] < +\infty\ . \tag{1.15}$$

[MT09, Theorem 15.4.1] montre que si P est irréductible, apériodique et (1.15) est satisfait avec un ensemble petite C, alors il existe $r > 1$ et $C \geq 0$ tels que pour tout $x \in \{V < +\infty\}$, et $n \geq 1$,

$$\|\delta_x P^n - \pi\|_V \leq CV(x) r^n\ .$$

Finalement nous mentionnons aussi qu'il existe deux types de condition de dérive proposés respectivement dans [TT94] et [Dou+04], pour avoir des convergences sous-géométriques en V-variation totale d'une chaîne de Markov irréductible et apériodique. Moralement ces deux conditions de dérive entraînent des moments sous géométriques pour les temps de retour à un ensemble, ce qui entraîne la convergence sous géométrique dans le cas où cet ensemble est un petite set.

1.5.2 Application à la convergence des méthodes MCMC

De nombreux travaux portent sur l'obtention des taux de convergence pour des noyaux de type Metropolis-Hastings sur $(\mathsf{E}, \mathcal{E})$, quand π admet une densité par rapport à une mesure de domination ν et le noyau de proposition, une densité de transition q par rapport aussi à ν, cf Section 1.3-(A).

Notons P_{MH} un noyau de Metropolis-Hastings vérifiant ces conditions et défini par (1.4)-(1.6). [Tie94, Corollary 2] montre que si P_{MH} est π-irréductible, alors il est Harris récurrent. Pour vérifier que P_{MH} est π-irréductible, une condition très simple est que pour tout $x \in \mathsf{E}$, $\pi(x) > 0$ implique que $q(y,x) > 0$ pour tout $y \in \mathsf{E}$, [MT96, Lemma 1.1]. Dans le cas où $\mathsf{E} = \mathbb{R}^d$ pour $d \geq 1$, et ν est la mesure de Lebesgue, cette condition est affaiblie par [RT96b, Theorem 2.2] qui établit que si π est positive et bornée sur $\mathbb{R}^d$, et il existe $\delta_q, \epsilon_q > 0$ tels que

$$q(x,y) \geq \epsilon_q \text{ pour tout } x,y \in \mathbb{R}^d,\ \|x-y\| \leq \delta_q\ , \tag{1.16}$$

alors P est irréductible par rapport à la mesure de Lebesgue et donc π-irréductible. De plus le théorème indique aussi que ces conditions entraînent que P est fortement apériodique et tout ensemble compact non vide est small. Notons que ce résultat est immédiat si π et q sont continus car P_{MH} est Feller et donc tout ensemble compact non vide est petite si il est irréductible[12]. Ainsi l'analyse des taux de convergence des algorithmes de Metropolis Hastings dans le cas dominé se divise en deux catégories.

Si le noyau satisfait une condition de minoration uniforme, alors il est uniformément ergodique. Par exemple, dans le cas de l'algorithme de Metropolis indépendant (voir Section 1.3-(A)), [MT96, Theorem 2.1] montre que si il existe $\beta_q > 0$ tel que pour tout $x \in \mathsf{E}$, $q(x)/\pi(x) > \beta_q$, alors P_{MH} est uniformément ergodique : pour tout $n \geq 1$,

$$\sup_{x \in \mathsf{E}} \|\delta_x P_{\mathrm{MH}} - \pi\|_{\mathrm{TV}} \leq (1-\beta_q)^n \; .$$

Si le noyau n'est pas uniformément ergodique, alors l'approche classique consiste à établir une condition de dérive géométrique ou sous-géométrique. C'est par exemple le cas pour l'analyse de l'algorithme de Metropolis avec marche aléatoire symétrique sur $\mathbb{R}^d$, et lorsque π admet une densité par rapport à la mesure de Lebesgue, qui soit positive et continue. Dans ce cadre, diverses conditions sur la géométrie des lignes de niveaux de π ont été proposées dans [RT96b] et [JH00] qui impliquent une condition de dérive géométrique de la forme (1.14). En particulier, ces conditions sont satisfaites pour des densités de la forme: il existe $\ell \in \mathbb{N}^*$, un polynôme positive homogène $\mathrm{p} : \mathbb{R}^d \to \mathbb{R}_+^*$ de degré ℓ, un polynôme $\mathrm{q} : \mathbb{R}^d \to \mathbb{R}$ de degré strictement plus petit que ℓ, et un polynôme positif $\mathrm{r} : \mathbb{R}^d \to \mathbb{R}_+^*$ tels que pour tout $x \in \mathbb{R}^d$,

$$\pi(x) \propto \mathrm{r}(x) \exp(-\mathrm{p}(x) - \mathrm{q}(x)) \; . \tag{1.17}$$

Par suite, si la densité de proposition satisfait (1.16), l'algorithme de Metropolis à marche aléatoire symétrique est géométriquement ergodique pour ce type de densité. Pour le cas de distributions avec des queues plus lourdes que celles de la forme (1.17), des conditions de dérive peuvent être établies mais qui n'entraînent que des convergences sous-géométriques de la chaîne P_{MH} vers π, cf [FM00], [FM03] et [JR07].

Pour des applications à des méthodes de Monte Carlo, avoir des bornes explicites de convergence pour les noyaux de de transition associés est important. La dérivation de bornes explicites en variation totale a fait l'objet de nombreux travaux. La plupart des ces résultats s'appuient encore sur une condition de dérive et une condition de minoration par l'existence d'un ensemble small, cf [Ros95], [RT99], ou [Bax05]. Ces bornes sont établies soit en utilisant la technique de scission de Nummelin précédemment introduite, soit par couplages. Cependant, il a été observé dans [JH01] que les bornes obtenues à partir de ces résultats sont difficilement utilisables pour une analyse fine de la convergence.

1.5.3 Analyse de ULA et MALA

[RT96a] a analysé la convergence des chaînes produites par les algorithmes MALA et ULA à pas constant $\gamma > 0$, présentés en Section 1.4.2. Cependant, aucun résultat n'est établi sur la convergence de ULA vers la densité cible π, seulement vers une distribution invariante qui, comme nous l'avons indiqué, est la plupart du temps différente de π.

[12] cf Proposition A.22

Notons par ailleurs que puisque la mesure invariante π_γ de la chaîne de Markov produite par ULA à pas constant diffère de la loi cible π, d'autres méthodes de preuve doivent être combinées à celles présentées en Section 1.5.1.

[RT96a] a établi un résultat de convergence géométrique pour MALA sous la condition que

$$\lim_{\|x\|\to+\infty} \int_{\mathtt{A}(x)} q(x,y)\mathrm{d}y = 0 \ ,$$

où q est la densité de transition associée à MALA, $\mathtt{A}(x)$ est une zone de rejet potentiel restreinte donnée par

$$\mathtt{A} = (\mathtt{B}(x) \cup \mathtt{C}(x)) \setminus (\mathtt{B}(x) \cap \mathtt{C}(x)) \ ,$$

et

$$\mathtt{B}(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, \pi(x)q(x,y) \leq \pi(y)q(y,x) \right\} \ , \quad \mathtt{C}(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, \|y\| \leq \|x\| \right\} \ .$$

Cependant vérifier cette condition s'avère très compliqué en pratique.

Un des objectifs de ce travail de thèse est de fournir des bornes explicites pour la convergence des méthodes de Monte Carlo par chaîne de Markov, MALA et ULA (à pas constants et tendant vers 0), en variation totale et en distance de Wasserstein. Pour des modèles particuliers de distributions cibles, nous avons analysé la dépendance de cette convergence en la dimension de l'espace d'état.

Par ailleurs, les résultats introduits en Section 1.5.1 supposent que la chaîne est irréductible. Or cette condition n'est souvent pas vérifiée en dimension infinie car les mesures de probabilité ont tendance à être singulières. Par exemple si $(B_t)_{t\geq 0}$ est un mouvement brownien pour tout $\sigma^2 > 0$, $\sigma^2 \neq 1$, les lois associées à $(B_t)_{t\geq 0}$ et $(B_t)_{t\geq 0}$ sont singulières sur l'espace des fonctions continues de $\mathbb{R}_+$ dans $\mathbb{R}^d$. Nous établirons au cours de ce manuscrit un résultat de convergence qui ne supposent pas de condition d'irréductibilité.

1.6 Échelonnage optimal d'algorithmes de type Metropolis-Hastings

Une autre approche pour l'étude des algorithmes de type Metropolis-Hastings en grande dimension est l'étude de l'échelonnage optimal de telles méthodes. Considérons une mesure cible π sur $\mathbb{R}^d$. Si on souhaite appliquer l'algorithme de Metropolis à marche aléatoire symétrique ou l'algorithme MALA, un paramètre a besoin d'être choisi : respectivement la taille des incréments de la marche aléatoires et le pas de discrétisation dans MALA. Les études d'échelonnage optimal ont pour but de trouver le meilleur choix possible pour ce paramètre (en un certain sens) et sa dépendance en la dimension d. Nous donnons dans cette section une présentation des résultats d'échelonnage optimal pour ces deux algorithmes.

1.6.1 L'échelonnage optimal de l'algorithme de Metropolis à marche aléatoire symétrique

L'analyse par échelonnage optimal des algorithmes de type Metropolis-Hastings a été proposée dans le papier fondateur [RGG97]. Soit π^d une densité cible positive ($\pi^d(x) > 0$ pour tout $x \in \mathbb{R}^d$) par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$, $d \geq 1$. Nous considérons l'algorithme de Metropolis à marche aléatoire symétrique avec incrément gaussien qui définit la chaîne de Markov $(Y_k^d)_{k\in\mathbb{N}}$ comme suit : pour une condition initiale Y_0^d donnée, pour tout $k \geq 0$,

$$Y_{k+1}^d = \begin{cases} Y_k^d + \sigma_d Z_{k+1}^d & \text{avec probabilité } \alpha\left(Y_k^d, Y_k^d + \sigma_d Z_{k+1}^d\right) \\ Y_k^d & \text{sinon} \ , \end{cases} \tag{1.18}$$

où $\sigma_d \in \mathbb{R}_+^*$, $(Z_i^d)_{i\geq 1}$ est une suite i.i.d. de gaussiennes centrées réduites d-dimensionnelles et pour tout $x, y \in \mathbb{R}^d$

$$\alpha^d(x, y) = \min\left(1, \pi^d(y)/\pi^d(x)\right) \ . \tag{1.19}$$

En pratique le choix du paramètre σ_d est laissé à l'utilisateur. On peut alors se demander si il existe un choix optimal pour ce paramètre et si oui qu'elle est sa valeur et sa dépendance en la dimension. A première vue, on pourrait penser à prendre σ_d le plus grand possible pour que la chaîne puisse visiter l'espace plus facilement. Cependant, il faut veiller à ce que les mouvements soient acceptés en proportion non négligeable. Cet aspect limite le choix de σ_d. Une première étape est d'identifier la dépendance en la dimension du paramètre σ_d de telle sorte que le ratio d'acceptation moyen, en régime stationnaire ($Y_0^d \sim \pi^d$), admette une limite lorsque la dimension tend vers l'infini dans l'intervalle $]0, 1[$, i.e.

$$\lim_{d\to+\infty} \mathbb{E}_{\pi^d}\left[\alpha\left(Y_0^d, Y_0^d + \sigma_d Z_1^d\right)\right] \in]0, 1[\ .$$

[RGG97] s'est intéressé au cas où la densité π^d est de la forme:

$$\pi^d(x^d) = \prod_{i=1}^{d} \exp\left(-\mathrm{U}(x_i^d)\right) \ , \qquad x^d = (x_1^d, \cdots, x_d^d) \ , \tag{1.20}$$

où $\mathrm{U} : \mathbb{R} \to \mathbb{R}$ est une fonction qui satisfait :

H 1 U est trois fois continûment dérivable et U′ est Lipschitz.

H 2 $\mathbb{E}[(\mathrm{U}'(W))^8] < \infty$ et $\mathbb{E}[(\mathrm{U}''(W))^4] < \infty$ où W est une variable aléatoire de loi π^1.

[RGG97, corollary 1.2] montre que en régime stationnaire si $\sigma_d = \ell d^{-1/2}$ avec $\ell > 0$, et Y_0^d est de loi π^d,

$$\lim_{d\to+\infty} \mathbb{E}_{\pi^d}\left[\alpha\left(Y_0^d, Y_0^d + \ell d^{-1/2} Z_1^d\right)\right] = 2\boldsymbol{\Phi}(-\ell/2\sqrt{I}) \ , \tag{1.21}$$

où $\mathbf{\Phi}$ est la fonction de répartition de la loi gaussienne standard et

$$I = \int_{\mathbb{R}} (\mathrm{U}')^2(x)\pi^1(x)\mathrm{d}x \; . \tag{1.22}$$

[RGG97, Theorem 1.1] montre que en régime stationnaire, chaque composante de la chaîne de Markov $(Y_i^d)_{i\in\mathbb{N}}$, correctement mise à l'échelle, converge faiblement vers la solution de l'equation de Langevin associée à π^1. Plus précisément, considérons la suite de chaînes de Markov $\{(Y_k^d)_{k\geq 0}\,;\, d\geq 1\}$ définies par (1.18) en prenant pour tout $d\geq 1$, Y_0^d est de loi π^d et

$$\sigma_d = \ell d^{-1/2} \; .$$

[RGG97, Theorem 1.1] définit la suite de processus de sauts markovien $\{(\mathbf{Y}_t^d)_{t\geq 0}\,;\, d\geq 1\}$ à partir de $\{(Y_i^d)_{i\in\mathbb{N}}\,;\, d\geq 1\}$ par

$$\text{pour tout } t\geq 0 \text{ et } d\geq 1 \; , \qquad \mathbf{Y}_t^d = Y_{S_t}^d \; ,$$

où $(S_t)_{t\geq 0}$ est un processus de Poisson homogène d'intensité d. Notons $\mathbf{Y}_{t,1}^d$ la première composante de $\mathbf{Y}_t^d$. Si π^d est donnée par (1.20) et U vérifie H1 et H2, la suite de processus $\{(\mathbf{Y}_{t,1}^d)_{t\geq 0}\,;\, d\geq 1\}$ converge faiblement dans l'espace de Skorokhod vers la solution de l'equation de Langevin

$$\mathrm{d}\mathbf{Y}_t = \sqrt{\mathrm{h}(\ell)}\mathrm{d}B_t^1 - \frac{1}{2}\mathrm{h}(\ell)\mathrm{U}'(\mathbf{Y}_t)\mathrm{d}t \; , \tag{1.23}$$

où $(B_t^1)_{t\geq 0}$ est un mouvement brownien unidimensionnel, $\mathbf{Y}_0$ est de loi π^1 et $\mathrm{h}(\ell)$ est donnée par

$$\mathrm{h}(\ell) = 2\ell^2\Phi\left(-\frac{\ell}{2}\sqrt{I}\right) \; , \tag{1.24}$$

et I est défini par (1.22).

Ce résultat permet de choisir en pratique le paramètre ℓ. Considérons la diffusion de Langevin suivante

$$\mathrm{d}\mathbf{Y}_t^c = -c\nabla\mathrm{U}(\mathbf{Y}_t^1)\mathrm{d}t + \sqrt{2c}\mathrm{d}B_t^1 \; ,$$

pour $c\in\mathbb{R}_+^*$ et $\mathbf{Y}_0^c = u\in\mathbb{R}$. Sous de bonnes conditions sur U, [Bha78] montre que pour toute fonction $f\in\mathrm{L}^2(\pi^1)$ telle que $\int_{\mathbb{R}} f(x)\pi^1(\mathrm{d}x) = 0$, le processus $t\mapsto t^{-1/2}\int_0^t f(\mathbf{Y}_s^1)\mathrm{d}s$ converge faiblement vers une loi normale centrée de variance $\mathbb{E}_\pi[(\int_0^{+\infty} f(\mathbf{Y}_s^1)\mathrm{d}s)^2]$. On peut alors observer par un simple changement de variable que plus c est grand, plus cette quantité est petite. Ainsi, on aimerait trouver la constante ℓ qui maximise $\mathrm{h}(\ell)$ dans (1.23). D'après (1.21), un calcul montre que la fonction $\ell\mapsto\mathrm{h}(\ell)$ est maximale en point tel que la limite (1.21) vaut 0.234.

Les résultats introduits sont établis sous la condition que le potentiel U est au moins trois fois différentiable. Nous présenterons une extension de ce résultat sous des conditions beaucoup moins fortes sur le potentiel.

1.6.2 L'échelonnage optimal du MALA

La même étude peut être menée pour MALA. Soit π^d une densité cible positive par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ sous la forme (1.20) avec $\mathrm{U} : \mathbb{R} \to \mathbb{R}$ une fonction continûment dérivable. Dans ce cas, MALA définit la chaîne de Markov $(Y_i^d)_{i\in\mathbb{N}}$ pour une condition initiale Y_0^d donnée, pour tout $k \geq 0$,

$$Y_{k+1}^d = \begin{cases} W_{k+1}^d = Y_k^d + \sigma_d^2 \nabla \log \pi^d(Y_k^d) + \sqrt{2}\sigma_d Z_{k+1}^d & \text{avec probabilité } \alpha\left(Y_k^d, W_{k+1}^d\right) \\ Y_k^d & \text{sinon} \end{cases}, \tag{1.25}$$

où $\sigma_d \in \mathbb{R}_+^*$, $(Z_i^d)_{i\geq 1}$ est une suite i.i.d. de gaussiennes centrées réduites d-dimensionnelle et pour tout $x, y \in \mathbb{R}^d$

$$\alpha^d(x,y) = \min\left\{1, \frac{\pi^d(y)\exp\left(-\left\|x - y - \sigma_d^2 \nabla \log \pi^d(y)\right\|^2 / (4\sigma_d^2)\right)}{\pi^d(x)\exp\left(-\left\|y - x - \sigma_d^2 \nabla \log \pi^d(x)\right\|^2 / (4\sigma_d^2)\right)}\right\}. \tag{1.26}$$

Comme pour l'algorithme de Metropolis à marche aléatoire symétrique, le choix du paramètre σ_d est laissé à l'utilisateur.

On peut alors avoir les mêmes considérations: il est naturel de chercher à établir la dépendance en la dimension de σ_d pour que $\mathbb{E}_{\pi^d}[\alpha^d(Y_0^d, W_1^d)]$ converge lorsque la dimension tend vers l'infini vers une limite dans $]0,1[$. Le paramètre σ_d doit cette fois-ci être de la forme $\sigma_d = \ell d^{-1/6}$ pour $\ell \in \mathbb{R}_+^*$. En effet, [RR98, Theorem 1] montre que si le potentiel U vérifie:

M 1 U est huit fois continûment dérivable et il existe un polynôme P satisfaisant pour tout $i \in \{0, \cdots, 8\}$ et $x \in \mathbb{R}$, $\left|\mathrm{U}^{(i)}(x)\right| \leq \mathrm{P}(x)$

M 2 Pour tout $p \geq 1$, $\int_{\mathbb{R}} |x|^p \pi_1(x)\mathrm{d}x < \infty$

alors

$$\lim_{d\to+\infty} \mathbb{E}_{\pi^d}\left[\alpha\left(Y_0^d, Y_0^d + \ell^2 d^{-1/3}\nabla \log \pi^d(Y_0^d) + \sqrt{2}\ell d^{-1/6} Z_1^d\right)\right] = 2\boldsymbol{\Phi}(-\ell^3\sqrt{2J}), \tag{1.27}$$

où $\boldsymbol{\Phi}$ est la fonction de répartition de la loi gaussienne standard, $\ell > 0$ et

$$J = (48)^{-1}\int_{\mathbb{R}} \left\{5(\mathrm{U}^{(3)}(x))^2 - 3(\mathrm{U}^{(2)}(x))^3\right\} \pi^1(x)\mathrm{d}x > 0. \tag{1.28}$$

[RR98, Theorem 2] montre qu'il existe un limite diffusive pour MALA. Considérons la suite de chaînes de Markov $\{(Y_i^d)_{i\in\mathbb{N}}\,;\, d \geq 1\}$ donnée pour tout $d \geq 1$ par (1.25) avec comme condition initiale Y_0^d de loi π^d, et

$$\sigma_d = \ell d^{-1/6}.$$

On définit la suite de processus de sauts markovien $\{(\mathbf{Y}_t^d)_{t\geq 0}\,;\, d \geq 1\}$ à partir de $\{(Y_i^d)_{i\in\mathbb{N}}\,;\, d \geq 1\}$ par

$$\text{pour tout } t \geq 0 \text{ et } d \geq 1\,, \qquad \mathbf{Y}_t^d = Y_{S_t}^d\,,$$

où $(S_t)_{t\geq 0}$ est un processus de Poisson homogène d'intensité $d^{1/3}$. Notons $\mathbf{Y}_{t,1}^d$ la première composante de $\mathbf{Y}_t^d$. Si π^d est donnée par (1.20) et U vérifie M1 et M2, la suite de processus $\{(\mathbf{Y}_{t,1}^d)_{t\geq 0}\,;\, d \geq 1\}$ converge faiblement dans l'espace de Skorokhod vers la solution de l'equation de Langevin

$$\mathrm{d}\mathbf{Y}_t = \sqrt{2\mathrm{g}(\ell)}\mathrm{d}B_t^1 - \mathrm{g}(\ell)\mathrm{U}'(\mathbf{Y}_t)\mathrm{d}t\,, \tag{1.29}$$

où $(B_t^1)_{t\geq 0}$ est un mouvement brownien unidimensionnel, $\mathbf{Y}_0$ est de loi π^1 et $\mathrm{g}(\ell)$ est donnée par

$$\mathrm{g}(\ell) = 8\ell^2\Phi\left(-\ell^3\sqrt{2J}\right)\,, \tag{1.30}$$

et J est défini par (1.28).

Comme pour l'algorithme de Metropolis à marche aléatoire symétrique, ce résultat permet aussi de choisir en pratique le paramètre ℓ. D'après (1.27), un calcul montre que la fonction $\ell \mapsto \mathrm{g}(\ell)$ est maximale pour une valeur moyenne du ratio d'acceptation de l'ordre de 0.574. On peut observer que l'algorithme MALA permet en régime stationnaire de prendre des pas plus grands vis à vis de la dimension que l'algorithme de Metropolis à marche aléatoire symétrique le permet. Nous présenterons dans un chapitre une alternative à l'algorithme MALA avec un meilleur échelonnage dimensionnel.

1.7 Plan du manuscrit et contributions

Cette thèse est divisée en trois parties qui correspondent aux différents thèmes et travaux que nous avons menés. Chacune est constituée de deux chapitres. À une exception près, ces chapitres sont des articles acceptés ou en révision.

Dans la première partie, nous établissons des résultats de convergence de chaînes de Markov sur un espace polonais en distance de Wasserstein, puis nous appliquons ces résultats à des algorithmes MCMC.

Dans la seconde partie, nous nous intéressons à des bornes explicites pour l'algorithme ULA en variation totale et en distance de Wasserstein sous différentes conditions sur la densité cible.

Enfin nous traitons dans la dernière partie de l'échelonnage optimal pour l'algorithme de Metropolis à marche aléatoire symétrique et des accélérations de l'algorithme MALA.

Nous présentons ci-dessous un résumé des contributions et des résultats des chapitres de ce manuscrit.

1.7.1 Partie I Chapitre 3

Dans ce premier chapitre, nous établissons tout d'abord un résultat de convergence géométrique en distance de Wasserstein pour un noyau de transition P sur un espace polonais $(\mathsf{E}, \mathbf{d})$ muni de sa tribu borélienne $\mathcal{E}$. Il est obtenu en combinant une condition de contraction locale en distance de Wasserstein pour le noyau et une condition de dérive géométrique de type Foster-Lyapounov. Plus précisément, on suppose que P vérifie une condition de contraction sur un ensemble $\mathtt{G} \in \mathcal{E} \otimes \mathcal{E}$ pour la distance de Wasserstein associée à la métrique de E, i.e. il existe $\epsilon \in]0,1]$ et $\ell \in \mathbb{N}^*$ tels que pour tout $x, y \in \mathtt{G}$,

$$W_{\mathbf{d}}(P^\ell(x,\cdot), P^\ell(y,\cdot)) \leq (1-\epsilon)\mathbf{d}(x,y) . \tag{1.31}$$

On suppose de plus qu'il existe une fonction mesurable $V : \mathsf{E} \to [1, +\infty[$, $b \in \mathbb{R}_+$ et $\lambda \in [0,1[$ tels que pour tout $x, y \in \mathsf{E}$,

$$PV(x) + PV(y) \leq \lambda(V(x) + V(y)) + b\mathbb{1}_{\mathtt{G}}(x,y) , \quad \sup_{(z,w)\in\mathsf{E}\times\mathsf{E}} \{V(z) + V(w)\} < +\infty . \tag{1.32}$$

Cette dernière condition est similaire à la condition de dérive géométrique (1.14) mais est formulée sur l'espace produit. Le résultat que nous déduisons de ces deux hypothèses donne un contrôle quantitatif sur la convergence de la chaîne lorsque cette dernière n'est pas irréductible. D'autre part, même dans le cas irréductible, il fournit des bornes de convergence en distance de Wasserstein qui peuvent être plus précises que pour la variation totale. Ce type de résultat avait déjà été établi dans [HMS11], cependant les bornes obtenues dans ce chapitre sont plus directement exploitables car elles dépendent de façon plus simple des constantes apparaissant dans (1.31) et (1.32). De plus, la méthode de preuve est complètement différente et s'effectue directement par couplage.

Nous appliquons ensuite ce résultat pour l'analyse d'un algorithme MCMC présenté ci-dessous. Considérons une densité cible donnée pour tout $x \in \mathbb{R}^d$ par

$$\pi(x) = \mathcal{Z}^{-1}\exp(-U(x) - \Gamma(x)) ,$$

avec $\mathcal{Z} = \int_{\mathbb{R}^d} \exp(-U(x) - \Gamma(x))\mathrm{d}x < \infty$, $\Gamma : \mathbb{R}^d \to \mathbb{R}$ et $U : \mathbb{R}^d \to \mathbb{R}$ de la forme pour tout $x \in \mathbb{R}^d$,

$$U(x) = (1/2)x^T Q x + \Upsilon(x) ,$$

où Q est une matrice définie positive et $\Upsilon : \mathbb{R}^d \to \mathbb{R}$ une fonction convexe de gradient Lipschitz. On considère alors la diffusion de Langevin associée à U:

$$\begin{aligned} \mathrm{d}\mathbf{Y}_t &= -\mathbf{Y}_t\mathrm{d}t - Q^{-1}\nabla\Upsilon(\mathbf{Y}_t)\mathrm{d}t + \sqrt{2}Q^{-1/2}\mathrm{d}B_t^d , \\ \mathbf{Y}_0 &= y_0 , \end{aligned}$$

où $(B_t^d)_{t\geq 0}$ est un mouvement brownien d-dimensionnel. Par hypothèse sur U, cette EDS admet une unique solution forte $(\mathbf{Y}_t)_{t\geq 0}$. Il est aisé de voir que pour tout $t \geq 0$ et $\delta > 0$,

$$\mathbf{Y}_{t+\delta} = \mathbf{Y}_t\mathrm{e}^{-\delta} - \int_t^{t+\delta} \mathrm{e}^{-(t+\delta-s)}Q^{-1}\nabla\Upsilon(\mathbf{Y}_s)\mathrm{d}s + \sqrt{2}Q^{-1/2}\int_t^{t+\delta}\mathrm{e}^{-(t+\delta-s)}\mathrm{d}B_s^d . \tag{1.33}$$

Une discrétisation de Euler-Maruyama de (1.33) avec un pas $\delta > 0$ mène alors à la chaîne de Markov $(Y_k)_{k\in\mathbb{N}}$, définie pour une condition initiale donnée pour tout $k \geq 0$ par

$$\bar{Y}_{k+1} = \bar{Y}_k \mathrm{e}^{-\delta} - (1 - \mathrm{e}^{-\delta}) Q^{-1} \nabla \Upsilon(\bar{Y}_k) + Z_{k+1} \;, \tag{1.34}$$

où $(Z_k)_{k\in\mathbb{N}^*}$ est une suite i.i.d. de variables aléatoires gaussiennes centrées et de matrice de covariance $(1-\mathrm{e}^{-2\delta})Q^{-1}$. Ce type de discrétisation est appelé intégrateur stochastique exponentiel d'Euler, voir [LR04]. La chaîne associée à la relation (1.34) définit un noyau de Markov qui est alors utilisé comme loi de proposition dans un algorithme de type Metropolis-Hastings avec comme loi cible π. L'algorithme ainsi défini sera appelé EI-MALA. Cet algorithme généralise un algorithme de Metropolis pour la simulation de diffusion proposé dans [Bes+08] : dans ce cas $h = 2(1 - \mathrm{e}^{-\delta})$ et $\Gamma = 0$. Cet algorithme a été analysé dans [Ebe14] sous certaines conditions sur Q et Υ. Nous complétons cette analyse en établissant sous des conditions appropriées, une convergence géométrique du noyau produit par l'algorithme avec des bornes explicites. En particulier, nous montrons une dépendance logarithmiquement polynomiale en la dimension dans le cas où $\Upsilon = 0$ et Γ est bornée. Aussi nous vérifions que les hypothèses de notre résultat principal sont satisfaites dans le cas d'un problème inverse bayesien. Enfin, nous donnons quelques simulations numériques pour illustrer les performances de la méthode.

Ce travail a fait l'objet de la publication [DM15b].

1.7.2 Partie I Chapitre 4

Nous complétons dans cette partie le chapitre 3 sur l'étude de chaînes de Markov qui ne sont pas nécessairement irréductibles, en particulier de chaînes dans des espaces d'état fonctionnels. Nous cherchons à établir

- des conditions d'existence et surtout d'unicité de la mesure invariante d'un noyau de Markov P sur un espace polonais $(\mathsf{E}, \mathbf{d})$, muni de sa tribu borélienne $\mathcal{E}$,
- des taux de convergences de la chaîne vers sa mesure stationnaire.

À l'instar du cas irréductible, on suppose aussi qu'il existe une fonction mesurable $V : \mathsf{E} \to [1, +\infty)$ et $b \in \mathbb{R}_+$ tels que pour tout $x, y \in \mathsf{E}$,

$$PV(x) + PV(y) \leq V(x) + V(y) - 1 + b\mathbb{1}_{\mathtt{G}}(x, y) \;, \qquad \sup_{(z,w)\in\mathtt{G}} \{V(z) + V(w)\} < +\infty \;,$$

où $\mathtt{G} \in \mathcal{E} \otimes \mathcal{E}$ est un ensemble pour lequel le noyau vérifie la condition de contraction (1.31). Cette condition est similaire à la condition (1.11) dans le cas irréductible mais est formulée sur l'espace produit.

Nous nous établlissons des bornes quantitatives de convergence en distance de Wasserstein. De telles bornes dans le cas géométrique ont été obtenues par [HMS11] et le travail présenté en chapitre 3. Nous établissons dans ce chapitre un résultat de convergence sous-géométrique en distance de Wasserstein sous des conditions de dérive sous-géométriques. Ces conditions sont similaires à celles proposées dans [TT94] et [Dou+04] pour établir des convergences sous-géométrique en variation totale. La convergence en distance de Wasserstein avait déjà été considéré dans [But14] mais les taux obtenus ne correspondaient pas aux taux connus pour la distance en variation totale.

Nos travaux, appliqués à la variation totale, conduisent aux même taux de convergences que ceux donnés dans [Dou+04] pour la distance en variation totale (dans le cas des taux polynomiaux et logarithmiques). Alors que la méthode de [But14] s'inspire de [HMS11], notre preuve s'inspire de celle dans le cas géométrique présentée dans le chapitre 3 et d'idées présentes dans [Dou+04].

Nous considérons deux applications de nos contributions théoriques. Nous étudions d'abord un modèle fonctionnel auto-régressif dans $\mathbb{R}^d$ de la forme

$$Y_{k+1} = \mathrm{h}(Y_k) + \xi_{k+1} \ , \tag{1.35}$$

où $\mathrm{h} : \mathbb{R}^d \to \mathbb{R}^d$ est une fonction mesurable et $(\xi_k)_{k\in\mathbb{N}^*}$ est une suite i.i.d. de vecteurs aléatoires. Notons que si la suite $(\xi_k)_{k\in\mathbb{N}^*}$ est à valeurs dans un espace discret, alors la chaîne n'est pas irréductible. On suppose que h est une contraction stricte au centre de l'espace, i.e. il existe une fonction $\varpi : \mathbb{R}^d \times \mathbb{R}^d \to]0,1[$ telle que pour tout $x, y \in \mathbb{R}^d$,

$$\|\mathrm{h}(x) - \mathrm{h}(y)\| \leq \varpi(x,y) \|x - y\| \ ,$$

et pour tout $R \in \mathbb{R}_+$, $\phi(R) = \sup\{\varpi(x,y) \mid \|x\| + \|y\| \leq R\} < 1$. Lorsque $\sup_{R\in\mathbb{R}_+} \phi(R) < 1$, alors le modèle est géométriquement ergodique en distance de Wasserstein. Nous nous intéressons aux cas où ϕ tend vers 1 lorsque R tend vers $+\infty$. Suivant des conditions sur le taux de cette convergence et des conditions de moments sur la loi de la suite $(\xi_k)_{k\in\mathbb{N}^*}$ nous établissons des taux de convergence sous géométriques pour le noyau associé à (1.35).

Le deuxième exemple traite de l'algorithme de Crank-Nicolson pré-conditionné. Ce dernier est un algorithme de type Metropolis-Hastings défini sur un espace de Hilbert H muni d'une mesure gaussienne μ_0 et pour une mesure cible ayant une densité positive $x \to \exp(-\Phi(x))$ par rapport à μ_0. Le noyau de proposition est celui associé à celui d'un modèle auto-regressif :

$$W_{k+1} = (1-\eta)^{1/2} W_k + \eta^{1/2} Z_{k+1} \ , \tag{1.36}$$

où $\eta \in [0,1[$ et $(Z_k)_{k\in\mathbb{N}^*}$ est une suite de variables aléatoires gaussiennes de loi μ_0. Observons que si G_1 et G_2 sont deux variables aléatoires gaussiennes de loi μ_0, alors pour tout $\eta \in]0,1[$, $(G_1, (1-\eta)^{1/2} G_1 + \eta^{1/2} G_2)$ et $((1-\eta)^{1/2} G_1 + \eta^{1/2} G_2, G_1)$ sont toujours deux gaussiennes de même loi. On en déduit que le noyau Q associé à la chaîne de Markov définie par (2.35) est réversible par rapport à μ_0. Donc d'après la discussion sur l'algorithme de Metropolis-Hastings en section 1.3-(B), le noyau de Metropolis-Hastings P défini par (1.4) associé au noyau Q et au ratio d'acceptation donné pour tout $x, y \in \mathsf{H}$ par

$$\alpha(x,y) = \min\left\{1, \exp\left(\Phi(x) - \Phi(y)\right)\right\} \ ,$$

est réversible par rapport à π. [HSV14] ont montré si que le ratio d'acceptation est uniformément minoré, alors le noyau produit par l'algorithme de Metropolis est géométriquement ergodique pour une certaine distance de Wasserstein. Nous affaiblissons cette condition et montrons que sous l'hypothèse que Φ est Lipschitz, alors le noyau produit par l'algorithme est sous-géométriquement ergodique.

Ce travail a été accepté pour publication dans les annales de l'IHP.

1.7.3 Partie II Chapitre 5

Ce chapitre a pour sujet l'étude fine de la convergence de l'algorithme ULA présenté en section 1.4.2 pour échantillonner suivant une loi π sur $\mathbb{R}^d$, $d \geq 1$, de la forme (1.8). Nous établissons des bornes explicites en variation totale entre les lois marginales de la chaîne de Markov produite et la mesure cible π. Nous obtenons des résultats à la fois pour des algorithmes à pas constants et à pas décroissants.

La méthode de preuve est très différente des approches classiques, notre objectif étant d'obtenir des résultats "utilisables". En effet, pour les diffusions de Langevin, de très nombreux résultats ont été établis récemment pour quantifier précisément la distance en variation totale entre la loi de la diffusion et sa limite. Ces résultats découlent soit de méthodes fines d'analyse fonctionnelles (inégalités de Poincaré ou de Sobolev logarithmique), voir par exemple [Bak+08] et [BGL14], soit par des méthodes de couplage [Ebe15]. Nous obtenons en particulier des résultats originaux de convergence en variation totale en utilisant le couplage par réflexion [LR86] et les méthodes quantitatives de contrôle de convergence que nous avons mises en évidence dans [DM15b].

Pour passer des résultats à temps continu à temps discret, nous utilisons une approche due à [Dal16] basée sur un couplage direct de la diffusion et d'une interpolation continue du processus discret (la loi de la diffusion est absolument continue par rapport à la loi de l'interpolation continue de l'algorithme, ce qui permet d'évaluer la formule de Girsanov pour évaluer l'entropie relative des deux lois).

À partir de ces résultats, nous comparons différentes stratégies à horizon fini (le nombre d'itérations de l'algorithme est fixé à l'avance) ou infini (l'algorithme peut être interrompu à n'importe quel instant). En particulier, suivant les conditions de courbures et/ou convexité et de régularité du potentiel associé à π, nous nous intéressons à la dépendance des bornes par rapport à la dimension de l'espace, ce qui est bien entendu crucial pour les applications statistiques. Nous considérons de nombreux cas: potentiel super-exponentiel, convexe, fortement convexe, fortement convexe à l'extérieur d'un ensemble compact, qui correspondent à des situations d'intérêt pratique. Pour chacun de ces cas, nous obtenons dans certains scénarios des contrôles dépendant de façon polynomiale de la dimension.

À pas constant γ, sous des conditions faibles sur π, la chaîne de Markov produite par ULA admet une mesure stationnaire π_γ. Nous donnons dans ce chapitre une borne explicite en V-variation totale entre la mesure cible π et π_γ d'ordre $\sqrt{\gamma}$.

Nous montrons que si la suite de pas $(\gamma_k)_{k\geq 0}$ converge vers 0 et $\sum_{k=1}^{+\infty} \gamma_k = +\infty$, alors la suite des lois marginales de la chaîne inhomogène produite converge vers la mesure cible π, toujours en variation totale.

Ces résultats complètent [Dal16], où seul le cas des mesures fortement log-concaves est considéré. Il permet de comprendre dans quelles situations il est préférable d'utiliser l'algorithme à pas fixe ou à pas décroissant (il donne dans ce dernier cas des indications précieuses sur la vitesse à laquelle la suite des pas doit tendre vers 0).

Ce travail a été accepté pour publication à Annals of Applied Probability.

1.7.4 Partie II Chapitre 6

Dans ce chapitre, nous complétons et améliorons les résultats du chapitre précédent dans le cas de densités fortement log-concaves π sur $\mathbb{R}^d$, $d \geq 1$, de la forme (1.8).

Tout d'abord la convergence en distance de Wasserstein est considérée. Des bornes explicites sont établies en utilisant le couplage synchrone entre la diffusion de Langevin et sa discrétisation d'Euler associée à une suite de pas $(\gamma_k)_{k\in\mathbb{N}^*}$. Ce couplage utilise le même bruit et est défini pour des conditions initiales données par :

$$\begin{cases} \mathbf{Y}_t = \mathbf{Y}_0 - \int_0^t \nabla U(\mathbf{Y}_s)\mathrm{d}s + \sqrt{2}B_t^d & \text{pour tout } t \geq 0 \\ Y_{k+1} = Y_k - \gamma_{k+1}\nabla U(Y_k) + \sqrt{2}(B^d_{\Gamma_{k+1}} - B^d_{\Gamma_k}) & \text{pour tout } k \in \mathbb{N}^* , \end{cases}$$

pour un mouvement brownien $(B_t^d)_{t\geq 0}$ d-dimensionnel et où $\Gamma_k = \sum_{i=1}^k \gamma_i$ pour tout $k \in \mathbb{N}^*$. Ces bornes permettent d'établir un taux explicite de convergence en distance de Wasserstein des lois marginales de la chaîne de la Markov inhomogène vers π, lorsque la suite des pas tend vers 0 et $\sum_{k=1}^{+\infty} \gamma_k = +\infty$.

À pas constant $\gamma > 0$, nous en déduisons une borne explicite de la distance de Wasserstein entre la mesure invariante π_γ associée au noyau de la discrétisation d'Euler π. Concernant la dépendance en la dimension, pour une précision $\epsilon > 0$ donnée, nous montrons que sous deux jeux d'hypothèses différentes (qui se distinguent par la régularité de U) nous obtenons que le nombre d'itérations est de l'ordre $d\epsilon^{-2}$ ou $\sqrt{d}\epsilon^{-1}$ pour que la distance de Wasserstein entre la loi marginale de l'algorithme et la loi cible π soit inférieure à ϵ.

Nous adaptons ensuite les résultats de [JO10] au cadre des chaînes considérées qui sont inhomogènes pour obtenir des bornes explicites sur l'erreur en moyenne quadratique et des inégalités exponentielles de déviations pour des estimateurs de Monte Carlo de $\int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ avec $f : \mathbb{R}^d \to \mathbb{R}$, Lipschitz.

Dans un second temps, nous utilisons les bornes en distance de Wasserstein pour établir des bornes en variation totale. En effet lorsque U est fortement convexe, alors le semi-groupe à un effet régularisant. Soit f une fonction mesurable bornée. Si on note $(\mathbf{P}_t)_{t\geq 0}$ le semi-groupe associé à la diffusion de Langevin, pour tout $t > 0$, $\mathbf{P}_t f$ est lipschitzienne avec un coefficient qui est explicite et dépend de t. Cependant, ce coefficient en temps court est de l'ordre de $1/\sqrt{t}$ et c'est pourquoi nous utilisons en partie la technique de preuve utilisée dans le chapitre 5. Le résultat que l'on déduit donne des taux de convergence en variation totale vers π, lorsque la suite de pas tend vers 0 et $\sum_{k=1}^{+\infty} \gamma_k = +\infty$. Ces taux améliorent très significativement les résultats que nous obtenons dans le chapitre 5. Lorsque les pas de discrétisation sont constants, les bornes obtenues impliquent que si U est suffisamment régulière, un nombre d'itérations de l'ordre de $\sqrt{d}\epsilon^{-1}$ est suffisant pour que la loi marginale de l'algorithme soit à une distance $\epsilon > 0$ de π en variation totale. Encore une fois, ce résultat améliore celui obtenu dans le chapitre 5 qui donnerait un nombre d'itérations de l'ordre de $d\epsilon^{-2}$.

Nous concluons ce travail en établissant des bornes explicites sur l'erreur en moyenne quadratique et des inégalités exponentielles de déviations pour des estimateurs de $\int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ lorsque $f : \mathbb{R}^d \to \mathbb{R}$ mesurable bornée. Cependant, nous ne pouvons pas appliquer directement la technique utilisée lorsque f est lipschitzienne. Nous observons d'abord que le noyau associé à la discrétisation d'Euler a lui-même un effet régularisant. Pour cela, on s'intéressera aux modèles fonctionnels auto-régressifs de la forme:

$$W_{k+1} = \mathrm{h}(W_k) + \sigma Z_{k+1} , \tag{1.37}$$

où $\mathrm{h} : \mathbb{R}^d \to \mathbb{R}$ est Lipschitz, $\sigma > 0$ et $(Z_k)_{k\in\mathbb{N}^*}$ est une suite de variables aléatoires gaussiennes centrées réduites d-dimensionnelles. Nous montrons que tout noyau associé à un tel modèle est régularisant en nous inspirant de la construction d'un couplage donné dans [BDJ98] entre deux variables aléatoires gaussiennes d dimensionnelles de même variance mais de moyennes différentes.

Ce travail a été mené en collaboration avec le professeur Éric Moulines.

1.7.5 Partie III Chapitre 7

Le résultat original d'échelonnage optimal pour l'algorithme de Metropolis à marche aléatoire symétrique de [RGG97] suppose que la mesure cible possède une densité positive par rapport à la mesure de Lebesgue et est trois fois continûment dérivable, cf section 1.6.1. Il ne s'applique donc pas aux densités qui sont différentiables en moyenne quadratique, comme la loi de Laplace par exemple (qui n'est pas différentiable en certains points). Néanmoins, lorsque l'on analyse le résultat de [RGG97], on se rend aisément compte que la seule quantité qui apparaît dans la limite du taux moyen d'acceptation dans le régime stationnaire est la matrice d'information de Fisher associée au score de translation. Cette quantité est bien définie dès que la densité est différentiable en moyenne quadratique. Une question qui se pose dès lors est la possibilité d'obtenir un résultat d'échelonnage optimal tout en affaiblissant autant que possible les hypothèses. Une des motivations de ce travail est l'analyse des méthodes bayésiennes utilisant des lois a priori non régulières (convexes et lipschitziennes mais pas différentiables): celles-ci apparaissent très naturellement en inférence bayésienne en grande dimension (comme par exemple les normes L^1 pondérées dans l'analyse bayésienne de l'algorithme LASSO).

Le premier résultat de ce chapitre est l'extension du résultat d'échelonnage optimal de [RGG97] à des densités positives sur $\mathbb{R}$ et différentiables en moyenne L^p pour un $p \geq 2$, donc qui peuvent ne pas être différentiables en certains points. De plus, nous démontrons que le taux d'acceptation moyen à la stationnarité admet une limite lorsque $\sigma_d = \ell d^{-1/2}$, voir (1.21). Nous montrons aussi que l'algorithme admet une limite diffusive: la diffusion limite est une diffusion de type Langevin mais qui est en général singulière (qui peut ne pas admettre de solutions fortes).

La méthode de preuve est différente de celle de [RGG97]: elle est basée sur une approche directe de la convergence en loi. Nous montrons la tension d'une interpolation continue du processus discret puis nous identifions la distribution limite en utilisant l'équivalence de l'EDS à un problème de martingale. En outre, la preuve s'appuie sur les méthodes permettant d'établir les conditions LAN pour des densités différentiables en moyennes quadratiques (et notamment un développement astucieux du rapport de vraisemblance qui permet de s'affranchir d'une dérivée d'ordre 2). Nous appliquons ce résultat à l'algorithme du Bayesian LASSO.

Nous étendons aussi ce résultat pour des densités qui possèdent un support inclus dans un intervalle de $\mathbb{R}$ tout en restant différentiable en moyenne quadratique.

Ce résultat est alors appliqué à des densités de type Beta et Gamma. Ce résultat permet de compléter les travaux de [**neal:roberts:2012**].

Ce travail a été soumis à publication dans *Advances in Applied Probability*

1.7.6 Partie III Chapitre 8

Nous avons vu en section 1.6.2 que l'échelonnage optimal de MALA était en $d^{-1/3}$. Une question naturelle est de savoir si il serait possible d'améliorer cette dépendance en la dimension en incorporant plus d'informations sur la mesure cible.

Une première idée serait d'utiliser un intégrateur d'ordre plus élevé que le schéma d'Euler comme proposition dans un algorithme de Metropolis-Hastings. Dans ce chapitre, nous observons que utiliser un meilleur schéma de discrétisation n'améliore pas l'échelonnage optimal de l'algorithme de Metropolis Hastings induit.

Au lieu de se concentrer sur l'erreur de discrétisation, nous montrons que la quantité importante à contrôler est l'ordre du ratio d'acceptation. En effet, expliquons l'idée principale des preuves d'échelonnage optimal pour des densités de la forme (1.20) et des propositions associés à des chaînes de Markov données pour tout $k \in \mathbb{N}$,

$$W_{k+1}^d = \mathrm{F}(W_k^d, \sigma_d^{1/2}) + \Sigma(W_k^d, \sigma_d^{1/2}) Z_{k+1}^d \; , \tag{1.38}$$

où $\sigma_d \in \mathbb{R}_+^*$, $\mathrm{F} : \mathbb{R}^d \times \mathbb{R}_+^* \to \mathbb{R}^d$, $\Sigma : \mathbb{R}^d \times \mathbb{R}_+^* \to M_d(\mathbb{R}^d)$, $M_d(\mathbb{R}^d)$ étant l'ensemble des matrices carrées de dimension d, et $(Z_k)_{k\geq}$ est une suite i.i.d. de variables aléatoires gaussiennes centrées réduites d dimensionnelles. Notons que pour le cas de l'algorithme de Metropolis à marche aléatoire symétrique

$$\mathrm{F}(x, \sigma_d^{1/2}) = x, \quad \Sigma(x, \sigma_d^{1/2}) = \sqrt{\sigma_d}\,\mathrm{Id}\;,$$

et pour MALA

$$\mathrm{F}(x, \sigma_d^{1/2}) = x - \sigma_d \nabla \mathrm{U}(x), \quad \Sigma(x, \sigma_d^{1/2}) = \sqrt{2\sigma_d}\,\mathrm{Id}\;.$$

La ratio d'acceptation peut alors s'écrire sous la forme pour tout $x, y \in \mathbb{R}^d$

$$\alpha(x, y) = \min\{1, \exp(R_d(x, y))\}\;, \tag{1.39}$$

pour une fonction $R_d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ qui dépend de la loi de proposition (1.38). Nous nous intéressons aussi à la valeur du ratio lorsque y est sous la forme

$$y = \mathrm{F}(x, \sigma_d^{1/2}) + \Sigma(x, \sigma_d^{1/2}) z\;.$$

Si on suppose $R_d(x,x) = 0$ et des conditions de régularité sur F, Σ et U, alors un développement de Taylor de R_d peut être effectué en σ_d au voisinage de 0:

$$R_d(x,y) = \sum_{i=1}^{k}\sum_{j=1}^{d} \sigma_d^{i/2} C_i(x_j,z_j) + \sigma_d^{(k+1)/2} L_{k+1}(x,\sigma_d^{1/2},z) \ . \tag{1.40}$$

Il s'avère que l'échelonnage optimal associé à un noyau de proposition de la forme (1.38) est directement relié au nombre de termes C_i qui s'annulent dans (1.40). Si $C_i = 0$, pour $i = 1,\cdots,p$, $p \in \mathbb{N}^*$ et $\sigma_d = \ell d^{1/(p+1)}$ alors le terme dominant dans (1.40) est

$$\frac{\ell^{p+1}}{\sqrt{d}} \sum_{j=1}^{d} C_{p+1}(x_j,z_j) \ .$$

C'est grâce à cette étude du taux d'acceptation que l'on peut conclure que

$$\lim_{d\to+\infty} \mathbb{E}[\alpha(W_0^d, W_1^d)] \in]0,1[\ ,$$

où W_0^d a pour loi π^d et W_1^d est donnée par (1.38).

Pour obtenir un nouvel algorithme de type Metropolis-Hastings avec un meilleur échelonnage optimal que MALA, on restreint tout d'abord la classe des noyaux de proposition associée à (1.38) en imposant la forme suivante pour les fonctions F et Σ : pour $x \in \mathbb{R}^d$ et $\sigma_d > 0$,

$$\mathrm{F}(x,\sigma_d^{1/2}) = x + \sigma_d \mathrm{F}_1(x) + \sigma_d^2 \mathrm{F}_2(x) \ , \qquad \Sigma(x,\sigma_d^{1/2}) = \sigma_d^{1/2}\Sigma_1(x) + \sigma_d^{3/2}\Sigma_2(x) \ ,$$

où pour $i = 1,2$, $\mathrm{F}_i(x) : \mathbb{R}^d \to \mathbb{R}^d$ et $\Sigma_i : \mathbb{R}^d \to M_d(\mathbb{R}^d)$. On effectue alors un développement de Taylor de R_d défini par (1.39) en σ_d au voisinage de 0 pour de telles fonctions F et Σ. On cherche ensuite des expressions explicites pour les fonctions $\mathrm{F}_1, \mathrm{F}_2, \Sigma_1, \Sigma_2$ de façon à imposer à ce que les quatre premiers termes $C_i(\cdot,\cdot)$, $i \in \{1,2,3,4\}$ dans (1.40) soient identiquement nuls. On obtient ainsi un système de quatre équations à quatre inconnues qui admet une unique solution. Nous obtenons par cette méthode les expressions suivantes pour les fonctions F et Σ

$$\begin{aligned}
\mathrm{F}(x,\sigma_d^{1/2}) &= x - \frac{\sigma_d}{2}\nabla U^d(x) - \frac{\sigma_d^2}{24}\left(\nabla^2 U^d(x)\nabla U^d(x) - \vec{\Delta}(\nabla U^d)(x)\}\right) , \\
\Sigma(x,\sigma_d^{1/2}) &= \sigma_d^{1/2}\,\mathrm{Id} - (\sigma_d^{3/2}/12)\nabla^2 U^d(x) \ ,
\end{aligned}$$

où $U^d = \log \pi^d$ et $\vec{\Delta}$ est le laplacien vectoriel. Nous appelons l'algorithme de Metropolis-Hastings associé à un tel choix de F et Σ le fast Metropolis Adjusted Langevin Algorithm (fMALA). On note que les corrections ne coïncident pas avec les schémas d'intégration des EDS d'ordres plus élevés. Dans un second temps, nous montrons que l'échelonnage optimal de l'algorithme de Metropolis-Hastings associé à la proposition définie par (1.38) pour F et Σ données par (1.41) est en $d^{-1/5}$ et donc améliore l'ordre de l'échelonnage optimal associé à MALA. Nous étudions aussi d'autres formes pour les fonctions F et Σ et obtenons ainsi deux nouveaux algorithmes dont l'échelonnage optimal est en $d^{-1/5}$.

Nous établissons ensuite des résultats de stabilité et d'ergodicité pour ces nouveaux algorithmes. En particulier, nous montrons que l'un d'entre eux est géométriquement ergodique pour une classe de densités proche de celle définie par (1.17) en établissant une inégalité de dérive géométrique (1.14). Nous donnons aussi un critère sur F et Σ qui implique qu'un algorithme de Metropolis-Hastings dont le noyau de proposition est associé à (1.38), n'est pas géométriquement ergodique. Ce résultat généralise [RT96a, Théorème 4.2]. On utilise ce critère pour établir que fMALA n'est pas géométriquement ergodique si $\lim_{\|x\|\to+\infty} \|\nabla U(x)\| / \|x\| = +\infty$.

Enfin nous présentons des simulations numériques pour illustrer nos résultats, ainsi que des stratégies algorithmiques pour gérer le régime transitoire de l'algorithme. En effet, notre résultat d'échelonnage optimal ne s'applique uniquement qu'aux chaînes stationnaires. Il est à noter que le même problème se pose pour MALA mais que l'algorithme de Metropolis-Hastings à marche aléatoire symétrique a un échelonnage optimal de l'ordre de d^{-1} même si la chaîne n'est pas stationnaire, voir [JLM15]. Nous proposons alors une stratégie hybride similaire à celle dans [CRR05] pour MALA. Avec probabilité 1/2, nous utilisons le noyau associé à l'algorithme de Metropolis-Hastings à marche aléatoire symétrique et le noyau associé à l'algorithme fMALA sinon. Nous comparons alors empiriquement les deux stratégies hybrides pour MALA et fMALA et observons que celle utilisant fMALA converge plus rapidement et présente de meilleurs fonctions d'autocorrélations. Ce travail a été mené en collaboration avec les professeurs Gareth Roberts, Gilles Vilmart et Konstantinos Zygalakis. Il est en révision majeure à Annals of App. Prob.

1.7.7 Liste des travaux

Publications dans des revues à comité de lecture internationales

A. Durmus and É. Moulines. "Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm". In: *Stat. Comput.* 25.1 (2015), pp. 5–19. ISSN: 0960-3174.

A. Durmus, G. Fort, and É. Moulines. *Subgeometric rates of convergence in Wasserstein distance for Markov chains.* Accepted for publication in Ann. Inst. H. Poincaré Probab. Statist.

A. Durmus and É. Moulines. *Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm.* Accepted for publication in Ann. Appl. Probab.

Pré-publications et travaux en préparation

A. Durmus and É. Moulines. *Sampling from strongly log-concave distributions with the Unadjusted Langevin Algorithm.* First version : arXiv:1605.01559, in preparation. 2016.

A. Durmus, G. O Roberts, G. Vilmart, and K. C Zygalakis. *Fast Langevin based algorithm for MCMC in high dimensions.* arXiv:1507.02166, submitted. 2015.

A. Durmus, S. Le Corff, É Moulines, and G. O. Roberts. *Optimal scaling of the Random Walk Metropolis algorithm under Lp mean differentiability.* arXiv:1604.06664, submitted. 2016.

Publications dans des conférences à comité de lecture internationales

A. Durmus, U. Simsekli, É. Moulines, R. Badeau, and G. Richard. "Stochastic Gradient Richardson-Romberg Markov Chain Monte Carlo". In: *Thirtieth Annual Conference on Neural Information Processing Systems (NIPS).* 2016.

1.7.8 Liste des communications orales

Les diapositives des présentations mentionnées ci-dessous sont disponibles à l'adresse suivante : `https://perso.telecom-paristech.fr/~durmus/talk.html`.

a) Juillet 2016, Présentation à International Conference on Monte Carlo techniques, Paris (FR), *Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm,*

b) Juin 2016, Présentation au workshop MCMC and diffusions techniques, Londres (UK), *Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm*

c) Mai 2016, Présentation aux journées des statistiques de la SFDS, Montpellier (FR), *Echantillonage de loi log-concave en grande dimension*

d) Avril 2016, Présentation au Colloque des jeunes prob. et stat., les Houches (FR), *Echantillonage de loi log-concave en grande dimension*

e) Février 2016, Séminaire des doctorants du CERMICS, École des Ponts ParisTech, *Efficient sampling from log-concave distributions over high-dimensional spaces*

f) Mai 2015, Présentation au séminaire de statistique de l'université de Bristol (UK), *Subgeometric ergodicity of Markov Chain in Wasserstein distance*

g) Février 2015, Séminaire des doctorants du LSTA, UPMC *New Langevin-based Metropolis algorithm*

h) Février 2015, Présentation au séminaire de stochastique de l'université de Helsinki (FI), *New Langevin-based Metropolis algorithm*

i) Février 2015, Présentation à Bayes in Paris à l'ENSAE (FR)
New Langevin-based Metropolis algorithm

j) Octobre 2014, Présentation au séminaire d'analyse numérique de l'université de Genève (CH), *Geometric ergodicity in Wasserstein distance*

k) Septembre 2014, Présentation à une journée YSP à l'institut Henri Poincaré à Paris (FR), *Introduction au scaling optimal des algorithmes de Metropolis*

Chapter 2

Introduction

2.1 Motivation: a short digression on Bayesian statistics

Sampling from a probability distribution is the main motivation of the work presented in this thesis. There are of course many different applications of sampling. In particular, Bayesian inference is based on the exploration of the a posteriori distribution of a model.

Bayesian statistics suppose a probabilistic model on some observed data w which is assumed to be a sample from a random variable W valued in a measurable space $(\mathsf{W}, \mathcal{W})$. Here, we consider a parametric model for W, $(\mathsf{W}, \mathcal{W}, \mathcal{P}_\Theta)$ where $\mathcal{P}_\Theta$ is a set of probability measure defined by:

$$\mathcal{P}_\Theta = \{K(\vartheta, \cdot) \mid \vartheta \in \Theta\} \ .$$

K is a Markov kernel on $(\mathsf{W}, \mathcal{W})$ and $(\Theta, \mathcal{F})$ is a measurable space. In most applications, Θ is either discrete or a subset of $\mathbb{R}^d$ for $d \geq 1$. Assume that the model is dominated by a measure μ on $(\mathsf{W}, \mathcal{W})$, i.e. K admits a transition density with respect to μ: there exists a measurable function $\mathcal{L} : \Theta \times \mathsf{W} \to \mathbb{R}_+$ such that for all $\vartheta \in \Theta$ and $w \in \mathsf{W}$,

$$\frac{\mathrm{d}K(\vartheta, \cdot)}{\mathrm{d}\mu}(w) = \mathcal{L}(w|\vartheta) \ .$$

The function $\vartheta \mapsto \mathcal{L}(W|\vartheta)$ is called the likelihood function of the model. Then while in frequentist statistics, the parameter ϑ would be inferred by maximizing the likelihood function, Bayesian statistics consider that the parameter ϑ is itself a sample from a random variable θ, whose the distribution is chosen and called the prior distribution. This law will be denoted by ν_θ. Bayes theorem [Sch95, Theorem 1.31] gives an expression of the conditional law of θ given W depending on the likelihood function and the prior distribution ν_θ. This conditional law admits a density with respect to ν_θ given for ν_θ-almost all ϑ and μ-almost all w by

$$\mathrm{p}_{\theta|W}(w, \vartheta) = \frac{\mathcal{L}(w|\vartheta)}{\mathrm{p}_W(w)} \ , \tag{2.1}$$

where p_W is the marginal density of W with respect to μ given for all $w \in \mathsf{W}$ by

$$\mathrm{p}_W(w) = \int_\Theta \mathcal{L}(w|\vartheta)\nu_\theta(\mathrm{d}\vartheta) \; . \tag{2.2}$$

If w are some observed data, the probability measure associated with density $\mathrm{p}_{\theta|W}(\mathrm{w}, \cdot)$ is called the posterior distribution.

2.2 Monte Carlo algorithms and Markov Chain Monte Carlo

Let π be a probability measure on a measurable space $(\mathsf{E}, \mathcal{E})$ and $f : \mathsf{E} \to \mathbb{R}$ be an integrable function with respect to a target distribution π. We are interested here in estimating the quantity $\int_{\mathsf{E}} f(x)\pi(\mathrm{d}x)$. Classical Monte Carlo methods are based on the strong law of large number for sequences of i.i.d. (independent and identically distributed) random variables with distribution π: let $(Y_i)_{i\in\mathbb{N}}$ be i.i.d. random variables with distribution π, then $\int_{\mathsf{E}} f(x)\pi(\mathrm{d}x)$ is approximated by the sequence of estimator defined for all $N \in \mathbb{N}$ by

$$\hat{f}_N = \frac{1}{N+1}\sum_{i=0}^{N} f(Y_i) \; . \tag{2.3}$$

By the strong law of large numbers $\hat{f}_N$ converges to $\int_{\mathsf{E}} f(x)\pi(\mathrm{d}x)$ almost surely, as N goes to infinity, and if f^2 is integrable with respect to π, then the central limit theorem provides a way to evaluate the asymptotic error. Although this procedure is quite simple, it requires i.i.d. samples with distribution π. Many methods are available for such purpose among which the two most popular ones are the inverse transform sampling and the accept-reject algorithm, see [RC10, Section 2.1.2] and [RC10, Section 2.3].

However as the dimension of the state space gets large, these methods become inefficient. Besides in Bayesian inference, π admits a density with respect to the prior distribution (2.1), given with the notations of Section 2.1 by:

$$x \mapsto \frac{\mathcal{L}(\mathrm{w}|x)}{\mathrm{p}_W(\mathrm{w})} \; ,$$

where w are some observed data. Note that this denisty is known up to a multiplicative constant. Indeed, except if conjugate distributions are used, the marginal density (2.2) has no closed form expression. This constraint make all the more difficult the application of the two methods presented below.

Another class of methods are Markov Chain Monte Carlo methods. The sequence of random variables $(Y_i)_{i\in\mathbb{N}}$ is no longer i.i.d., but $(Y_i)_{i\in\mathbb{N}}$ is a Markov chain with Markov kernel P, with invariant distribution π. An estimator of $\int_{\mathsf{E}} f(x)\mathrm{d}x$ is still defined by (2.3). As in the case of i.i.d. samples, under appropriate assumptions on π and the Markov chain $(Y_i)_{i\in\mathbb{N}}$, a strong law of large numbers can be established [MT09, Chapter 17], which justifies these methods. Indeed, [MT09, Theorem 17.1.7] shows that if P is

Harris recurrent[1] and admits π as invariant distribution, then for all $x \in \mathsf{E}$,

$$N^{-1} \sum_{i=0}^{N-1} f(Y_i) = \int_{\mathsf{E}} f(y) \mathrm{d}y \;, \qquad \mathbb{P}_x\text{-almost surely} \;,$$

where $\mathbb{P}_x$ is the induced law by the Markov kernel P and the initial distribution δ_x on $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ and $(Y_i)_{i \in \mathbb{N}}$ is the canonical chain on $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \mathbb{P}_x)$. A first question regarding these methods is to find a Markov kernel which admits π as invariant distribution. We now present Metropolis-Hastings type algorithms which are generic methods to build such Markov kernels.

2.3 Metropolis-Hastings type samplers

Metropolis-Hastings type algorithms have been first introduced in [Met+53], and then generalized in [Has70] and [Tie98]. Consider a Markov kernel P on $(\mathsf{E}, \mathcal{E})$ of the form for all $x \in \mathsf{E}$ and $\mathtt{A} \in \mathcal{E}$,

$$P(x, \mathtt{A}) = \int_{\mathtt{A}} \alpha(x,y) Q(x, \mathrm{d}y) + \delta_x(\mathtt{A}) \int_{\mathsf{E}} (1 - \alpha(x,y)) Q(x, \mathrm{d}y) \;, \tag{2.4}$$

where Q is a Markov kernel, called the proposal kernel, and $\alpha : \mathsf{E} \times \mathsf{E} \to [0,1]$ is a measurable function, called the acceptance ratio. We can easily sample a Markov chain with Markov kernel P if it is the case for the kernel Q. Let $x \in \mathsf{E}$ and W a random variable distributed according to $Q(x, \cdot)$. Define the random variable Y by

$$Y = \begin{cases} W & \text{with probability } \alpha(x, W) \\ x & \text{otherwise} \;. \end{cases}$$

Then, we easily get that Y is distributed according to $P(x, \cdot)$. Note that if P is of the form (2.4) then by definition, it is reversible with respect to the probability distribution π if and only if for any bounded and measurable function $g : \mathsf{E}^2 \to \mathbb{R}$,

$$\int_{\mathsf{E}^2} g(x,y) \pi(x) P(\mathrm{d}x, \mathrm{d}y) = \int_{\mathsf{E}^2} g(x,y) \pi(\mathrm{d}y) P(y, \mathrm{d}x) \;,$$

which is equivalent to

$$\int_{\mathsf{E}^2} g(x,y) \alpha(x,y) \pi(\mathrm{d}x) Q(x, \mathrm{d}y) = \int_{\mathsf{E}^2} g(x,y) \alpha(y,x) \pi(\mathrm{d}y) Q(y, \mathrm{d}x) \;.$$

Define the two measures on $(\mathsf{E}^2, \mathcal{E}^{\otimes 2})$, μ and μ^{T} for all $\mathtt{A} \in \mathcal{E}^{\otimes 2}$ by[2]:

$$\mu(\mathtt{A}) = \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\mathtt{A}}(y,z) \pi(\mathrm{d}y) Q(y, \mathrm{d}z)$$
$$\mu^{\mathrm{T}}(\mathtt{A}) = \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\mathtt{A}}(z,y) \pi(\mathrm{d}y) Q(y, \mathrm{d}z) \;.$$

[1] see Definition A.14
[2] see also Definition-Proposition A.1

Then, P is reversible with respect to π if there exists a symmetric set $\mathsf{S} \in \mathcal{E}^{\otimes 2}$ such that α is zero on S^c, $\mu(\cdot \cap \mathsf{S})$ is absolutely continuous with respect to $\mu^{\mathrm{T}}(\cdot \cap \mathsf{S})$ with density m and μ-almost everywhere

$$\alpha(x,y)m(x,y) = \alpha(y,x) \ .$$

Two significant examples which we deal with in this manuscript are the following.

(A) Assume that there is a measure ν on $(\mathsf{E}, \mathcal{E})$ which dominated π and Q, i.e. π admits a density still denoted by π with respect to ν and Q admits a kernel density denoted by q with respect to ν. Then if we define,

$$\mathsf{S} = \left\{ (x,y) \in \mathsf{E}^2 \,\middle|\, \pi(x)q(x,y) > 0 \ , \pi(y)q(y,x) > 0 \right\} \ ,$$

and

$$\alpha(x,y) = \begin{cases} \min\left(1, \dfrac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) & \text{pour } (x,y) \in \mathsf{S} \\ 0 & \text{otherwise} \ , \end{cases} \tag{2.5}$$

the conditions for P to be reversible are satisfied with

$$m(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) \ ,$$

defined on S.

A simple choice for Q is the dominating measure ν, if it is a probability measure. In practice, we need to be able to sample from ν because the proposed samples are i.i.d. random variables with distribution ν. In this case, note that the acceptance ratio is of the form,

$$\alpha(x,y) = \min\left(1, \pi(y)/\pi(x)\right) \ . \tag{2.6}$$

This algorithm is called the Metropolis independent sampler. Another possibility to define a Metropolis-Hastings kernel is, when E is a group and ν is a translation invariant measure invariant, to consider the random walk associated with a symmetric probability measure on $(\mathsf{E}, \mathcal{E})$ which admits a density $\tilde{q}$ with respect to ν. In this case, the random variables proposed in the algorithm are of the form

$$W_{k+1} = Y_k + Z_{k+1} \ ,$$

where Y_k is the current state of the Markov chain at iteration k and $(Z_i)_{i\geq 1}$ are i.i.d. random variables with density $\tilde{q}$ with respect to ν. In addition, the proposal transition density is given for all $x, y \in \mathsf{E}$ by $q(x,y) = \tilde{q}(x-y)$, since $\tilde{q}$ is assumed to be symmetric, and the acceptance ratio is of the form (2.6) again. This method is called the symmetric random walk Metropolis algorithm. An example of such algorithm is the symmetric random walk Metropolis algorithm defined on $\mathbb{R}^d$ with the Lebesgue measure and where the proposal kernel is the Markov kernel associated with the symmetric random walk with zero-mean Gaussian increments and covariance matrix $\varsigma^1 \,\mathrm{Id}$, $\varsigma^2 > 0$.

(B) It can be observed that it is essential for the previous algorithms that π and Q are dominated by a common measure. We can weaken this condition when π admits a positive density, still denote by π, with respect to a probability measure ν for which Q is reversible. Then setting $\mathsf{S} = \mathsf{E}$ and for all $(x, y) \in \mathsf{E}^2$,

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right) ,$$

(2.4) defines a Metropolis-Hastings kernel P reversible with respect to π.

2.4 The overdamped Langevin diffusion

2.4.1 Continuous Markovian dynamics

Instead of considering discrete dynamics, we can think of using continuous Markovian processes associated with a semi-group $(\mathbf{P}_t)_{t\geq 0}$ on $(\mathsf{E}, \mathcal{E})$ and for which π is invariant. For this, assume that E is a locally compact Polish space and that $(\mathbf{P}_t)_{t\geq 0}$ is Feller[3]. The Markov semi-group $(\mathbf{P}_t)_{t\geq 0}$ is associated with an operator, called its generator and denoted by $\mathscr{A}$, which is defined as follows. Denote by $C_0(\mathsf{E})$ the set of functions vanishing at infinity[4]. The domain of definition of $\mathscr{A}$, denoted by $\mathscr{D}(\mathscr{A})$ is the set of all functions $h \in C_0(\mathsf{E})$ for which there exists a function $g_h \in C_0(\mathsf{E})$ such that for all $x \in \mathsf{E}$,

$$g_h(x) = \lim_{t\to 0} t^{-1}\left\{\mathbf{P}_t h(x) - h(x)\right\} .$$

Then define for all $h \in \mathscr{D}(\mathscr{A})$, $\mathscr{A}h = g_h$. The study of the properties of the generator associated with a Markov semi-group allows to deduce a lot of information on this semi-group. Let $\mathbb{A}$ be an algebra included in $\mathscr{D}(\mathscr{A})$ and dense in $C_0(\mathsf{E})$. In particular if $\mathsf{E} = \mathbb{R}^d$, we can take for $\mathbb{A}$, the set of k-times differentiable functions with compact support from $\mathbb{R}^d$ to $\mathbb{R}$, for $k \in \mathbb{N} \cup \{\infty\}$. By [RY99, Proposition 1.5, Chapter VII] and [EK86, Theorem 9.17, Chapter 3], if for all $h \in \mathbb{A}$, $\int_{\mathsf{E}} \mathscr{A}h(x)\mathrm{d}\pi(x) = 0$, then π is invariant for $(\mathbf{P}_t)_{t\geq 0}$. In addition if π is an invariant distribution for $(\mathbf{P}_t)_{t\geq 0}$ and that $(\mathbf{P}_t)_{t\geq 0}$ is Harris recurrent[5], then a strong law of large numbers holds [RY99, Theorem 3.12, Chapter X]: for all $x \in \mathsf{E}$

$$\lim_{T\to+\infty} T^{-1}\int_0^T f(\mathbf{Y}_s)\mathrm{d}s = \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x) , \qquad \mathbb{P}_x\text{-almost surely} ,$$

where $(\mathbf{Y}_t)_{t\geq 0}$ is the Markovian canonical process associated with $(\mathbf{P}_t)_{t\geq 0}$ and $\mathbb{P}_x$ is the probability measure induced by the semi-group and the initial distribution δ_x[6]. We consider here Markov semi-groups associated with solutions of homogeneous stochastic differential equation (SDE) on $\mathbb{R}^d$.

[3] see Definition A.25
[4] see Definition A.24
[5] see Definition A.30
[6] see Theorem A.26

Let $b : \mathbb{R}^d \to \mathbb{R}$ and σ be a function from $\mathbb{R}^d$ to $M_{d,m}(\mathbb{R})$, the set of matrices of dimension $d \times m$, such that for all $x \in \mathbb{R}^d$, $\sigma(x)\sigma(x)^{\mathrm{T}}$ is definite positive. Assume that these two functions are locally Lipschitz and consider the SDE:

$$\mathrm{d}\mathbf{Y}_t = b(\mathbf{Y}_t)\mathrm{d}t + \sigma(\mathbf{Y}_t)\mathrm{d}B_t^m \ .$$

where (B_t^m) is a m-dimensional standard Brownian motion. By [IW89, Theorem 2.3, Theorem 3.1, Chapter 4], for all initial condition $\mathbf{Y}_0 = x \in \mathbb{R}^d$, this SDE admits a unique solution $(\mathbf{Y}_t)_{t\in 0}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ endowed with a Brownian motion $(B_t^d)_{t\geq 0}$ until an explosion time ξ which is the stopping time defined by

$$\xi = \inf \{ t \geq 0 \,|\, \mathbf{Y}_t = \infty \} \ .$$

Assume that almost surely $\xi = +\infty$. Then by [SV79, Corollary 10.1.5] and [IW89, Theorem 6.1, Chapter 4], the distribution of the process $(\mathbf{Y}_t)_{t\in 0}$ defines a Feller Markov semi-group $(\mathbf{P}_t)_{t\geq 0}$ by for all $\mathtt{A} \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, $\mathbf{P}_t(x, \mathtt{A}) = \mathbb{P}_x[\mathbf{Y}_t \in \mathtt{A}]$. Besides, [Bha78, Lemma 2.4] shows that this semi-group is irreducible[7] for the Lebesgue measure. Finally, [Bha78, Theorem 3.3] gives a criteria for the diffusion to be Harris recurrent.

When $\mathsf{E} = \mathbb{R}^d$ and π admits a positive density with respect to the Lebesgue measure of the form

$$\pi(x) = \mathrm{e}^{-U(x)} \Big/ \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)} \mathrm{d}x \ \ ,$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a continuously differentiable function, the overdamped Langevin diffusion defined by

$$\mathrm{d}\mathbf{Y}_t^{\mathrm{L}} = -\nabla U(\mathbf{Y}_t^{\mathrm{L}}) + \sqrt{2}\mathrm{d}B_t^d \ , \tag{2.7}$$

admits π as invariant distribution. Its generator $\mathscr{A}^{\mathrm{L}}$ is given for all function $h \in C_c^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$\mathscr{A}^{\mathrm{L}} h(x) = -\langle \nabla U(x), \nabla h(x) \rangle + \Delta h(x) \ .$$

By [MT93b, Theorem 2.1], this process is non explosive if for all $x \in \mathbb{R}^d$,

$$\langle \nabla U(x), x \rangle \geq -a_1 \left\| x \right\|^2 - a_2 \ ,$$

for some constants $a_1, a_2 \in \mathbb{R}_+$. By a simple change of variable, for any function $h \in C_c^2(\mathbb{R}^d)$, we have

$$\int_{\mathbb{R}^d} \left\{ \mathscr{A}^{\mathrm{L}} h \right\} (x) \pi(x) \mathrm{d}x = 0 \ ,$$

which shows that π is invariant for the Markov semi-group $(\mathbf{P}_t^{\mathrm{L}})_{t\geq 0}$ associated with (2.7). Therefore, by [MT09, Proposition 10.1.1] and Corollary A.34, $(\mathbf{P}_t^{\mathrm{L}})_{t\geq 0}$ is Harris recurrent.

With a very few exceptions, it does not exist simple methods to sample a solution of the overdamped Langevin equation. Whereas some exact simulation algorithms have been proposed, see for example [**beskos:roberts:2005**], their implementation seems to be very costly.

[7] see Definition A.28

2.4.2 Two MCMC algorithms based on the overdamped Langevin equation: the ULA and the MALA

We consider in this thesis the Euler-Maruyama discretization associated with (2.7) and defined for a given initial condition by: for all $k \geq 0$,

$$Y_{k+1} = Y_k - \gamma \nabla U(Y_k) + \sqrt{2\gamma} Z_{k+1} , \tag{2.8}$$

where $\gamma > 0$ is the step size of the discretization and $(Z_k)_{k \geq 1}$ is a sequence of i.i.d. d dimensional standard normal random variables. The discretization $(Y_n)_{n \in \mathbb{N}}$ can be seen as a approximate path of $(\mathbf{Y}_t)_{t \geq 0}$ and used to sample from π. This algorithm has been first proposed by [Erm75] and [Par81] for molecular dynamics applications. Then it has been popularized in machine learning by [Gre83], [GM94] and computational statistics by [Nea93] and [RT96a]. As in [RT96a], this algorithm will be called in this manuscript the Unadjusted Langevin Algorithm.

The drawback of this method is that even if the Markov chain $(Y_n)_{n \in \mathbb{N}}$ has a unique stationary distribution π_γ and is ergodic (which is guaranteed under mild assumptions on U), π_γ is most of the time different from π. Therefore, using this method introduces a bias in the computation of $\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x)$, i.e.

$$\lim_{N \to +\infty} N^{-1} \sum_{k=0}^{N-1} f(Y_k) = \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi_\gamma(x) \neq \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x) .$$

However, [TT90] shows that under appropriate conditions on f, the chain $(Y_n)_{n \in \mathbb{N}}$ and the diffusion $(\mathbf{Y}_t)_{t \geq 0}$, there exists a constant C which depends on f and π such that for all $\gamma > 0$ in a neighborhood of 0,

$$\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi_\gamma(x) - \int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x) = C\gamma + \mathcal{O}(\gamma^2) .$$

Therefore, taking a step size sufficiently small implies that the error in the computation of $\int_{\mathbb{R}^d} f(x) \mathrm{d}\pi(x)$ is small as well. To suppress the bias of the method, it has been proposed by [RDF78] and [RT96a] to use the Markov kernel defined by the Euler-Maruyama discretization (2.8) as a proposal kernel in a Metropolis-Hastings algorithm. Following [RT96a], this algorithm will be referred to as the Metropolis Adjusted Langevin Algorithm (MALA).

Another method to suppress the bias of ULA is to use a sequence of non-increasing step sizes $(\gamma_k)_{k \geq 1}$ satisfying $\lim k \to +\infty \gamma_k = 0$ and $\sum_{k=1}^{+\infty} \gamma_k = +\infty$. Then, we define the inhomogeneous Markov chain $(Y_n)_{n \in \mathbb{N}}$ associated with this sequence, for a given initial condition and all $k \geq 0$ by

$$Y_{k+1} = Y_k - \gamma_{k+1} \nabla U(Y_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} ,$$

where $(Z_k)_{k\geq 1}$ is a sequence of i.i.d.d-dimensional standard normal random variables. It has been proved in [LP02, Theorem 6] that under appropriate assumptions on f and U, for all $x \in \mathbb{R}^d$,

$$\lim_{N\to+\infty} \frac{\sum_{k=0}^{N-1} \gamma_{k+1} f(Y_k)}{\sum_{k=1}^{N} \gamma_k} = \int_{\mathbb{R}^d} f(x)\mathrm{d}\pi(x) , \qquad \mathbb{P}_x\text{-almost surely} .$$

In the sequel, this algorithm is still referred to as the ULA algorithm.

2.5 Convergence of MCMC methods

Let $(Y_k)_{k\in\mathbb{N}}$ be a homogeneous Markov chain, associated with the initial distribution μ_0 and the Markov kernel P, on a Polish space E endowed with a distance $\mathbf{d}$ and its Borel σ-field $\mathcal{E}$. Define the sequence of marginal laws $(\mu_0 P^k)_{k\in\mathbb{N}^*}$ of the chain $(Y_k)_{k\in\mathbb{N}}$ by induction for all $k \in \mathbb{N}^*$ and $\mathtt{A} \in \mathsf{E}$ by:

$$\mu_0 P^k(\mathtt{A}) = \int_{\mathsf{E}} \mathbb{1}_{\mathtt{A}}(y)\mu_0 P^{k-1}(\mathrm{d}x)P(x, \mathrm{d}y) .$$

We are interested in this section in the existence and especially in the uniqueness of an invariant probability measure for P. In addition, some results on the convergence of the sequence of probability measures $(\mu_0 P^k)_{k\in\mathbb{N}^*}$ to the unique stationary distribution of P, when it exists, are given.

Some methods have been presented in Section 2.3 and Section 2.4.2 to build some Markov chains for which π is an invariant distribution or is very close to be. The convergence analysis of the marginal laws of the associated Markov chains is justified by the study of the convergence of the estimator $\hat{f}_N$, defined by (1.3) to $\int_{\mathbb{R}^d} f(x)\mathrm{d}\pi(x)$ as N goes to infinity. Indeed, as $(Y_k)_{k\in\mathbb{N}}$ is not i.i.d., a first step is to measure the bias of the estimation given by

$$\left| \mathbb{E}_x \left[\hat{f}_N \right] - \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x) \right| = \left| \frac{1}{N+1} \sum_{i=0}^{N} \left\{ \mathbb{E}_x \left[f(Y_i) \right] - \int_{\mathsf{E}} f(x)\mathrm{d}\pi(x) \right\} \right| .$$

The convergence of the marginal laws $(\mu_0 P^k)_{k\in\mathbb{N}}$ to π has been the subject of numerous studies [MT09], [Num84], [HMS11]. This convergence is established under different distance on the set of probability measures on $(\mathsf{E}, \mathcal{E})$, denoted by $\mathcal{P}(\mathsf{E})$. We consider in this manuscript two kinds of distances: V-total variations distances and Wasserstein distances.

2.5.1 Convergence of Markov chains

Distances on the set of probability measures

Let $V : \mathsf{E} \to [1, \infty)$ be a measurable function. We define the V-norm of a measurable function $h : \mathsf{E} \to \mathbb{R}$ by

$$\|h\|_V = \sup_{x\in\mathsf{E}} |h(x)|/V(x) .$$

Let μ be a bounded signed measure on $(\mathsf{E}, \mathcal{E})$. The V-total variation of μ is given by

$$\|\mu\|_V = (1/2) \sup_{\|h\|_V \le 1} \left| \int_{\mathbb{R}^d} h(x) \mathrm{d}\mu(x) \right| .$$

If $V \equiv 1$, then $\|\cdot\|_V$ is the total variation of μ and is denoted by $\|\cdot\|_{\mathrm{TV}}$. For two measures of probability $\mu, \nu \in \mathcal{P}(\mathsf{E})$, the V-total variation distances between μ and ν is the V-total variation of the measure $\mu - \nu$. Similarly, the total variation distance between μ and ν is the total variation of the measure $\mu - \nu$. [DMS14, Proposition 6.16] shows that the set of probability measure $\{\mu \in \mathcal{P}(\mathsf{E}) \,|\, V \in \mathrm{L}^1(\mu)\}$ is a Banach space if it is endowed with the V-total variation distance. Another kind of distance we consider, is the Wasserstein distance associated with the distance $\mathbf{d}$ on E. Let $h : \mathsf{E} \to \mathbb{R}$ be a Lipschitz function i.e. there exists $C \ge 0$ such that for all $x, y \in \mathsf{E}$, $|h(x) - h(y)| \le C\mathbf{d}(x, y)$. Let $h : \mathsf{E} \to \mathbb{R}$ be a Lipschitz function and denote by

$$\|h\|_{\mathrm{Lip}} = \sup_{x,y \in \mathsf{E}} \left\{ \frac{|h(x) - h(y)|}{\mathbf{d}(x, y)} \right\} .$$

Define the set $\mathcal{P}_1(\mathsf{E})$ of probability measure on $(\mathsf{E}, \mathcal{E})$ by

$$\mathcal{P}_1(\mathsf{E}) = \{\mu \in \mathcal{P}(\mathsf{E}) \,|\, \int_{\mathsf{E}} \mathbf{d}(x, x_0) \mathrm{d}\mu(x) < +\infty\} ,$$

for a fixed element $x_0 \in \mathsf{E}$. The Wasserstein distance is defined for all $\mu, \nu \in \mathcal{P}_1(\mathsf{E})$ by

$$W_{\mathbf{d}}(\mu, \nu) = \sup_{\|h\|_{\mathrm{Lip}} \le 1} \left| \int_{\mathsf{E}} h(x) \mathrm{d}\mu(x) - \int_{\mathsf{E}} h(x) \mathrm{d}\nu(x) \right| .$$

By [Vil09, Theorems 6.8 and 6.16], $\mathcal{P}_1(\mathsf{E})$ is a Polish space if it is endowed with $W_{\mathbf{d}}$.

The Wasserstein distance and the total variation distance can not be comparable in general, except when $\mathbf{d}$ is bounded. In such a case, for all $\mu, \nu \in \mathcal{P}(\mathsf{E})$,

$$W_{\mathbf{d}}(\mu, \nu) \le \sup_{x,y \in \mathsf{E}^2} \{\mathbf{d}(x, y)\} \|\mu - \nu\|_{\mathrm{TV}} .$$

Note that the convergence in one of these distances implies the weak convergence.

By the Monge-Kantorovich Theorem [Vil09, Theorem 5.10], the Wasserstein distance and the total variation distance between two probability measures μ and ν on $(\mathsf{E}, \mathcal{E})$, have dual forms in terms of couplings between these μ and ν. A probability measure $\zeta \in \mathcal{P}(\mathsf{E} \times \mathsf{E})$ is a transference plan between μ and ν if its first marginal $\zeta(\cdot \times \mathsf{E})$ is equal to μ and its second $\zeta(\mathsf{E} \times \cdot)$ is equal to ν. The set of all transference plan between μ and ν is denoted by $\Pi(\mu, \nu)$. A coupling between μ and ν is any couple of random variables (X, Y) with distribution $\zeta \in \Pi(\mu, \nu)$. The total variation distance can be written of the form: for all $\mu, \nu \in \mathcal{P}(\mathsf{E})$

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf_{\zeta \in \Pi(\mu, \nu)} \int_{\mathsf{E} \times \mathsf{E}} \mathbb{1}_{\Delta_{\mathsf{E}}}(x, y) \zeta(\mathrm{d}x, \mathrm{d}y) = \inf_{(\mathrm{X}, \mathrm{Y}) \in \tilde{\Pi}(\mu, \nu)} \mathbb{P}(\mathrm{X} \neq \mathrm{Y}) ,$$

where $\Delta_{\mathsf{E}} = \{(x, y) \in \mathsf{E} \times \mathsf{E} \,|\, x = y\}$. As regards to the Wasserstein, it can be written of the form: for all $\mu, \nu \in \mathcal{P}(\mathsf{E})$

$$W_{\mathbf{d}}(\mu, \nu) = \inf_{\zeta \in \Pi(\mu,\nu)} \int_{\mathsf{E}\times\mathsf{E}} \mathbf{d}(x, y)\zeta(\mathrm{d}x, \mathrm{d}y) = \inf_{(\mathrm{X},\mathrm{Y}) \in \tilde{\Pi}(\mu,\nu)} \mathbb{E}[\mathbf{d}(\mathrm{X}, \mathrm{Y})] \;.$$

In addition, the infimum in the two equations is reached.

Convergence in V-total variation distances

Historically, the analysis of chains began with the study of chains which have an accessible atom. A measurable set $\boldsymbol{\alpha} \in \mathcal{E}$ is an atom for the Markov kernel, P on $(\mathsf{E}, \mathcal{E})$ if there exists a probability measure ν on $(\mathsf{E}, \mathcal{E})$ such that for all $x \in \boldsymbol{\alpha}$, $P(x, \cdot) = \nu$. Indeed in this case, the existence of an invariant measure and the analysis of the convergence to it, boil down to the study of the return time to this atom, see [MT09, section 10.2, 13.2, 14.1, 15.1].

Consider $(Y_k)_{k\in\mathbb{N}}$ and $(\mathcal{F}_k)_{k\in\mathbb{N}}$ the canonical chain and filtration associated with P. For all $\mathtt{A} \in \mathcal{E}$, the return time of the chain to $\mathtt{A}$ is defined by:

$$\boldsymbol{\sigma}_{\mathtt{A}} = \inf \{ k \in \mathbb{N}^* \,|\, Y_k \in \mathtt{A} \} \;.$$

The successive return times to $\mathtt{A}$ are the sequence of random variables $(\boldsymbol{\sigma}_{\mathtt{A}}^{(m)})_{m\in\mathbb{N}^*}$ defined recursively by for $m = 1$, $\boldsymbol{\sigma}_{\mathtt{A}}^{(m)} = \boldsymbol{\sigma}_{\mathtt{A}}$ and for $m \geq 2$,

$$\boldsymbol{\sigma}_{\mathtt{A}}^{(m)} = \inf \{ k \in \mathbb{N}^* \,|\, Y_{\boldsymbol{\sigma}_{\mathtt{A}}+k} \in \mathtt{A} \} \;.$$

Note that for all $\mathtt{A} \in \mathcal{E}$ and $m \in \mathbb{N}^*$, $\boldsymbol{\sigma}_{\mathtt{A}}^{(m)}$ is a $(\mathcal{F}_k)_{k\in\mathbb{N}}$-stopping time.

Except if the state space E is discrete, the existence of an atom is a very strong condition which is rarely satisfied. A weaker condition is the existence of a small set for P. Let $\mathtt{n} \in \mathbb{N}^*$. The set $\mathtt{C} \in \mathcal{E}$ is $(\mathtt{n})$-small for P if there exists a σ-finite non trivial measure ν on $(\mathsf{E}, \mathcal{E})$ such that for all $x \in \mathtt{C}$, $P^{\mathtt{n}}(x, \dots) \geq \nu(\cdot)$. If $\mathtt{C} \in \mathcal{E}$ is 1-small for P, then the splitting technique from [Num78] allows to build a Markov kernel $\check{P}$ on a extended state space $(\mathsf{E} \times \{0, 1\}, \mathcal{E} \otimes \mathcal{B}(\{0, 1\}))$, which admits $\mathtt{C} \times \{1\}$ as an atom and has P for marginal, i.e. for all $x \in \mathsf{E}$, $a \in \{0, 1\}$ and $\mathtt{A} \in \mathcal{E}$,

$$\check{P}((x, a), \mathtt{A} \times \{0, 1\}) = P(x, \mathtt{A}) \;.$$

In addition, the return time to the atom $\mathtt{C} \times \{1\}$ of $\check{P}$ are closely related to the successive return time of P to the small set $\mathtt{C}$, and the analysis of P is made through the renewal process defined by these stopping times. If $\mathtt{C} \subset \mathcal{E}$ is just petite [8], it is 1-small for a sample kernel associated with P and the splitting technique can still be applied. To summarize, the study of irreducible chains which have a petite set $\mathtt{C} \in \mathcal{E}$, boils down to the analysis of successive return times to this set and using the results for atomic chains.

Bounds on moments of successive return times to a measurable set $\mathtt{C} \in \mathcal{E}$ can be obtained from Foster-Lyapunov drift inequalities.

[8] see Definition A.8

(1) The first drift condition which is presented is the following: there exist a proper measurable function $V : \mathsf{E} \to \bar{\mathbb{R}}_+$, $a \in \mathbb{R}_+$ such that

$$PV \leq V - 1 + a\,\mathbb{1}_{\mathsf{C}}\ , \qquad \sup_{x\in\mathsf{C}} V(x) < +\infty\ . \tag{2.9}$$

By Dynkin's theorem[9], for all $x \in \mathsf{E}$,

$$\mathbb{E}_x\left[\boldsymbol{\sigma}_{\mathsf{C}}\right] \leq V(x) + a\mathbb{1}_{\mathsf{C}}(x)\ , \qquad \sup_{x\in\mathsf{C}} \mathbb{E}_x\left[\boldsymbol{\sigma}_{\mathsf{C}}\right] < +\infty\ . \tag{2.10}$$

Furthermore, [MT09, p. 13.0.1] shows that if C is a petite set and P is irreducible, aperiodic and Harris recurrent, then (2.10) is equivalent to the existence of a unique invariant probability measure π for P and for all $x \in \mathbb{R}^d$,

$$\lim_{n\to+\infty} \|\delta_x P^n - \pi\|_{\mathrm{TV}} = 0\ .$$

If P is not Harris recurrent but only irreducible, aperiodic and recurrent, then by [MT09, Theorem 9.1.5][10], P has still a unique invariant probability measure π and in addition there exists a set $\mathsf{N} \in \mathcal{E}$ such that $\pi(\mathsf{N}) = 0$ and for all $x \in \mathsf{N}^{\mathrm{c}}$,

$$\lim_{n\to+\infty} \|\delta_x P^n - \pi\|_{\mathrm{TV}} = 0\ .$$

(2) The second drift inequality presented below implies a convergence in potentially stronger distance than the total variation distance. Assume that there exist a measurable proper function $V : \mathsf{E} \to \bar{\mathbb{R}}_+$, a measurable function $\mathrm{f} : \mathsf{E} \to [1, +\infty)$ and $a \in \mathbb{R}_+$ such that

$$PV \leq V - \mathrm{f} + a\,\mathbb{1}_{\mathsf{C}}\ , \qquad \sup_{x\in\mathsf{C}} V(x) < +\infty\ .$$

By Dynkin's Theorem, we have

$$\mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \mathrm{f}(Y_k)\right] \leq V(x) + a\mathbb{1}_{\mathsf{C}}(x)\ , \qquad \sup_{x\in\mathsf{C}} \mathbb{E}_x\left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \mathrm{f}(Y_k)\right] < +\infty\ . \tag{2.11}$$

[MT09, Theorem 14.0.1] shows that if C is a petite set, P is irreducible, aperiodic and recurrent, (2.11) implies that P admits a unique invariant probability measure π and for all t $x \in \{V < +\infty\}$,

$$\lim_{n\to+\infty} \|\delta_x P^n - \pi\|_{\mathrm{f}} = 0\ .$$

(3) We now present a drift condition which implies a geometric rate of convergence of $(\delta_x P^n)_{n\in\mathbb{N}}$ to π in V-total variation. Assume that there exist a proper measurable function $V : \mathsf{E} \to [1, +\infty]$, $\lambda \in [0, 1)$ and $a \in \mathbb{R}_+$ such that

$$PV \leq \lambda V + a\,\mathbb{1}_{\mathsf{C}}\ , \qquad \sup_{x\in\mathsf{C}} V(x) < +\infty\ . \tag{2.12}$$

[9] see Corollary A.20
[10] see Theorem A.15

Dynkin's theorem implies that for all $\kappa \in (1, \lambda^{-1})$, there exists $C \geq 0$ such that

$$\mathbb{E}_x \left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \kappa^k V(Y_k) \right] \leq C(V(x) + a\mathbb{1}_{\mathsf{C}}(x)) , \qquad \sup_{x \in \mathsf{C}} \mathbb{E}_x \left[\sum_{k=0}^{\boldsymbol{\sigma}_{\mathsf{C}}-1} \kappa^k V(Y_k) \right] < +\infty . \tag{2.13}$$

[MT09, Theorem 15.4.1] shows that if P is irreducible, aperiodic and (2.13) is satisfied for a petite set C, then there exists $r > 1$ and a constant $C \geq 1$ such that for all $x \in \{V < +\infty\}$ and $n \geq 1$,

$$\|\delta_x P^n - \pi\|_V \leq CV(x) r^n .$$

Finally, we mention that there exist two drift conditions proposed in [TT94] and [Dou+04] respectively, to get sub-geometric convergence rates in total variation distance and for some V-total variation distances. Roughly speaking, these two conditions imply sub-geometric moments for the return times to a set and therefore yields to sub-geometric convergence when this set is petite.

2.5.2 Application to the convergence of MCMC methods

We now deal with convergence of Metropolis-Hastings type kernels on $(\mathsf{E}, \mathcal{E})$, as π admits a density with respect to a dominating measure ν, and the proposal kernel admits a transition density with respect to ν as well, see Section 2.3-(A).

Denote by P_{MH} a Metropolis-Hastings kernel satisfying these conditions and defined by (2.4)-(2.5). A significant result is [Tie94, Corollary 2], which shows that if P_{MH} is π-irreducible, then it is Harris recurrent. Moreover, to check that P_{MH} is π-irreducible, a very simple condition is that for all $x \in \mathsf{E}$, if $\pi(x) > 0$ then $q(y, x) > 0$ for all $y \in \mathsf{E}$, [MT96, Lemma 1.1]. In the case $\mathsf{E} = \mathbb{R}^d$ for $d \geq 1$ and ν is the Lebesgue measure, this condition is weakened by [RT96b, Theorem 2.2], which establishes that if π is positive, bounded on $\mathbb{R}^d$, and there exist $\delta_q, \epsilon_q > 0$ such that

$$q(x, y) \geq \epsilon_q \text{ pour tout } x, y \in \mathbb{R}^d,\ \|x - y\| \leq \delta_q , \tag{2.14}$$

then P is irreducible with respect to the Lebesgue measure and consequently is π-irreducible. In addition, this result shows that under the same conditions, P is strongly aperiodic and any non-empty compact set is small. Therefore the analysis of convergences rates for Metropololis-Hastings type kernels in the dominated case is divided in two categories.

If the kernel satisfies a uniform minorization conditions, then it is uniformly ergodic. For instance, for the Metropolis independent sampler (see Section 2.3-(A)), [MT96, Theorem 2.1] shows that if there exists $\beta_q > 0$ such that for all $x \in \mathsf{E}$, $q(x)/\pi(x) > \beta_q$, then P_{MH} is uniformly ergodic: for all $n \geq 1$,

$$\sup_{x \in \mathsf{E}} \|P_{\mathrm{MH}}(x) - \pi\|_{\mathrm{TV}} \leq (1 - \beta_q)^n .$$

If the kernel is not uniformly ergodic, then a common approach is to establish a geometric or sub-geometric drift condition. It is the case for the symmetric random walk Metropolis algorithm, as $\mathsf{E} = \mathbb{R}^d$ and π admits a positive, continuous density with respect to the Lebesgue measure, still denoted by π. In this context, some conditions on the geometry of the level sets of π have been proposed in [RT96b] and [JH00], which imply a geometric drift condition of the form (2.12). In particular, these conditions are satisfied for densities of the following form: there exist $\ell \in \mathbb{N}^*$, a positive homogeneous polynomial $\mathrm{p} : \mathbb{R}^d \to \mathbb{R}_+^*$ of degree ℓ, a polynomial $\mathrm{q} : \mathbb{R}^d \to \mathbb{R}$ of degree strictly smaller than ℓ, and a positive polynomial $\mathrm{r} : \mathbb{R}^d \to \mathbb{R}_+^*$ such that for all $x \in \mathbb{R}^d$,

$$\pi(x) \propto \mathrm{r}(x) \exp(-\mathrm{p}(x) - \mathrm{q}(x)) \ . \tag{2.15}$$

Then, if the proposal density satisfies (2.14), the symmetric random walk Metropolis is geometrically ergodic for densities of the form (2.15). Note that for heavy tail distributions, some drift conditions which imply sub-geometric convergence to π can be established, see [FM00], [FM03] and [JR07].

For applications to MCMC, having explicit bounds on the convergence of the associated Markov kernels can be important. Derivation of explicit bounds in total variation has been the object of numerous results. Most of these works rely on a drift condition again, and a minorization condition by the existence of a small set, see [Ros95], [RT99], or [Bax05]. These bounds are then established using either the splitting technique of Nummelin introduced in Section 2.5.1 or coupling techniques. However, it has been observed in [JH01] that the bounds derived from these results can not be used for a detailed analysis of the convergence.

2.5.3 Analysis of ULA and MALA

Regarding the algorithms ULA and MALA presented in Section 2.4.2, [RT96a] studied the convergence of the produced Markov chains for both algorithms at fixed step size $\gamma > 0$. However, no result is established concerning the convergence of ULA to the target density π, only to its stationary distribution which, as mentioned earlier, is in general different from π. On the other hand, note that since the Markov chain produced by ULA with fixed step sizes does not have the correct stationary distribution, other techniques than the one presented in Section 2.5.1 have to be used, as it is done in [Dal16].

Concerning the MALA algorithm, [RT96a] shows that the produced Markov kernel is geometrically ergodic under the condition that

$$\lim_{\|x\| \to +\infty} \int_{\mathtt{A}(x)} q(x,y) \mathrm{d}y = 0 \ ,$$

where q is the transition density associated with MALA,

$$\mathtt{A}(x) = (\mathtt{B}(x) \cup \mathtt{C}(x)) \setminus (\mathtt{B}(x) \cap \mathtt{C}(x)) \ ,$$

and

$$\mathtt{B}(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, \pi(x) q(x,y) \leq \pi(y) q(y,x) \right\} \ , \quad \mathtt{C}(x) = \left\{ y \in \mathbb{R}^d \,\middle|\, \|y\| \leq \|x\| \right\} \ .$$

However this condition is difficult to check in practice.

One of the purpose the work presented in this manuscript is to provide explicit bounds for the convergence of MCMC methods, in total variation and Wasserstein distance. In particular, for particular models of target distributions, we have analysed the dependence of the convergence on the dimension of the state space.

Besides, the results introduced in Section 2.5.1 assume that the chain is irreducible, which is barely true in infinity dimension. We also address this problem in a part of this manuscript.

2.6 Optimal scaling for Metropolis-Hastings type algorithms

Another approach to study the Metropolis-Hastings type algorithms in a high dimensional setting is the optimal scaling of these methods.

Consider a target distribution π on $\mathbb{R}^d$. If we want to apply the symmetric random walk Metropolis algorithm or MALA, we can observe that a parameter need to be chosen. For the symmetric random walk Metropolis, it is the size of the increments and for the MALA algorithm, it is the step size of the disretization. Optimal scaling results aim to find the best possible choice for this parameter (in some sense) and its dependency on the dimension d. We give in the two next sections a presentation of these results.

2.6.1 Optimal scaling of the symmetric random walk Metropolis algorithm

The first optimal scaling result for Metropolis-Hastings algorithms has been presented in the pioneer work of [RGG97] and concerns the symmetric random walk Metropolis algorithm on $\mathbb{R}^d$. Let π^d be a positive target density on $\mathbb{R}^d$ with respect to the Lebesgue measure ($\pi^d(x) > 0$ for all $x \in \mathbb{R}^d$). Recall that the symmetric random walk Metropolis algorithm with zero-mean Gaussian increments defines a Markov chain $(Y_k^d)_{k\in\mathbb{N}}$ as follows: for a given initial condition Y_0^d , for all $k \geq 0$,

$$Y_{k+1}^d = \begin{cases} Y_k^d + \sigma_d Z_{k+1}^d & \text{with probability } \alpha\left(Y_k^d, Y_k^d + \sigma_d Z_{k+1}^d\right) \\ Y_k^d & \text{otherwise} \ , \end{cases} \tag{2.16}$$

where $\sigma_d \in \mathbb{R}_+^*$, $(Z_i^d)_{i\geq 1}$ is an sequence of i.i.d. d-dimensional standard normal random variables and for all $x, y \in \mathbb{R}^d$

$$\alpha^d(x, y) = \min\left(1, \pi^d(y)/\pi^d(x)\right) \ . \tag{2.17}$$

Observe that the choice of the parameter σ_d is left to the user. Then, we can wonder if there exists an optimal choice for this parameter and if so what is its dependence on the dimension d. At first sight, we could think to take σ_d as large as possible for the chain to be able to visit the state space more easily. Nevertheless, we have to ensure that some moves has to be accepted in a non negligible proportion. This aspect limits the choice of the parameter σ_d, and therefore a first step is to find a parameter σ_d such that the mean acceptance ratio at stationarity admits a limit belonging to $(0,1)$, i.e.

$$\lim_{d\to+\infty} \mathbb{E}_{\pi^d}\left[\alpha^d\left(Y_0^d, Y_0^d+\sigma_d Z_1^d\right)\right] \in (0,1) \ .$$

[RGG97] have been interested in this problem when π is of the form:

$$\pi^d(x^d) = \prod_{i=1}^d \exp\left(-\mathrm{U}(x_i^d)\right) \ , \qquad x^d = (x_1^d, \cdots, x_d^d) \ , \tag{2.18}$$

where $\mathrm{U} : \mathbb{R} \to \mathbb{R}$ is a function which satisfies:

H 1 U is three times continuously differentiable and U' is Lipschitz.

H 2 $\mathbb{E}[(\mathrm{U}'(W))^8] < \infty$ and $\mathbb{E}[(\mathrm{U}''(W))^4] < \infty$ where W is a random variable with distribution π^1.

[RGG97, corollary 1.2] shows that if for all $d \geq 1$, $\sigma_d = \ell d^{-1/2}$ for $\ell \in \mathbb{R}_+^*$, Y_0^d is distributed according to π^d, then

$$\lim_{d\to+\infty} \mathbb{E}_{\pi^d}\left[\alpha^d\left(Y_0^d, Y_0^d+\ell d^{-1/2} Z_1^d\right)\right] = 2\mathbf{\Phi}(-\ell/2\sqrt{I}) \ , \tag{2.19}$$

where $\mathbf{\Phi}$ is the cumulative distribution function of the standard normal distribution and

$$I = \int_{\mathbb{R}} (\mathrm{U}')^2(x)\pi^1(x)\mathrm{d}x \ . \tag{2.20}$$

Furthermore, [RGG97, Theorem 1.1] implies that still at stationarity, each component of the Markov chain $(Y_i^d)_{i\in\mathbb{N}}$, properly scaled, weakly converges to the solution of the overdamped Langevin equation associated with π^1. More precisely, consider the sequence of Markov chains $\{(Y_k^d)_{k\geq 0}\,;\, d \geq 1\}$ defined by (2.16) satisfying for all $d \geq 1$, Y_0^d is distributed according to π^d and

$$\sigma_d = \ell d^{-1/2} \ .$$

Define the sequence of jump processes $\{(\mathbf{Y}_t^d)_{t\geq 0}\,;\, d \geq 1\}$ from $\{(Y_k^d)_{k\geq 0}\,;\, d \geq 1\}$ by

$$\text{for all } t \geq 0 \text{ and } d \geq 1 \ , \qquad \mathbf{Y}_t^d = Y_{S_t}^d \ ,$$

where $(S_t)_{t\geq 0}$ is a Poisson process with rate d. Denote by $\mathbf{Y}_{t,1}^d$ is the first component of $\mathbf{Y}_t^d$. Then if π^d is of the form (2.18) and U satisfies H1 et H2, the sequence $(\mathbf{Y}_{t,1}^d, t \geq 0)$ weakly converges in the Skorokhod space to the solution of the Langevin equation

$$\mathrm{d}\mathbf{Y}_t = \sqrt{\mathrm{h}(\ell)}\mathrm{d}B_t^1 - \frac{1}{2}\mathrm{h}(\ell)\mathrm{U}'(\mathbf{Y}_t)\mathrm{d}t \ , \tag{2.21}$$

where $(B_t^1)_{t\geq 0}$ is a unidimensional Brownian motion, $\mathbf{Y}_0$ is distributed according to π^1, $\mathrm{h}(\ell)$ is given for all $\ell > 0$ by

$$\mathrm{h}(\ell) = 2\ell^2\Phi\left(-\frac{\ell}{2}\sqrt{I}\right) \ , \tag{2.22}$$

and I is defined by (2.20).

In addition, this result allows to tune in practice the parameter ℓ. Consider the following Langevin equation

$$\mathrm{d}\mathbf{Y}_t^c = -c\nabla \mathrm{U}(\mathbf{Y}_t^1)\mathrm{d}t + \sqrt{2c}\mathrm{d}B_t^1 \ ,$$

where $c \in \mathbb{R}_+^*$ and $\mathbf{Y}_0^c = u \in \mathbb{R}$. Under appropriate conditions on U, [Bha78] shows that for any function $f \in \mathrm{L}^2(\pi^1)$ such that $\int_{\mathbb{R}} f(x)\pi^1(\mathrm{d}x) = 0$, the process $(t^{-1/2}\int_0^t f(\mathbf{Y}_s^1)\mathrm{d}s)_{t\geq 0}$ weakly converges to zero-mean normal random variable with variance $\mathbb{E}_\pi[(\int_0^{+\infty} f(\mathbf{Y}_s^1)\mathrm{d}s)^2]$. Then by a simple change of variable, we can observe that the larger c is, the smaller this variance is. Therefore, we aim to find the constant ℓ which maximizes the function $\ell \to \mathrm{h}(\ell)$ dans (1.23). However using (1.21), a calculation shows that this function is maximal for a value of ℓ such that the limit (2.19) is equal to 0.234 (up to 4 digits).

The introduced results are established under the condition that the potential U is at least three times continuously differentiable on $\mathbb{R}$. We will present some extensions of this result under weaker assumptions on the potential U.

2.6.2 Optimal scaling of MALA

The same study can be lead for MALA. Let π^d be a probability measure on $\mathbb{R}^d$ with a positive density of the form (2.18) where $\mathrm{U} : \mathbb{R} \to \mathbb{R}$ is continuously differentiable. Recall that MALA defines the Markov chain $(Y_i^d)_{i\in\mathbb{N}}$ for a given initial condition Y_0^d by for all $k \geq 0$,

$$Y_{k+1}^d = \begin{cases} W_{k+1}^d = Y_k^d + \sigma_d^2 \nabla \log \pi^d(Y_k^d) + \sqrt{2}\sigma_d Z_{k+1}^d & \text{with probability } \alpha\left(Y_k^d, W_{k+1}^d\right) \\ Y_k^d & \text{otherwise} \ , \end{cases} \tag{2.23}$$

where $\sigma_d \in \mathbb{R}_+^*$, $(Z_i^d)_{i\geq 1}$ is a sequence of i.i.d. d-dimensional standard normal random variables, and for all $x, y \in \mathbb{R}^d$

$$\alpha^d(x,y) = \min\left\{1, \frac{\pi^d(y)\exp\left(-\left\|x - y - \sigma_d^2\nabla\log\pi^d(y)\right\|^2/(4\sigma_d^2)\right)}{\pi^d(x)\exp\left(-\left\|y - x - \sigma_d^2\nabla\log\pi^d(x)\right\|^2/(4\sigma_d^2)\right)}\right\} \ . \tag{2.24}$$

As in the case of the symmetric random walk Metropolis, the choice of the parameter σ_d is left to the choice of the user. Therefore, we can have the same considerations as for the symmetric random walk Metropolis on this choice: what is the good dependence of σ_d on the dimension so that $\mathbb{E}_{\pi^d}[\alpha^d(Y_0^d, W_1^d)]$ converges as the dimension d goes to infinity, to a constant in $(0,1)$. This time the parameter σ_d has to scale as $\ell d^{-1/6}$ for $\ell \in \mathbb{R}_+^*$. Indeed [RR98, Theorem 1] shows that if the potential satisfies:

M 1 U is eight times continuously differentiable on $\mathbb{R}$ and there exits a real polynomial P satisfying for all $i \in \{0, \cdots, 8\}$ and $x \in \mathbb{R}$, $\left|\mathrm{U}^{(i)}(x)\right| \leq \mathrm{P}(x)$

M 2 For any $p \geq 1$, $\int_{\mathbb{R}} |x|^p \, \pi_1(x)\mathrm{d}x < \infty$

then

$$\lim_{d \to +\infty} \mathbb{E}_{\pi^d} \left[\alpha^d \left(Y_0^d, Y_k^d + \ell^2 d^{-1/3} \nabla \log \pi^d(Y_k^d) + \sqrt{2}\ell d^{-1/6} Z_{k+1}^d \right) \right] = 2\mathbf{\Phi}(-\ell^3 \sqrt{2J}) \,, \tag{2.25}$$

where $\mathbf{\Phi}$ is still the cumulative distribution function of the standard normal distribution and

$$J = (48)^{-1} \int_{\mathbb{R}} \left\{ 5(\mathrm{U}^{(3)}(x))^2 - 3(\mathrm{U}^{(2)}(x))^3 \right\} \pi^1(x)\mathrm{d}x > 0 \,. \tag{2.26}$$

[RR98, Theorem 2] shows that a diffusion limit holds for MALA as well. Consider the sequence of Markov chains $\{(Y_k^d)_{k \geq 0} \,;\, d \geq 1\}$ defined by (2.16) satisfying for all $d \geq 1$, Y_0^d is distributed according to π^d and

$$\sigma_d = \ell d^{-1/6} \,.$$

Define the sequence of jump Markov processes $\{(\mathbf{Y}_t^d)_{t \geq 0} \,;\, d \geq 1\}$ from $\{(Y_k^d)_{k \geq 0} \,;\, d \geq 1\}$ by:

$$\text{for all } t \geq 0 \text{ and } d \geq 1 \,, \qquad \mathbf{Y}_t^d = Y_{S_t}^d \,,$$

where $(S_t)_{t \geq 0}$ is a Poisson process with rate $d^{1/3}$. Denote by $\mathbf{Y}_{t,1}^d$ is the first component of $\mathbf{Y}_t^d$. If π^d is of the form (2.18) and U satisfies M1 and M2, the sequence of processes $\{(\mathbf{Y}_{t,1}^d, t \geq 0) \,;\, d \geq 1\}$ weakly converges in the Skorokhod space to the solution of the Langevin equation:

$$\mathrm{d}\mathbf{Y}_t = \sqrt{2\mathrm{g}(\ell)}\mathrm{d}B_t^1 - \mathrm{g}(\ell)\dot{\mathrm{U}}(\mathbf{Y}_t)\mathrm{d}t \,, \tag{2.27}$$

where $(B_t^1)_{t \geq 0}$ is a unidimensional Brownian motion, $\mathbf{Y}_0$ is distributed according to π^1 and $\mathrm{g}(\ell)$ is given for all $\ell > 0$ by

$$\mathrm{g}(\ell) = 8\ell^2 \Phi\left(-\ell^3 \sqrt{2J} \right) \,, \tag{2.28}$$

and J defined by (2.26).

Finally as in the case of the symmetric random walk Metropolis, this diffusion limit allows to tune the parameter ℓ in practice. By (2.25), a calculation shows that the function $\ell \mapsto \mathrm{g}(\ell)$ has a unique minimizer at $\ell^\star > 0$, for which the limit mean acceptance ratio given in (2.25) has to be about 0.574.

Note that compared to the symmetric random walk Metropolis, MALA allows to take at stationarity larger step-sizes with respect to the dimension d. In a chapter, we will investigate if it is possible to improve the optimal scaling of MALA and propose an alternative to it.

2.7 Outlines and contributions

This manuscript is divided into three parts according to the different topics we have investigated. Each part is split into two chapters. With one exception, these chapters are accepted or submitted papers.

In the first part, convergence results of Markov chains in Polish space will be established, and will be applied to the study of MCMC algorithms.

In a second part, the ULA algorithm will be studied in depth. Explicit bounds are established in total variation distance and Wasserstein distance, depending on the assumptions on the potential associated with the target measure.

Finally, the last part consists in new optimal scaling results for the symmetric random walk Metropolis algorithm and a new Metropolis-Hastings type algorithm.

We present below an outline of each chapter of this thesis.

2.7.1 Part I-Chapter 3

In this first chapter, we establish geometric convergence result in Wasserstein distance for a Markov kernel P on a Polish space $(\mathsf{E}, \mathbf{d})$ endowed with its corresponding Borel σ-field $\mathcal{E}$. This result is obtained by combining a geometric drift condition and a contraction condition for the kernel P on a subset of $\mathsf{E} \times \mathsf{E}$ in the Wasserstein distance associated with the metric on E. More precisely, it is assumed that there exist a measurable set $\mathtt{G} \in \mathcal{E} \otimes \mathcal{E}$, $\ell \in \mathbb{N}^*$ and $\epsilon \in (0,1]$ such that P satisfies for all $(x,y) \in \mathtt{G}$,

$$W_{\mathbf{d}}(P^{\ell}(x,\cdot), P^{\ell}(y,\cdot)) \leq (1-\epsilon)\mathbf{d}(x,y) \ . \tag{2.29}$$

In addition, it is assumed that there exist a measurable function $V : \mathsf{E} \to [1,+\infty[$, $b \in \mathbb{R}_+$ and $\lambda \in [0,1[$ such that for all $x, y \in \mathsf{E}$,

$$PV(x) + PV(y) \leq \lambda(V(x)+V(y)) + b\mathbb{1}_{\mathtt{G}}(x,y) \ , \qquad \sup_{(z,w)\in\mathsf{E}\times\mathsf{E}} \{V(z)+V(w)\} < +\infty \ . \tag{2.30}$$

This condition is similar to the geometric drift condition (1.14) but must hold on the product space. The result derived from these two conditions gives a quantitative control on the convergence of the kernel for (possibly) non irreducible kernels. On the other hand, even if the kernel is irreducible, our result provides convergence bounds in Wasserstein distance, which can be more precise than in total variation. This kind of result has been already established in [HMS11], but the bounds obtained in this chapter depend on the constants appearing in (2.29) and (2.30) in a simpler way. Furthermore, the technique of proof is completely different and relies on an appropriate coupling of the chain.

This result is then applied to a MCMC algorithm defined as follows. Let π be a target density on $\mathbb{R}^d$ given for all $x \in \mathbb{R}^d$ by

$$\pi(x) = \mathcal{Z}^{-1} \exp(-U(x) - \Gamma(x)) \ ,$$

with $\mathcal{Z} = \int_{\mathbb{R}^d} \exp(-U(x) - \Gamma(x))\mathrm{d}x < \infty$, $\Gamma : \mathbb{R}^d \to \mathbb{R}$ and $U : \mathbb{R}^d \to \mathbb{R}$. Assume that U is of the form for all $x \in \mathbb{R}^d$,

$$U(x) = (1/2)x^T Q x + \Upsilon(x) \ ,$$

where Q is a positive definite matrix and $\Upsilon : \mathbb{R}^d \to \mathbb{R}$ is a convex, gradient Lipschitz function. Then consider the Langevin diffusion associated with U:

$$\begin{aligned} \mathrm{d}\mathbf{Y}_t &= -\mathbf{Y}_t \mathrm{d}t - Q^{-1}\nabla\Upsilon(\mathbf{Y}_t)\mathrm{d}t + \sqrt{2}Q^{-1/2}\mathrm{d}B_t^d \ , \\ \mathbf{Y}_0 &= y_0 \ , \end{aligned}$$

where $(B_t^d)_{t\geq 0}$ is a d-dimensional Brownian motion. By assumption on U, this SDE admits a unique strong solution $(\mathbf{Y}_t)_{t\geq 0}$. It is an easy task to show that for all $t \geq 0$ and $\delta > 0$,

$$\mathbf{Y}_{t+\delta} = \mathbf{Y}_t \mathrm{e}^{-\delta} - \int_t^{t+\delta} \mathrm{e}^{-(t+\delta-s)} Q^{-1}\nabla\Upsilon(\mathbf{Y}_s)\mathrm{d}s + \sqrt{2}Q^{-1/2}\int_t^{t+\delta} \mathrm{e}^{-(t+\delta-s)}\mathrm{d}B_s^d \ . \tag{2.31}$$

Therefore, the Euler scheme for (2.31) with a step size $\delta > 0$ defines the Markov chain $(Y_k)_{k\in\mathbb{N}}$, given for all $k \geq 0$ by

$$\bar{Y}_{k+1} = \bar{Y}_k \mathrm{e}^{-\delta} - (1-\mathrm{e}^{-\delta})Q^{-1}\nabla\Upsilon(\bar{Y}_k) + Z_{k+1} \ , \tag{2.32}$$

where $\bar{Y}_0$ is the starting point, $(Z_k)_{k\in\mathbb{N}^*}$ are i.i.d. zero-mean Gaussian random variables with covariance matrix $(1-\mathrm{e}^{-2\delta})Q^{-1}$. This kind of discretization is called a stochastic exponential Euler integrator, see [LR04]. The chain associated with the relation (2.32) defines a Markov kernel which can be used combined with a Metropolis-Hastings acceptance step to target π. The algorithm so defined is called the EI-MALA algorithm. It generalises an algorithm proposed in [Bes+08], setting $h = 2(1-\mathrm{e}^{-\delta})$ and $\Gamma = 0$. This algorithm was analyzed in [Ebe14] under some assumptions on Q and Υ. We complete here this analysis and establish under appropriate conditions geometric convergence of the Markov kernel produced by the algorithm with explicit bounds. In particular, we show a logarithmic dependence on the dimension when $\Upsilon = 0$ and Γ is bounded. Moreover, we check that the assumptions of our result hold for a Bayesian inverse problem. Finally, some numerical simulations are given to support our findings.

This work has been published in Statistics and Computing [DM15b].

2.7.2 Part I Chapter 4

We complete in this part Chapter 3 on the study of Markov kernels which are potentially non-irreducible, in particular chains defined on functional state space. We aim to establish

- conditions to get existence and especially uniqueness of an invariant distribution of Markov kernel P on a Polish space $(\mathsf{E}, \mathbf{d})$, with $\mathcal{E}$ the corresponding Borel σ-field.
- convergence rate of the chain to its stationary distribution.

As well as in the irreducible case, it is assumed that there exist a measurable function $V : \mathsf{E} \to [1, +\infty)$ and a constant $b \in \mathbb{R}$ such that for all $x, y \in \mathsf{E}$,

$$PV(x) + PV(y) \leq V(x) + V(y) - 1 + b\mathbb{1}_{\mathsf{G}}(x,y) \ , \qquad \sup_{(z,w)\in\mathsf{G}} \{V(z) + V(w)\} < +\infty \ ,$$

where $\mathtt{G} \in \mathcal{E} \otimes \mathcal{E}$ is a set on which the kernel P satisfied the contraction condition (2.29). This condition is similar to the condition (2.9) in the irreducible case but is formulated on the product space $\mathsf{E} \times \mathsf{E}$.

Besides, we establish quantitative bounds for the convergence in Wasserstein distance. Such bounds in the geometric case have been obtained in [HMS11] and the work presented inChapter 3. In this chapter, we establish sub-geometric convergence rates in Wasserstein distance under sub-geometric drift conditions. These conditions are similar to the one given in [TT94] and [Dou+04] to establish sub-geometric convergence rates in total variation. The convergence in Wasserstein distance has been already been considered in [But14] but the obtained convergence rates do not fit the established one for the total variation distance. We obtain in this chapter the same convergence rates reported in [Dou+04] for the total variation distance. Whereas the method of proof used in [But14] is inspired by [HMS11], our proof is inspired by the method used in the geometric case presented in Chapter 3, and some ideas from [Dou+04].

We consider two applications of our theoretical contributions. We first apply our results to a functional auto-regressive model on $\mathbb{R}^d$ of the form

$$Y_{k+1} = \mathrm{h}(Y_k) + \xi_{k+1} \ , \tag{2.33}$$

where $\mathrm{h} : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable function and $(\xi_k)_{k\in\mathbb{N}^*}$ are i.i.d. random variables. Note that if the sequence $(\xi_k)_{k\in\mathbb{N}^*}$ is valued in a discrete denumerable alphabet, then the Markov chain is not irreducible. It is assumed that h is a strict contraction at the center of the space, i.e. there exists a function $\varpi : \mathbb{R}^d \times \mathbb{R}^d \to]0,1[$ such that for all $x, y \in \mathbb{R}^d$,

$$\|\mathrm{h}(x) - \mathrm{h}(y)\| \leq \varpi(x,y) \|x - y\| \ ,$$

and for all $R \in \mathbb{R}_+$, $\phi(R) = \sup\{\varpi(x,y) \,|\, \|x\| + \|y\| \leq R\} < 1$. If $\sup_{R\in\mathbb{R}_+} \phi(R) < 1$, then the model is geometrically ergodic in Wasserstein distance. We are interested in the case where $\limsup_{R\to+\infty} \phi(R) = 1$. Depending on the conditions on the rate of this convergence and moment conditions on the distribution of the sequence $(\xi_k)_{k\in\mathbb{N}^*}$, we establish sub-geometric convergence rate for the Markov kernel associated with (2.33).

The second example deals with the pre-conditioned Crank-Nicolson algorithm [Bes+08]. It is a Metropolis-Hastings algorithm defined on a Hilbert space H endowed with a Gaussian measure μ_0, and applied to a target distribution π which admits a positive density $x \to \exp(-\Phi(x))$ with respect to μ_0. The proposal kernel is associated with the auto-regressive model:

$$W_{k+1} = (1-\eta)^{1/2} W_k + \eta^{1/2} Z_{k+1} \ , \tag{2.34}$$

where $\eta \in [0,1[$ and $(Z_k)_{k\in\mathbb{N}^*}$ is a sequence of i.i.d. Gaussian random variables with distribution μ_0. Note that if G_1 and G_2 are two Gaussian random variables with distribution μ_0, then for all $\eta \in]0,1[$, $(G_1, (1-\eta)^{1/2} G_1 + \eta^{1/2} G_2)$ and $((1-\eta)^{1/2} G_1 + \eta^{1/2} G_2, G_1)$ are two Gaussian random variables on $\mathsf{H} \times \mathsf{H}$ with the same covariance operator, therefore have the same distribution. It follows that the Markov kernel Q associated with the Markov chain defined by (2.34) is reversible with respect to μ_0. Following the discussion on the Metropolis-Hastings type algorithms in Section 1.3-(B), the Metropolis-Hastings kernel P_{pCN} defined by (1.4), associated with the kernel Q and with the acceptance ratio given for all $x, y \in \mathsf{H}$ by

$$\alpha(x, y) = \min\{1, \exp(\Phi(x) - \Phi(y))\} ,$$

is reversible with respect to π. [HSV14] have shown that if the acceptance ratio is uniformly lower bounded then P_{pCN} is geometrically ergodic for a particular Wasserstein distance. We weaken this condition and show that if Φ is Lipschitz, then the produced Markov kernel is sub-geometrically ergodic for a particular Wasserstein distance.

This work has accepted for publication in les Annales de l'IHP.

2.7.3 Part II-Chapter 5

This chapter consists in a detailed study of the ULA algorithm presented in Section 1.4.2 to sample from a target distribution π on $\mathbb{R}^d$, $d \geq 1$, of the form (1.8). We establish explicit bounds in total variation distance between the marginal laws of the Markov chain produced by ULA and π. We obtain results for both fixed and decreasing step size sequence.

The method of proof is very different from classical approach because our aim is to get quantitative results. Indeed, for Langevin diffusions, numerous results have been recently established to quantify the distance in total variation between the law of the diffusion and its stationary measure. These results follows from either sharp functional inequalities, such Poincaré or logarithmic Sobolev inequality see [Bak+08] and [BGL14], or coupling, see [Ebe15]. In particular, we obtain original convergence result for the Langevin diffusion combining the reflection coupling proposed in [LR86] and quantitative method of convergence introduced in Chapter 3. These results complete the work of [Ebe15] which deals with convergence of diffusion processes in Wasserstein distance.

To compare the diffusion and its discretization, we use an approach coming from [Dal16] based on the Girsanov Theorem which, combined with the Pinsker inequality, allows to bound the total variation distance between the laws of the diffusion and the continuous interpolation of its discretization.

Based on these results, two strategies are analyzed: in the first one the number of iterations of the algorithm is set and the second supposes that the algorithm can be stopped at any time. In particular, depending on the curvature and the regularity of the potential associated with π, we are interested in the dependence of the obtained bounds on the dimension of the state space, which is important for statistical applications. Numerous conditions on the potential are considered: super-exponential, convex, perturbations of a strongly convex function, which correspond to practical situations. For each case, we obtain explicit bounds, which depend in some scenarios polynomially on the dimension.

At fixed step size, under mild assumption on the potential associated with π, the Markov chain produced by ULA admits a unique invariant distribution π_γ. We give in this chapter an explicit bound in V-total variation between the target measure π and π_γ of order $\sqrt{\gamma}$.

We show that if the sequence of step sizes $(\gamma_k)_{k\geq 0}$ goes to 0 and $\sum_{k=1}^{+\infty}\gamma_k = +\infty$, then the sequence of marginal laws of the produced inhomogeneous Markov chain converges to the target measure π, still in total variation distance.

The results of this chapter complete [Dal16], where only the case of strongly convex potentials is considered. It is a joint work with the professor Éric Moulines. It has been accepted for publication in Annals of App. Prob.

2.7.4 Part II-Chapter 6

In this chapter, we complete and improve the results of the previous chapter in the case the target density π is strongly log-concave on $\mathbb{R}^d$, $d \geq 1$, and of the form (1.8). First the convergence in Wasserstein distance is studied. Explicit bounds are established using the synchronous coupling between the Langevin diffusion and its discretization associated with a sequence of step sizes $(\gamma_k)_{k\in\mathbb{N}^*}$. This coupling uses the same Brownian motion for the two processes and is given for any initial conditions by:

$$\begin{cases} \mathbf{Y}_t = \mathbf{Y}_0 - \int_0^t \nabla U(\mathbf{Y}_s)\mathrm{d}s + \sqrt{2}B_t^d & \text{for all } t \geq 0 \\ Y_{k+1} = Y_k - \gamma_{k+1}\nabla U(Y_k) + \sqrt{2}(B^d_{\Gamma_{k+1}} - B^d_{\Gamma_k}) & \text{for all } k \in \mathbb{N}^* \ , \end{cases}$$

where $(B_t^d)_{t\geq 0}$ is a d-dimensional Brownian motion and $\Gamma_k = \sum_{i=1}^k \gamma_i$ for all $k \in \mathbb{N}^*$. When the sequence of step sizes goes to 0 and $\sum_{k=1}^{+\infty}\gamma_k = +\infty$, we derive using these bounds, explicit convergence rate in Wasserstein distance between the marginal laws of the produced inhomogeneous Markov chain and the target distribution π.

At fixed step size $\gamma > 0$, we deduce an explicit bound between π and the stationary distribution π_γ of the homogeneous Markov chain produced by the ULA algorithm. Regarding the dependence on the dimension, for a target precision $\epsilon > 0$, we show that a number of iterations of order $d\epsilon^{-2}$ or $d^{1/2}\epsilon^{-1}$ is sufficient for the Wasserstein distance between the marginal law of the algorithm and the target measure to be smaller than ϵ. The difference between these two results comes from different regularity conditions on U (it is assumed for one of the two that U is three times continuously differentiable).

Then, we adapt results from [JO10] to the considered Markov chain which can be potentially inhomogeneous to get explicit bounds on the Mean Square Error and exponential deviation inequalities for estimators of $\int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ when $f : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz.

In a second time, we use the derived bounds in Wasserstein distance to get explicit bounds in total variation distance. Indeed, as U is convex, then the semi-group associated with the Langevin diffusion has a regularizing effect. More precisely, let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable and bounded function. Then for all $t > 0$, $\mathbf{P}_t f$ is a Lipschitz function with a Lipschitz coefficient which is explicit and depends on t. However this coefficient in short time is of order $t^{-1/2}$. That is why we in part use the method presented in Chapter 6 to deal with this problem. The resulting bounds gives convergence rate in total variation to π, when the sequence of step sizes goes 0 and $\sum_{k=1}^{+\infty}\gamma_k = +\infty$. These rates are significantly better than the one obtained in Chapter 6. As the discretization step sizes are held constant, the derived bounds imply that if U is sufficiently regular then a number of iterations of order $\sqrt{d}\epsilon^{-1}$ is sufficient for the marginal law of the algorithm to be in a ball centered at π and with radius $\epsilon > 0$ for the total variation distance. Again, this result improves the bounds obtained in Chapter 5 which would give a number of iterations of order $d\epsilon^{-2}$.

We conclude this work by establishing explicit bounds on the Mean Square Error and exponential deviation inequalities for estimators of $\int_{\mathbb{R}^d} f(x)\pi(x)\mathrm{d}x$ where $f : \mathbb{R}^d \to \mathbb{R}$ is a measurable and bounded function. However, we cannot apply directly the method used when f is Lipschitz. First, we show that the Markov kernel associated with the Euler discretization has a regularizing effect as well. For this, we are interested in functional auto-regressive models of the form

$$W_{k+1} = \mathrm{h}(W_k) + \sigma Z_{k+1} , \tag{2.35}$$

where $\mathrm{h} : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz, $\sigma > 0$ and $(Z_k)_{k\in\mathbb{N}^*}$ is a sequence of i.i.d. d dimensional standard normal random variables. We prove that the Markov kernel $\mathbf{K}_\sigma$ associated with this model is regularizing using a generalization of a coupling which was originally proposed in [BDJ98]. More precisely, let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable and bounded function. Under appropriate conditions on h, $\mathbf{K}_\sigma f$ is Lipschitz with Lipschitz coefficient of order σ^{-1}.

It is a joint work with the professor Éric Moulines.

2.7.5 Part III-Chapter 7

The original optimal scaling result for the symmetric random walk Metropolis [RGG97] assumes that the target distribution has a positive and three times continuously differentiable density with respect to the Lebesgue measure, see Section 2.6.1. Therefore, it can not be applied if the density is not regular as for example the Laplace distribution (which is not differentiable at some points). Nevertheless, if we observe the result of [RGG97], the only quantity which appears in the limit of the mean acceptance ratio at stationarity is the Fisher information matrix associated with the translation model. This quantity is well defined if the density is differentiable in quadratic mean. Therefore, the question that arises is if we can use results and techniques for quadratic mean differentiable models to weaken the assumptions in [RGG97]. One of the main motivation of this work is the analysis of Bayesian methods using prior distributions which have convex, Lipschitz and non-smooth potentials. This kind of prior distributions naturally appear in high-dimensional Bayesian inference. For example in the Bayesian analysis of the LASSO algorithm, log prior densities are chosen to be weighted L^1 norms.

The first result of this chapter is the generalization of the optimal scaling result of [RGG97] to positive densities on $\mathbb{R}$ and differentiable in L^p mean for $p \geq 2$. Therefore, this result can be applied to densities which are not differentiable everywhere. We show that the mean acceptance ratio at stationarity admits a limit if $\sigma_d = \ell d^{-1/2}$, see (1.21). Besides, we show that the algorithm admits a diffusive limit: the limit diffusion is a Langevin diffusion which can be singular and does not admit strong solutions.

The method of proof is different from [RGG97]. As in [JLM15], we directly show the weak convergence of the sequence of interpolated processes associated with the algorithm. For this, we first show that this sequence is tight in the Wiener space. Then we show that any limit point is solution of a well-posed martingale problem. Finally, we use the equivalence between this martingale problem and the associated Langevin diffusion. Besides as already emphasized, the proof relies on methods used to establish LAN conditions for densities differentiable in quadratic mean. In particular, by this method, we can make an expansion of the acceptance ratio which does not rely on the existence of a second derivative of the density.

In a second time, we extend this result to densities which can have a bounded support on $\mathbb{R}$ but are still differentiable in quadratic mean. This result is applied to densities similar to Beta and Gamma densities. This result completes the work of [**neal:roberts:2012**] which considers rougher discontinuous densities.

This work has been submitted for publication.

2.7.6 Part III-Chapter 8

We have seen in Section 2.6.2 that the optimal scaling of MALA was in $d^{-1/3}$. It is then natural to wonder if it could be possible to improve this scaling and therefore the dependence on the dimension by using more information on the geometry of the target distribution.

A first idea would be to use a higher order integrator than the Euler-Maruyama discretization, as a proposal in a Metropolis-Hastings algorithm. However, we will see in this chapter that a better discretization scheme for the SDE do not improve the scaling of the associated Metropolis-Hastings algorithm. Instead of focusing on the discretization error, it will be shown that it is the order of the acceptance ratio in the parameter σ_d which is important to control. Indeed to illustrate this statement, we give the main ideas for the proofs of the optimal scaling results for the class of densities of the form (2.18) and proposal kernels associated with Markov chains given for all $k \in \mathbb{N}$ by,

$$W_{k+1}^d = \mathrm{F}(W_k^d, \sigma_d^{1/2}) + \Sigma(W_k^d, \sigma_d^{1/2}) Z_{k+1}^d \;, \tag{2.36}$$

where $\sigma_d \in \mathbb{R}_+^*$, $\mathrm{F} : \mathbb{R}^d \times \mathbb{R}_+^* \to \mathbb{R}^d$, $\Sigma : \mathbb{R}^d \times \mathbb{R}_+^* \to M_d(\mathbb{R}^d)$[11], and $(Z_k)_{k\geq}$ are i.i.d. d-dimensional standard normal random variables. Note that for the symmetric random walk Metropolis algorithm, we have

$$\mathrm{F}(x, \sigma_d^{1/2}) = x, \quad \Sigma(x, \sigma_d^{1/2}) = \sqrt{\sigma_d}\,\mathrm{Id} \;,$$

and for MALA

$$\mathrm{F}(x, \sigma_d^{1/2}) = x - \sigma_d \nabla \mathrm{U}(x), \quad \Sigma(x, \sigma_d) = \sqrt{2\sigma_d}\,\mathrm{Id} \;.$$

The acceptance ratio can be written of the form for all $x, y \in \mathbb{R}^d$

$$\alpha(x, y) = \min\{1, \exp(R_d(x, y))\} \;, \tag{2.37}$$

[11] $M_d(\mathbb{R}^d)$ is the set of square matrices of dimension d

for a function $R_d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which depends on the proposal kernel associated with (2.36). Besides, we are interested in the value of the ratio as y is of the form

$$y = \mathrm{F}(x, \sigma_d^{1/2}) + \Sigma(x, \sigma_d^{1/2}) z \ .$$

If we assume that $R_d(x,x) = 0$ and regularity conditions on F, Σ and U, then a Taylor expansion of R_d can be made in power of $\sigma_d^{1/2}$ in a neighborhood of 0:

$$R_d(x,y) = \sum_{i=1}^{k} \sum_{j=1}^{d} \sigma_d^{i/2} C_i(x_j, z_j) + \sigma_d^{(k+1)/2} L_{k+1}(x, \sigma_d^{1/2}, z) \ . \tag{2.38}$$

Also, the optimal scaling associated with proposals of the form (2.36) (1.38) are crucially linked with the number of terms C_i which are zeros in (2.38). If $C_i = 0$, for $i = 1, \cdots, p$, $p \in \mathbb{N}^*$ and $\sigma_d = \ell d^{1/(p+1)}$ then the dominating term in (2.38) is

$$\frac{\ell^{p+1}}{\sqrt{d}} \sum_{j=1}^{d} C_{p+1}(x_j, z_j) \ .$$

Following this study of the acceptance ratio, we can conclude up to technicalities that

$$\lim_{d \to +\infty} \mathbb{E}[\alpha(W_0^d, W_1^d)] \in]0,1[\ ,$$

where W_0^d is random variable distributed according to π^d and W_1^d is given by (2.36).

To obtain a new Metropolis-Hastings type algorithm with a better optimal scaling than MALA, we first restrict the class of proposal densities associated with (2.36) by setting the following form for the functions F and Σ : for $x \in \mathbb{R}^d$ and $\sigma_d > 0$,

$$\mathrm{F}(x, \sigma_d^{1/2}) = x + \sigma_d \mathrm{F}_1(x) + \sigma_d^2 \mathrm{F}_2(x) \ , \qquad \Sigma(x, \sigma_d^{1/2}) = \sigma_d^{1/2} \Sigma_1(x) + \sigma_d^{3/2} \Sigma_2(x) \ ,$$

where for $i = 1, 2$, $\mathrm{F}_i(x) : \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma_i : \mathbb{R}^d \to M_d(\mathbb{R}^d)$. Next, we make a Taylor expansion of R_d given by (2.37) in $\sigma_d^{1/2}$ in a neighborhood of 0 for such functions F and Σ. Then, we try to have explicit expressions for the functions $\mathrm{F}_1, \mathrm{F}_2, \Sigma_1, \Sigma_2$ to get that the first fourth terms $C_i(\cdot,\cdot)$, $i \in \{1,2,3,4\}$ in (2.38) are identically zeros. Therefore, we obtain a system of four equations in four unknowns which admits a unique solution. We obtain by this method the following expression for the functions F and Σ

$$\begin{aligned}
\mathrm{F}(x, \sigma_d^{\sigma_d}) &= x - \frac{\sigma_d}{2} \nabla U^d(x) - \frac{\sigma_d^2}{24} \left(\nabla^2 U^d(x) \nabla U^d(x) - \vec{\Delta}(\nabla U^d)(x) \} \right) , \\
\Sigma(x, \sigma_d^{1/2}) &= \sigma_d^{1/2} \,\mathrm{Id} - (\sigma_d^{3/2}/12) \nabla^2 U^d(x) \ ,
\end{aligned}$$

where $U^d = \log \pi^d$ and $\vec{\Delta}$ is the vector Laplacian. The Metropolis-Hastings type algorithm associated with this choice of functions F and Σ will be called the fast Metropolis Adjusted Langevin Algorithm (fMALA). Note that this proposals does not lead to a discretization algorithm for SDEs of a higher order then the Euler-Maruyama scheme.

In a second time, we show that the optimal scaling of the fMALA algorithm is of order $d^{-1/5}$ and therefore improves the optimal scaling of MALA. Besides, we study other forms for the functions F and Σ to obtain two other proposals for which the optimal scaling of the associated Metropolis-Hastings algorithm is of order $d^{-1/5}$.

Then, we study the stability and ergodicity of these new Metropolis-Hastings algorithms. In particular, we show that one of them is geometrically ergodic for a class of densities close to the ones defined in (2.15) by establishing a geometric drift conditions (2.12). Furthermore, we give a criteria on F and Σ which implies that a Metropolis-Hastings algorithm associated with a proposal of the form (2.36), is not geometrically ergodic. This result completes [RT96a, Théorème 4.2]. We use this criteria to show that fMALA is not geometrically ergodic if $\lim_{\|x\|\to+\infty} \|\nabla U(x)\| / \|x\| = +\infty$.

Finally, we present some numerical simulations to support our findings, and besides some algorithmic strategies to handle the transient phase of the algorithm. Indeed, our optimal scaling result can be applied only to stationary chains. Note that this problem also occurs for MALA but for the symmetric random walk Metropolis algorithm, the optimal scaling is still of order d^{-1} even if the chain is not at stationarity, see [JLM15].

Then, we propose an hybrid strategy, similar to the one in [CRR05] for MALA. With probability 1/2, the proposal associated with the symmetric random walk Metropolis algorithm is used and the proposal associated with fMALA otherwise. We empirically compare the hybrid strategies of MALA and fMALA, and observe that the one associated with fMALA converges faster and presents better autocorrelation functions.

This work has been submitted for publication.

Part I

Convergence of Markov chains and applications to MCMC

Chapter 3

Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis Adjusted Langevin Algorithm

Alain Durmus[1], Éric Moulines [2]

Abstract

In this paper, we establish explicit convergence rates for Markov chains in Wasserstein distance. Compared to the more classical total variation bounds, the proposed rate of convergence leads to useful insights for the analysis of MCMC algorithms, and suggests ways to construct sampler with good mixing rate even if the dimension of the underlying sampling space is large. We illustrate these results by analyzing the Exponential Integrator version of the Metropolis Adjusted Langevin Algorithm (EI-MALA). We illustrate our findings using a Bayesian linear inverse problem.

The derivation of explicit bounds in total variation distance has received much attention in recent years, motivated mainly by control of convergence for Markov chain Monte Carlo. Most of these bounds are based on a *geometric drift condition* and an associated *minorization condition* for the underlying Markov chain, which together imply

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

geometric ergodicity under weak additional conditions [MT09, Chapters 15, 16]. Once drift and minorization have been established, the results presented in [Ros95], [RT99], or [Bax05] can be employed to calculate a bound on the number of iterations needed to get within a pre-specified (total variation) distance of the target distribution. These explicit bounds are based either on the Nummelin splitting construction and the renewal Theorem (see [Bax05] which extends some earlier results by [MT94]), or on the coupling construction and the Lindvall coupling inequality (see [Ros95], [RT99]). As evidenced in [JH01], these bounds are known to be rather pessimistic, especially when the dimension of the sampling space is large.

In this work, we first provide explicit bound for convergence in Wasserstein distance. Let P be a Markov transition kernel defined on a Polish space $(E, d_\star)$. Denote by $\mathcal{B}(E)$ the associated Borel σ-algebra and $\mathcal{P}(E)$ the set of probability measures on $(E, \mathcal{B}(E))$.

Let $\mu, \nu \in \mathcal{P}(E)$; ξ is a coupling of μ and ν if ξ is a probability on the product space $(E \times E, \mathcal{B}(E \times E))$, such that $\xi(A \times E) = \mu(A)$ and $\xi(E \times A) = \nu(A)$ for all $A \in \mathcal{B}(E)$. The set of couplings of $\mu, \nu \in \mathcal{P}(E)$ is denoted by $\Pi(\mu, \nu)$. Let d be a distance on E, topologically equivalent to $d_\star$, i.e. d generates the same topology on E than $d_\star$. The Wasserstein distance associated with d is defined by:

$$W_d(\mu, \nu) = \inf_{\xi \in \Pi(\mu,\nu)} \int_{E \times E} d(x, y) \mathrm{d}\xi(x, y) \; . \tag{3.1}$$

When d is the trivial metric $d_0(x, y) = \mathbb{1}_{x \neq y}$, the associated Wasserstein metric is, up to a multiplicative factor, the total variation d_{TV} (see ([Vil09, Chapter 6]). When d is bounded, the Monge-Kantorovich duality Theorem implies (see [Vil09, Remark 6.5]) that the lower bound in (3.1) is reached. In addition, W_d is a metric on $\mathcal{P}(E)$ and $\mathcal{P}(E)$ equipped with W_d is a Polish space; see [Vil09, Theorems 6.8 and 6.16]. Finally, the convergence in Wasserstein distance W_d implies the weak convergence (see e.g., [Vil09, Corollary 6.11]).

Our goal is to find explicit bounds on rates of convergence of $W_d(\mu P^n, \nu P^n)$ to zero. In the special case in which P has a stationary distribution π, this corresponds to bounding the convergence of μP^n to π. Our results extend the nonquantitative results developed for example by [MT09, Chapters 15 and 16] to Markov chains which are not necessarily ϕ-irreducible. It also complements some of the results presented recently in [HSV14] and [Cot+13].

The bound on $W_d(\mu P^n, \nu P^n)$ is also based on a drift condition but we replace the minorization condition by the existence of a *coupling set*. More precisely, we assume the existence of a *coupling kernel* $\mathbf{K}((x, y), \cdot)$ of $P(x, \cdot)$ and $P(y, \cdot)$, i.e. $\mathbf{K}((x, y), A \times E) = P(x, A)$ and $\mathbf{K}((x, y), E \times A) = P(y, A)$, for any $A \in \mathcal{B}(E)$. We assume that this coupling kernel $\mathbf{K}((x, y), \cdot)$ is weakly contracting on the whole state space, i.e. $\mathbf{K}d(x, y) \leq d(x, y)$ for all $(x, y) \in E \times E$, and strongly contracting when (x, y) belongs to a *coupling set* Δ, i.e. $\mathbf{K}d(x, y) \leq (1 - \epsilon)d(x, y)$, $(x, y) \in \Delta$. This assumption is combined with a drift condition for the coupling kernel outside the coupling set, which allows to control the exponential moment of the hitting time of the coupled chain to the coupling set. Under these assumptions, we prove that the Wasserstein distance between $P^n(x, \cdot)$ and $P^n(y, \cdot)$

decreases at a geometric rate with a rate depending explicitly from the constants in the drift conditions and in the coupling set. Our results are closely related to the general form of the Harris Theorem stated in [HMS11, Theorem 1.8] (see also [But14]), but the assumptions are slightly weaker; the techniques of proof are completely different, and share many similarities with the methods used to establish explicit rate of convergence in total variation via the coupling construction.

As mentioned above, the rate of convergence obtained for the total variation distance is practically useless to analyze MCMC algorithm when the dimension of the state space becomes large. Despite the fact that the bounds are explicit, these results say in practice little more than 'the chain converges for large n'; see [JH01] and [RR04]. On the contrary, as observed in several recent works in this direction (see [Cot+13] and the references therein), the Wasserstein bounds are much more informative, at least for appropriately designed MCMC algorithms. One of the key to the success of our approach for Wasserstein distance (under of course appropriate assumptions on the transition kernel and particular choice of the distance) is the ability to couple MCMC algorithms "naturally" by simply running two versions of the algorithm with the same random numbers. This in contrast with the "general" coupling construction used for total variation convergence where an attempt to make the two components equal is made only when the two versions of the chain meet in a coupling set (the probability meeting in a coupling set is typically not large and the probability of successfully coupling the chain is on the top of this vanishingly small in large dimension). We provide an illustration of this fact on a version of the Metropolis Adjusted Langevin Algorithm using an exponential integrator (EI-MALA) originally proposed by [Ebe14].

The remainder of this article is organized as follows. In Section 3.1 coupling set and drift are defined. We derive a *contraction inequality*, which is the key result for deriving convergence rate bounds for Markov chains. A theorem that allows one to use drift and coupling set to get exact upper bounds on the Wasserstein distance to stationarity is stated. In Section 3.2 we explain how to use these results to compute explicit bounds for the convergence for the EI-MALA. A limited Monte Carlo experiment on a linear Bayesian inverse problem is presented in Section 3.3 to support our findings.

3.1 Quantitative bounds for geometric convergence in Wasserstein distance

Definition 3.1 (Coupling set). *Let $\Delta \in \mathcal{B}(E \times E)$, $\epsilon \in (0,1)$ and d be a distance on E topologically equivalent to $\mathbf{d}$. Δ is a (ϵ, d)-coupling set for the Markov kernel P on $(E, \mathcal{B}(E))$ if there exists a kernel $\mathbf{K}$ on $(E \times E, \mathcal{B}(E \times E))$ satisfying the following conditions*

(i) for all $x, y \in E$, $\mathbf{K}((x,y), \cdot) \in \Pi(P(x,\cdot), P(y,\cdot))$.

(ii) $\mathbf{K}$ is a weak contraction, i.e., for all $x, y \in E$, $\mathbf{K}d(x,y) \le d(x,y)$.

(iii) $\mathbf{K}$ is a strict contraction on Δ, i.e., for all $(x,y) \in \Delta$, $\mathbf{K}d(x,y) \le (1-\epsilon)d(x,y)$.

Remark 3.2. *Let d be a distance topologically equivalent to $\mathbf{d}$ bounded by 1. If*

$$\begin{aligned} &\textit{for all } (x,y) \in E^2, \quad W_d(P(x,\cdot),P(y,\cdot)) \le d(x,y) \\ &\textit{for all } (x,y) \in \Delta, \quad W_d(P(x,\cdot),P(y,\cdot)) \le (1-\epsilon)d(x,y) \end{aligned}$$

for some $\epsilon \in (0,1)$, then [Vil09, corollary 5.22] implies that there exists a Markov kernel $\mathbf{K}$ on $(E \times E, \mathcal{B}(E \times E))$ which is such that Δ is a (ϵ, d)-coupling set. In this sense, the existence of the such coupling set is ultimately related to the contraction properties of the kernel.

We preface the statements of the main results by recalling some properties of the Wasserstein distance that are repeatedly used in the sequel. For any measurable function $l : E \times E \to \mathbb{R}_+$, we define the optimal transportation for $\mu, \nu \in \mathcal{P}(E)$ by:

$$W_l(\mu,\nu) = \inf_{\xi \in \Pi(\mu,\nu)} \int_{E\times E} l(x,y) \mathrm{d}\xi(x,y) \ . \tag{3.2}$$

Note that we may have $W_l(\mu,\nu) = +\infty$, and for all $x, y \in E \times E$, $W_l(\delta_x, \delta_y) = l(x,y)$. We consider the case when the function l is a distance-like function (see also [HMS11])

Definition 3.3. *A function $l : E \times E \to \mathbb{R}_+$ is said to be a* distance-like *if*

1. *For all (x,y) in E^2, $l(x,y) = 0$ if and only if $x = y$.*
2. *l is lower semicontinuous.*
3. *For all (x,y) in E^2, $l(x,y) = l(y,x)$.*

The following lemma establishes the convexity of W_l, when l is a distance-like function.

Lemma 3.4. *Let (E,d) be a Polish space. Let P be a Markov kernel on $(E, \mathcal{B}(E))$ and $l : E \times E \to \mathbb{R}_+$ be a distance-like function. For any $\mu, \nu \in \mathcal{P}(E)$*

$$W_l(\mu P, \nu P) \le \inf_{\xi \in \Pi(\mu,\nu)} \int_{E\times E} W_l(P(x,\cdot),P(y,\cdot))\, \xi(\mathrm{d}x,\mathrm{d}y) \ .$$

Proof. Let ξ be a coupling of μ and ν. We get

$$\begin{aligned} \mu P(\mathrm{d}z) &= \int_E P(x,\mathrm{d}z)\mu(\mathrm{d}x) = \int_{E\times E} P(x,\mathrm{d}z)\xi(\mathrm{d}x,\mathrm{d}y). \\ \nu P(\mathrm{d}z) &= \int_{E\times E} P(y,\mathrm{d}z)\xi(\mathrm{d}x,\mathrm{d}y) \ . \end{aligned}$$

Therefore,

$$W_l(\mu P, \nu P) = W_l\left(\int_{E\times E} P(x,\cdot)\xi(\mathrm{d}x,\mathrm{d}y), \int_{E\times E} P(y,\cdot)\xi(\mathrm{d}x,\mathrm{d}y)\right) \ .$$

Since l is lower semicontinuous and $l \ge 0$, by [Vil09, Theorem 4.8]

$$W_l(\mu P, \nu P) \le \int_{E\times E} W_l(P(x,\cdot),P(y,\cdot))\ \xi(\mathrm{d}x,\mathrm{d}y) \ .$$

The proof is concluded since this inequality holds for all couplings ξ. □

Lemma 3.5. *Let (E, d) be a Polish space and let P be a Markov kernel on $(E, \mathcal{B}(E))$. Assume there exists a Markov kernel $\mathbf{K}$ on $(E \times E, \mathcal{B}(E \times E)$ satisfying:*

(i) for all $x, y \in E$, $\mathbf{K}((x, y), \cdot)$ is a coupling kernel for $(P(x, \cdot), P(y, \cdot))$.

(ii) for all $x, y \in E$, $\mathbf{K}d(x, y) \leq d(x, y)$.

Then for all $x, y \in E$, $W_d(P(x, \cdot), P(y, \cdot)) \leq d(x, y)$ and for all probability measures $\mu, \nu \in \mathcal{P}(E)$,

$$W_d(\mu P, \nu P) \leq W_d(\mu, \nu) \ . \tag{3.3}$$

Proof. By assumption and the definition of the Wasserstein distance (3.1), we have for all $x, y \in E$,

$$W_d(P(x, \cdot), P(y, \cdot)) \leq \mathbf{K}d(x, y) \leq d(x, y) \ .$$

The second statement follows from Lemma 3.4 upon writing

$$\begin{aligned} W_d(\mu P, \nu P) &\leq \inf_{\alpha \in \Pi(\mu, \nu)} \int_{E \times E} W_d(P(x, \cdot), P(y, \cdot)) \alpha(\mathrm{d}x, \mathrm{d}y) \\ &\leq \inf_{\alpha \in \Pi(\mu, \nu)} \int_{E \times E} d(x, y) \alpha(\mathrm{d}x, \mathrm{d}y) = W_d(\mu, \nu) \ . \end{aligned}$$

□

We provide sufficient conditions for the existence of an invariant probability measure π for the Markov kernel P and for geometric ergodicity in Wasserstein distance, based on a Foster-Lyapounov drift condition on the product space $E \times E$ outside a coupling set. Consider the following assumption:

H1. *(a) There exists a measurable function $V : E \to [1, +\infty)$, $\lambda \in [0, 1)$ and $b \in \mathbb{R}$ such that for all $x \in E$,*

$$PV(x) \leq \lambda V(x) + b \ . \tag{3.4}$$

(b) For some $\delta > 0$, the subset

$$\Delta \overset{def}{=} \{(x, y) \in E \times E, V(x) + V(y) \leq (2b + \delta)/(1 - \lambda)\} \ , \tag{3.5}$$

is an (ϵ, d)-coupling set for a distance d topologically equivalent to $\mathbf{d}$ and bounded by 1.

The following Lemma shows that the drift condition (3.4) for P implies a drift condition for the coupling kernel $\mathbf{K}$ toward the coupling set Δ.

Lemma 3.6. *Assume **H1**. Then, for all $(x, y) \in E \times E$,*

$$PV(x) + PV(y) \leq \tilde{\lambda}\{V(x) + V(y)\} + 2b\mathbb{1}_\Delta(x, y) \ , \tag{3.6}$$

where Δ is defined in (3.5) and

$$\tilde{\lambda} = \frac{2b}{(2b + \delta)}(1 - \lambda) + \lambda < 1 \ . \tag{3.7}$$

Proof. Note that $\lambda < \tilde{\lambda} < 1$. For all $(x, y) \in E \times E$,

$$PV(x) + PV(y) \leq \tilde{\lambda}\{V(x) + V(y)\} - (\tilde{\lambda} - \lambda)\{V(x) + V(y)\} + 2b\,. \quad (3.8)$$

For $(x, y) \notin \Delta$, $V(x) + V(y) > (2b + \delta)/(1 - \lambda)$, so that

$$(\tilde{\lambda} - \lambda)\{V(x) + V(y)\} > \frac{2b}{(2b+\delta)}(1-\lambda)\frac{2b+\delta}{1-\lambda} = 2b\,.$$

The proof follows. □

Let $\mathbf{K}$ be the coupling kernel under which Δ is a (ϵ, d)-coupling set. Note that this implies that for any $n \in \mathbb{N}^\star$ and $x, y \in E$, $\mathbf{K}^n((x, y), \cdot)$ is a coupling of the probabilities $(P^n(x, \cdot), P^n(y, \cdot))$. Therefore, by (3.1),

$$W_d(P^n(x, \cdot), P^n(y, \cdot)) \leq \widetilde{\mathbb{E}}_{x,y}\left[d(X_n, Y_n)\right]$$

where $((X_n, Y_n), n \geq 0)$ is the canonical Markov chain on the product space $(E \times E)^{\mathbb{N}}$ equipped with the product σ-field with Markov kernel $\mathbf{K}$ and $\widetilde{\mathbb{E}}_{x,y}$ is the associated canonical expectation when the initial distribution is the Dirac mass at (x, y). We denote by $\{\tilde{\mathcal{F}}_n = \sigma((X_k, Y_k), k \leq n)\}$ the associated σ-field. Define by T_1 the first return time to the coupling set Δ,

$$T_1 \overset{\text{def}}{=} \inf\left\{n > 0 | (X_n, Y_n) \in \Delta\right\}\ , \quad (3.9)$$

with $\inf \emptyset = +\infty$. Define recursively the successive hitting times to Δ by

$$T_j = T_1 \circ \theta_{T_{j-1}} + T_{j-1}\ , j \geq 2\,, \quad (3.10)$$

where θ is the shift operator on the canonical space $(E \times E)^{\mathbb{N}}$. The following Proposition, adapted from [JT01a] relates the contraction of the Markov chain with the strict contraction coefficient in the coupling set and the number of visits to the coupling set.

Proposition 3.7. *Assume **H**1. Then, for all $x, y \in E$, and $n, m \in \mathbb{N}$, $m \geq 1$:*

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n, Y_n)\right] \leq (1-\epsilon)^{m-1} + \widetilde{\mathbb{P}}_{x,y}\left[T_m \geq n\right]\ . \quad (3.11)$$

Proof. Set $Z_n = d(X_n, Y_n)$; under **H** 1, for all $(x, y) \in E \times E$, $\widetilde{\mathbb{E}}_{x,y}\left[Z_1\right] \leq d(x, y)$. Therefore, $\{(Z_n, \tilde{\mathcal{F}}_n)\}_{n \geq 0}$ is a bounded non-negative supermartingale. Denote by Z_∞ its $\widetilde{\mathbb{P}}_{x,y}$-a.s limit. By the optional stopping theorem, we have for every $m \geq 0$:

$$\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_{m+1}} \middle| \tilde{\mathcal{F}}_{T_m+1}\right] \leq Z_{T_m+1}\ . \quad (3.12)$$

On the other hand, the strong Markov property imply for every $m \geq 0$

$$\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_m+1} \middle| \tilde{\mathcal{F}}_{T_m}\right] \leq (1-\epsilon) Z_{T_m}\ . \quad (3.13)$$

By (3.12) and (3.13), it yields $\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_{m+1}} \middle| \tilde{\mathcal{F}}_{T_m}\right] \leq (1-\epsilon) Z_{T_m}$. Under **H**1, Z_n is upper bounded by 1 and the proof follows from [JT01a, lemma 3.1]. □

The drift inequality allows to control the exponential moments of the return time to the coupling set Δ; this will allow us to control the number of returns to Δ of the coupled chain.

Proposition 3.8. *Assume* $\boldsymbol{H}1$. *Then,*

$$\widetilde{\mathbb{E}}_{x,y}\left[\widetilde{\lambda}^{-T_1}\right] \leq \begin{cases} V(x)+V(y) & (x,y)\notin\Delta \\ M & (x,y)\in\Delta\,. \end{cases} \tag{3.14}$$

where

$$M \stackrel{def}{=} \sup_{(x,y)\in\Delta}\{V(x)+V(y)\} + 2\tilde{\lambda}^{-1}b\,. \tag{3.15}$$

Corollary 3.9. *For all* $(x,y)\in E\times E$, *and all* $m\in\mathbb{N}$,

$$\widetilde{\mathbb{E}}_{x,y}\left[\widetilde{\lambda}^{-T_m}\right] \leq M^{m-1}\{V(x)+V(y)+2\tilde{\lambda}^{-1}b\}\,, \tag{3.16}$$

Proof. See [MT09, Chapter 15]. □

Combining the contraction inequality in Proposition 3.7 with the explicit control of the moment of the return times to the coupling set given in Proposition 3.8, we obtain an explicit expression of the contraction of the coupled Markov chain.

Proposition 3.10. *Assume* $\boldsymbol{H}1$. *Then, for all* $(x,y)\in E\times E$,

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right] \leq C\tau^n\{V(x)+V(y)\}\,,$$

with

$$\ln(\tau) = \ln(\widetilde{\lambda})\,\frac{-\ln(1-\epsilon)}{\ln(M)-\ln(1-\epsilon)}\,, \tag{3.17}$$

$$C = 1/2 + (1-\epsilon)^{-2}\,. \tag{3.18}$$

Proof. Proposition 3.7 and Corollary 3.9 imply that

$$\begin{aligned}\widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right] &\leq \left\{\widetilde{\mathbb{P}}_{x,y}\left[T_m\geq n\right] + (1-\epsilon)^{m-1}\right\} \\ &\leq \left\{M^m\widetilde{\lambda}^n + (1/2)(1-\epsilon)^{m-1}\right\}\{V(x)+V(y)\}\,.\end{aligned}$$

The proof is concluded by setting

$$m \stackrel{\text{def}}{=} \left\lfloor \frac{-n\ln(\widetilde{\lambda})}{\ln(M)-\ln(1-\epsilon)} \right\rfloor\,.$$

□

Using the contraction property, we may first establish the existence and uniqueness of a stationary distribution and then obtain an explicit control of convergence of the iterate of the Markov chain to stationarity.

Theorem 3.11. *Assume* $\boldsymbol{H}1$*. Then* P *admits a unique invariant probability* π *and for all* $x \in E$

$$W_d(P^n(x,\cdot),\pi) \leq C\,\tau^n V(x)\;, \tag{3.19}$$

with τ *given by* (3.17) *and*

$$C = (1/2 + (1-\epsilon)^{-2})(1 + b/(1-\lambda))\;.$$

Proof. Since d is continuous, according to [Vil09, Corollary 5.22], the function $(x,y) \mapsto W_d(P^n(x,\cdot),P^n(y,\cdot))$ is measurable.

We first show the uniqueness of the invariant probability. Assume that there exist two invariant distributions π and ν, and let ξ be a coupling of π and ν. According to Lemma 3.4, we have for every integer n :

$$W_d(\pi,\nu) = W_d(\pi P^n, \nu P^n) \leq \int_{E\times E} W_d(P^n(x,\cdot),P^n(y,\cdot))\,\xi(\mathrm{d}x,\mathrm{d}y)\;.$$

By Definition 3.1 and Proposition 3.10, there exist a constant C and $\tau \in (0,1)$, such that for all $x, y \in E$ and $n \geq 0$,

$$g_n(x,y) \overset{\text{def}}{=} W_d(P^n(x,\cdot),P^n(y,\cdot)) \leq C\tau^n\left(V(x)+V(y)\right)\;. \tag{3.20}$$

Eq. (3.20) shows that the sequence of functions $(g_n)_{n\in\mathbb{N}}$ converges pointwise to 0 and is bounded by 1. Therefore, by the Lebesgue dominated convergence theorem, we have:

$$\int_{E\times E} W_d(P^n(x,\cdot),P^n(y,\cdot))\,\xi(\mathrm{d}x,\mathrm{d}y) \underset{n\to+\infty}{\longrightarrow} 0\;,$$

showing that $W_d(\pi,\nu)=0$, or equivalently $\nu=\pi$ since W_d is a distance on $\mathcal{P}(E)$.

We now establish that under $\mathbf{H}1$, P admits one invariant probability measure. To that goal, let $x_0 \in E$, let us show that $\{P^n(x_0,\cdot), n\in\mathbb{N}\}$ is a Cauchy sequence for W_d. By Lemma 3.4, (3.20) and (3.4), for $n \in \mathbb{N}^*$ we have

$$\begin{aligned}
W_d(P^n(x_0,\cdot),P^{n+1}(x_0,\cdot)) &\leq \inf_{\xi\in\,\mathcal{C}_0} \int_{E\times E} W_d(P^n(z,\cdot),P^n(t,\cdot))\xi(\mathrm{d}z,\mathrm{d}t) \\
&\leq \inf_{\xi\in\,\mathcal{C}_0} \left\{ \int_{E\times E} C\tau^n\left(V(z)+V(t)\right)\xi(\mathrm{d}z,\mathrm{d}t) \right\} \\
&\leq C\tau^n\left(2V(x_0)+b\right)\;,
\end{aligned} \tag{3.21}$$

where $\mathcal{C}_0 \overset{\text{def}}{=} \Pi(\delta_{x_0}, P(x_0,\cdot))$. Therefore, the series

$$\sum_{n\geq 1} W_d(P^n(x_0,\cdot),P^{n+1}(x_0,\cdot))$$

is convergent which implies that $\{P^n(x_0,\cdot), n\in\mathbb{N}\}$ is a Cauchy sequence in $(\mathcal{P}(E),W_d)$. Since under $\mathbf{H}1$, $(\mathcal{P}(E),W_d)$ is Polish, there exists $\pi\in\mathcal{P}(E)$ such that $\lim_{n\to+\infty} W_d(P^n(x_0,\cdot),\pi) = 0$. The second step is to prove that π is invariant. As W_d is continuous on $\mathcal{P}(E)\times\mathcal{P}(E)$,

$$W_d(\pi,\pi P) = \lim_{n\to+\infty} W_d(P^n(x_0,\cdot),\pi P)\;.$$

By the triangle inequality and Lemma 3.4, it holds

$$\begin{aligned} W_d(\pi, \pi P) &\leq \lim_{n\to+\infty} \{W_d(P^n(x_0,\cdot), \delta_{x_0}P^n P) + W_d(\delta_{x_0}P^n P, \pi P)\} \\ &\leq \lim_{n\to+\infty} \{W_d(P^n(x_0,\cdot), P^{n+1}(x_0,\cdot)) + W_d(\delta_{x_0}P^n, \pi)\} \ . \end{aligned}$$

By definition of π and (3.21), the RHS is equals to 0, and therefore $\pi P = \pi$.

Finally, let us prove the claimed rate of convergence. Note that, the drift condition implies that $\pi(V) \leq b/(1-\lambda)$. Let $x \in E$, since π is invariant by Lemma 3.4 and Proposition 3.10,

$$\begin{aligned} W_d(P^n(x,\cdot), \pi) = W_d(P^n(x,\cdot), \pi P^n) &\leq \inf_{\xi\in\Pi(\delta_x,\pi)} \int_{E\times E} W_d(P^n(z,\cdot), P^n(t,\cdot))\xi(\mathrm{d}z, \mathrm{d}t) \\ &\leq C\left(1 + b/(1-\lambda)\right) V(x)\tau^n \ . \end{aligned}$$

□

3.2 Application to the EI-MALA algorithm

The Metropolis-Adjusted Langevin Algorithm (MALA), proposed by [RT96a], is a technique to sample high dimensional probability distributions. The MALA algorithm is a special instance of the Metropolis Hastings method. The main idea of MALA is to construct the proposal moves from the forward Euler discretization of the Langevin diffusion whose invariant measure is the target distribution. Let Q be a full-rank matrix and Υ be a gradient-Lipshitz convex function. Set

$$U(x) = (1/2)x^T Qx + \Upsilon(x) \ . \tag{3.22}$$

Let $\Gamma : \mathbb{R}^d \to \mathbb{R}$ be a gradient Lipshitz (not necessarily convex) function. Assume that $\mathcal{Z} = \int_{\mathbb{R}^d} \exp(-U(x) - \Gamma(x))\mathrm{d}x < \infty$ and let π denote the probability measure on $\mathbb{R}^d$ with density proportional to $\exp(-U - \Gamma)$. With a slight abuse in notations, we use the same letter π for the probability and its density, that is,

$$\pi(\mathrm{d}x) = \pi(x)\mathrm{d}x = \mathcal{Z}^{-1}\exp(-U(x) - \Gamma(x))\ \mathrm{d}x \ .$$

Below, we focus on the case where the potential Consider the over-damped Langevin stochastic differential equation

$$\begin{cases} \mathrm{d}Y_t &= -Y_t\mathrm{d}t - Q^{-1}\nabla\Upsilon(Y_t)\mathrm{d}t + \sqrt{2}Q^{-1/2}\mathrm{d}B_t \ , \\ Y_0 &= y_0 \ . \end{cases} \tag{3.23}$$

where $\{B_t, t \geq 0\}$ is the standard Brownian Motion. Because the gradient of U is locally Lipschitz and is at most of linear growth, (3.23) has a unique strong solution; see [RW00, Theorem 12.1]. A nice property of the Langevin diffusion making it interesting as a proposal mechanism for MCMC is that its stationary distribution has a density

$\exp(-U)$ w.r.t. the Lebesgue measure [Ken78, Theorem 10.1]. It is plain to see that, for any $t \geq 0$ and $\delta > 0$,

$$Y_{t+\delta} = Y_t \mathrm{e}^{-\delta} - \int_t^{t+\delta} \mathrm{e}^{-(t+\delta-s)} Q^{-1} \nabla \Upsilon(Y_s) \mathrm{d}s + \sqrt{2} Q^{-1/2} \int_t^{t+\delta} \mathrm{e}^{-(t+\delta-s)} \mathrm{d}B_s \,. \tag{3.24}$$

A forward Euler discretization of (3.24) yields to the following equation

$$\bar{Y}_{t+\delta} = \bar{Y}_t \mathrm{e}^{-\delta} - (1-\mathrm{e}^{-\delta}) Q^{-1} \nabla \Upsilon(\bar{Y}_t) + Z_{t,\delta} \,, \tag{3.25}$$

where $Z_{t,\delta}$ is a Gaussian variable with zero-mean and covariance $(1-\mathrm{e}^{-2\delta})Q^{-1}$. This discretization scheme is referred to as, for stochastic partial differential equation, the stochastic Euler exponential integrator; see [LR04] and the references therein.

Setting $h = 2(1-\mathrm{e}^{-\delta})$ and $t = 0$, $\bar{Y}_h(y) = \mathscr{O}_h(y, Z_0)$ where Z_0 is a standard Gaussian random variable and

$$\mathscr{O}_h(x,z) = x - (h/2) Q^{-1} \nabla U(x) + \tilde{h}\, Q^{-1/2} z \,, \tag{3.26}$$

with $\tilde{h} = \sqrt{h - h^2/4}$ yields to a proposal which can be used in a Metropolis-Hastings algorithm. This proposal was alluded to in [Bes+08] to sample diffusion bridge and further investigated in [Ebe14] (in these two works, Q is the identity matrix). The acceptance ratio is given by $\alpha_h(x,y) = \exp(-G_h(x,y)^+)$ where

$$\begin{aligned} G_h(x,y) &= \Upsilon(y) - \Upsilon(x) + \Gamma(y) - \Gamma(x) - \langle (y-x)/2, \nabla\Upsilon(x) + \nabla\Upsilon(y)\rangle \\ &+ \frac{h}{8-2h}\left[\langle y + x, \nabla\Upsilon(y) - \nabla\Upsilon(x)\rangle + \left\| Q^{-1/2}\nabla\Upsilon(y)\right\|^2 - \left\| Q^{-1/2}\nabla\Upsilon(x)\right\|^2 \right] . \end{aligned} \tag{3.27}$$

We denote by P the Markov kernel defined by Algorithm 1. As mentioned in the introduction, for $x, y \in \mathbb{R}^d$, the basic coupling between $P(x,\cdot)$ and $P(y,\cdot)$ is obtained by

Algorithm 1: EI-MALA Algorithm

Data: $h \in (0,2)$
Result: $(\mathbf{X}_n)_{n\in\mathbb{N}}$
begin
  Initialize $\mathbf{X}_0$
  for $n \geq 0$ **do**
    Generate $Z_n \sim \mathcal{N}(0, \mathrm{I}_d)$, and set
    $\mathbf{S}_n = (1-h/2)X_n + h/2 Q^{-1}\nabla\Upsilon(X_n) + \tilde{h}Q^{-1/2}Z_n$
    Generate $U_n \sim \mathcal{U}([0,1])$
    if $U_n \leq \alpha_h(\mathbf{X}_n, \mathbf{S}_n)$ **then**
      $\mathbf{X}_{n+1} = \mathbf{S}_n$
    else
      $\mathbf{X}_{n+1} = \mathbf{X}_n$

using the same Gaussian variable Z_0 and the same uniform U_0 to sample

$$\mathbf{X}_1 = x + \mathbb{1}(U_0 \leq \alpha_h(x, \mathscr{O}_h(x, Z_0)))(\mathscr{O}_h(x, Z_0) - x)$$
$$\mathbf{Y}_1 = y + \mathbb{1}(U_0 \leq \alpha_h(y, \mathscr{O}_h(y, Z_0)))(\mathscr{O}_h(y, Z_0) - y)$$

We denote by $\mathbf{K}_{\mathrm{M}}$ the Markov kernel induced by this construction. For Z_0 a standard Gaussian random variable, define

$$X_1(x) = \mathscr{O}_h(x, Z_0) \text{ and } Y_1(y) = \mathscr{O}_h(y, Z_0)\ , \tag{3.28}$$

which are the two proposals corresponding to $\mathbf{K}_{\mathrm{M}}$. To apply Theorem 3.11, we want to find a distance for which $X_1(x)$ is closer to $Y_1(y)$ than x to y. We could think to use the canonical euclidean norm. However, even assuming Υ is convex, it seems that such a contraction fails to be established since $\nabla\Upsilon$ is multiplied by Q^{-1} in (3.26). A better choice is the norm $\|\cdot\|_Q$ associated with the scalar product $\langle x, y\rangle_Q = \langle Qx, y\rangle$, for all $x, y \in \mathbb{R}^d$. Indeed, Lemma 3.17 establishes a contraction under the following standard assumptions.

M1. *(1) The function Υ belongs to $C^1(\mathbb{R}^d)$, is convex and there exists C_Υ such that for all $x, y \in \mathbb{R}^d$,*

$$\left\|Q^{-1}(\nabla\Upsilon(x) - \nabla\Upsilon(y))\right\|_Q \leq C_\Upsilon \|x - y\|_Q\ . \tag{3.29}$$

(2) The function Γ belongs to $C^1(\mathbb{R}^d)$ and there exists C_Γ such that for all $x, y \in \mathbb{R}^d$,

$$\left\|Q^{-1}(\nabla\Gamma(x) - \nabla\Gamma(y))\right\|_Q \leq C_\Gamma \|x - y\|_Q\ . \tag{3.30}$$

Eq. (3.29) and (3.30) are not restrictive since it holds if $\nabla\Upsilon$ and $\nabla\Gamma$ are $\|\|$-Lipschitz continuous, since in $\mathbb{R}^d$ all the norms are equivalent. Here is the other main assumption we make to establish our results.

M2. *There exists $h_\ell \in (0, 2)$ such that for all $h \in (0, h_\ell]$ there exists three positive real numbers a_h, R_h and r_h such that for all $x \in \mathbb{R}^d$, $\|x\|_Q \geq R_h$,*

$$\inf\left\{\alpha_h(x, z)\ ,\ z \in \mathrm{B}_Q\left(\mathscr{O}_h(x, 0), r_h\right)\right\} > a_h\ . \tag{3.31}$$

We first establish the drift condition (a) of **H**1.

Proposition 3.12. *Assume **M**1 and **M**2, and let $h \in (0, h_\ell \wedge (4/(C_\Upsilon^2 + 1)))$. Set $\mathscr{V}(x) = 1 \vee \|x\|_Q$. Then there exist $b \in \mathbb{R}_+$ and $\lambda \in (0, 1)$ such that for all $x \in \mathbb{R}^d$*

$$P\mathscr{V}(x) \leq \lambda\mathscr{V}(x) + b\mathbb{1}_{\{\mathscr{V} \leq u\}}(x)\ .$$

Proof. The proof is postponed to Section 3.4. □

Now, let us deal with the condition (b) of **H**1. The main difficulty is to control the probability that one proposal is accepted and the other not. We can see in Lemma 3.19 that this probability is not globally Lipschitz in function of x, y but is locally Lipschitz continuous with Lipschitz constant which grows linearly with $\|x\|_Q \vee \|y\|_Q$. To get the contraction of **H**1-(b), we need to define distances on $\mathbb{R}^d$ which are in some sense equivalent to $(x,y) \mapsto \|x-y\|_Q (1 + \|x\|_Q \vee \|y\|_Q)$. That is why in the following we construct such distances, by adapting arguments first introduced in [HSV14]. For $x, y \in \mathbb{R}^d$ and $T > 0$, let us define

$$\mathscr{F}^T_{x,y} = \left\{ \psi \in C^1([0,T], \mathbb{R}^d), \psi(0) = x,\ \psi(T) = y, \|\psi'(s)\|_Q = 1\ \forall s \in [0,T] \right\} ,$$

and for $\epsilon, \eta > 0$, $d_{\eta,\epsilon} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ defined for $x, y \in \mathbb{R}^d$ by

$$d_{\eta,\epsilon}(x,y) = 1 \wedge \epsilon^{-1} \times \inf \left\{ \eta \int_0^T \|\psi(s)\|_Q \, \mathrm{d}s + T; T \in \mathbb{R}_+, \psi \in \mathscr{F}^T_{x,y} \right\} . \tag{3.32}$$

It is proved in Lemma 3.20 that $d_{\eta,\epsilon}$ is a metric on $\mathbb{R}^d$, topologically equivalent to $\|\cdot\|_Q$; see Section 3.4. The level sets of the function $\mathscr{V}$ are the ball associated to $\|\cdot\|_Q$. We show in the next proposition that they are coupling sets for some iterate of P.

Proposition 3.13. *Assume **M**1 and **M**2.*
For all $h \in (0, h_\ell \wedge (4/(C_\Upsilon^2 + 1)))$, there exist two constants $\epsilon, \eta > 0$ satisfying the following property: for all $R > 0$, one may find $n \in \mathbb{N}^*$, $\tau \in (0,1)$ such that $\mathrm{B}_Q(0,R)^2$ is a $(\tau, d_{\eta,\epsilon})$-coupling set for P^n, with $d_{\eta,\epsilon}$ given by (3.32).

Proof. The proof is postponed to Section 3.4. □

By Proposition 3.12 and Proposition 3.13, there exists $N \in \mathbb{N}$ such that P^N satisfies **H**1. Therefore we deduce from by Theorem 3.11 the following result.

Theorem 3.14. *Assume **M**1, **M**2 and let $h \in (0, h_\ell \wedge (4/(C_\Upsilon^2 + 1)))$. Then, there exist positive constants ϵ, η and C, $\rho \in (0,1)$ (whose explicit expressions are given in the proof) such that for all $x \in \mathbb{R}^d$ and $n \in \mathbb{N}^*$*

$$W_{d_{\eta,\epsilon}}(P^n(x,\cdot), \pi) \le C\rho^n \{\mathscr{V}(x) + \mathscr{V}(y)\} ,$$

with $\mathscr{V}(x) = 1 \vee \|x\|_Q$ and $d_{\eta,\epsilon}$ is given by (3.32).

Example 3.15 (Bounded perturbations of Gaussian distributions)**.** *To illustrate our bounds, assume that $\Upsilon \equiv 0$ and that Γ is bounded on $\mathbb{R}^d$ by M_Γ and gradient Lipschitz. It is easily checked that **M**1 and **M**2 are satisfied. We can provide an explicit expression of the constant appearing in Theorem 3.14. The constants appearing in the drift condition Proposition 3.12 might be chosen to be:*

$$\lambda = (1 - (1 - h/2)\mathrm{e}^{-2M_\Gamma}) \text{ and } b = \tilde{h}\sqrt{d}$$

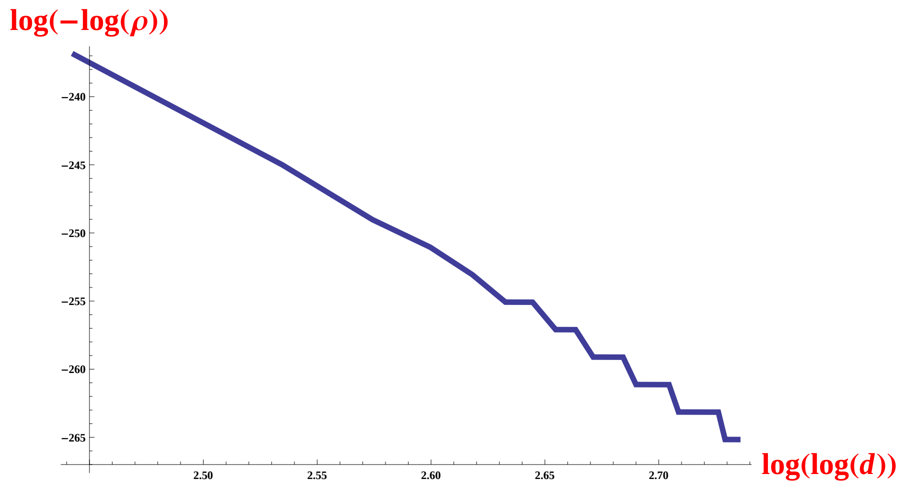


Figure 3.1: Evolution of the rate of convergence ρ given by Theorem 3.14 in function of the dimension d.

So, to apply Theorem 3.11 we need to consider the balls $B(0, R)$ for $R = (2b+\delta)(1-\lambda)^{-2}$ and arbitrary positive δ. Proposition 3.13 gives that $B(0, R)$ is $(\tau, d_{\eta,\epsilon})$-coupling set for P^n with

$$\begin{aligned}
\eta &= (2((1 - h/2) + \tilde{h}\sqrt{d}))^{-1} \\
\epsilon &= (1 - h/2)(4C_\Gamma(\eta^{-1} + 1 + \sqrt{d}))^{-1} \\
n &= \lceil (\log(\epsilon/2) - \log(2R(2\eta R + 1)))/\log((1 - h/2)) \rceil \vee 1 \\
\tau &= (1/2)\mathrm{e}^{-2M_\Gamma n}\mathbb{P}\left(\|Z_0\| \leq R/(n\tilde{h})\right)^n
\end{aligned}$$

Finally explicit expressions of C and ρ Theorem 3.14 are given by Theorem 3.11. We can see on Figure 3.1 that $\log(-\log(\rho))$ *depends linearly on* $\log(\log(d))$.

3.3 Simulation

We now illustrate our results with a Monte Carlo experiment. We have considered for simplicity an ill-conditioned Bayesian linear inverse problem; see [KS05] and the references therein. The aim is to invert $b \approx Ax$, where A is a linear operator from $\mathbb{R}^d$ to $\mathbb{R}^p$, assuming that the observations b are contaminated by some additive Gaussian noise, assumed, for simplicity, to have zero-mean and unit covariance matrix. In this problem, the dimension d can be very large [1] and $p << d$; see [Stu10], [Cot+13], [Das+13] and the references therein. For conciseness and simplicity, we study a toy inverse problems in which the prior distribution π_X of the parameter of interest x is given to be a small

[1]the underlying problem is typically infinite-dimensional; the problem is finite dimensional after truncation

pertubation of a exponential power distribution (see [BTbk]):

$$\pi_X(x) \propto \exp\left(-\lambda_1(x^Tx+\delta)^\beta - (\lambda_2/2)(x^Tx)\right) ,$$

with $\beta \in (1/2, 1)$, $\lambda_1, \lambda_2, \delta \in \mathbb{R}_+^*$. In this setting, the posterior distribution π is proportional to $\exp(-U)$ on $\mathbb{R}^d$, where $\Gamma = 0$ and the potential U is on the form (3.22),

$$Q = A^TA + \lambda_2 \mathrm{I}_d \text{ and } \Upsilon(x) = \lambda_1(x^Tx+\delta)^\beta - \langle b, Ax\rangle \; . \tag{3.33}$$

Lemma 3.16. *Let $\pi(x) \propto \exp\left\{-(1/2)x^TQx - \Upsilon(x)\right\}$, with Q and Υ given by (3.33). Then* ***M1*** *and* ***M2*** *are satisfied.*

Proof. First since for $x \in \mathbb{R}^d$,

$$\nabla\Upsilon(x) = \lambda_1\beta(x^Tx+\delta)^{\beta-1}x - A^Tb \, , \tag{3.34}$$

it is straightforward to see that Υ satisfied **M1** with $C_\Upsilon = |||Q^{-1}|||\,\beta\lambda_1/\delta$. To show that **M2** is satisfied, it is enough to show that for h small enough, there exist $r_z, R > 0$ such that $G_h(x, y_x)$, given by (3.27) are bounded from above for

$$y_x = (1-h/2)x - (h/2)Q^{-1}\nabla\Upsilon(x) + \tilde{h}z \text{ with } z \in \mathrm{B}_Q(0, r_z) \; ,$$

and x is outside a ball $\mathrm{B}_Q(0, R)$. To do so, since all the norm are equivalent in finite dimension, we prove in fact that for all h small enough $\lim_{\|x\|\to+\infty} G_h(x, y_x) = -\infty$. Indeed we separately study three terms in the expression of G_h. Since $\lim_{\|x\|\to+\infty}\langle x, \nabla\Upsilon(x)\rangle/(x^Tx) = 0$, we have the following asymptotic relations:

$$\Upsilon(y_x) - \Upsilon(x) - \langle (y_x - x)/2, \nabla\Upsilon(y_x) - \nabla\Upsilon(x)\rangle \underset{\|x\|\to+\infty}{\sim} \lambda_1\phi(h)(x^Tx)^\beta \tag{3.35}$$

with

$$\phi(h) = (1-(h/2))^{2\beta} - 1 + (h/2)\beta + (h/2)\beta(1-(h/2))^{2\beta-1} \, . \tag{3.36}$$

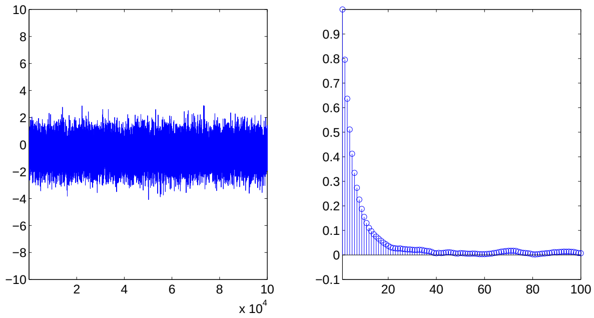
$$\langle y_x + x, \nabla\Upsilon(y_x) - \nabla\Upsilon(x)\rangle \underset{\|x\|\to+\infty}{\sim} \lambda_1\beta\left((1-(h/2))^\beta - 1\right)(x^Tx)^\beta \tag{3.37}$$

$$\begin{aligned}&\left\|Q^{-1/2}\Upsilon(y_x)\right\|^2 - \left\|Q^{-1/2}\Upsilon(x)\right\| \\ &\underset{\|x\|\to+\infty}{\sim} \lambda_1\beta((1-(h/2))^{2(\beta-1)} - 1)(x^Tx)^{2(\beta-1)}\,x^TQx \, .\end{aligned} \tag{3.38}$$

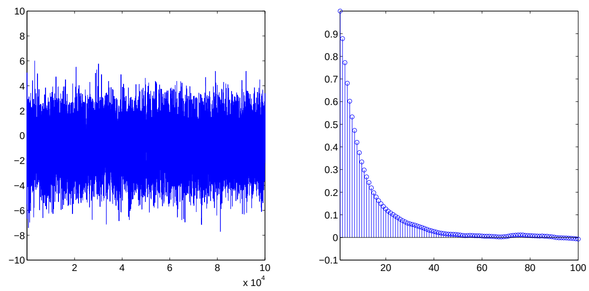
Then the terms (3.37) and (3.38) tends to $-\infty$ for all $h \in (0, 2)$. It remains to show that for h small enough, it is the case for (3.35). Therefore, we compute the asymptotic expansion of ϕ given by (3.36),

$$\phi(h) \underset{h\to 0^+}{\sim} \beta(2\beta-1)(2\beta-2)(\beta/2-\beta/3)(h/2)^3 \, .$$

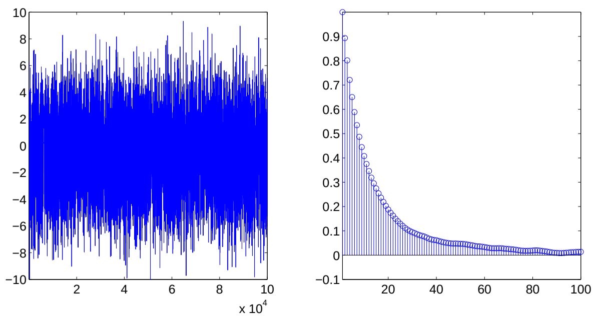
As $\beta \in (1/2, 1)$, ϕ is negative for all h small enough, which concludes the proof. $\square$



(a) $d = 100$

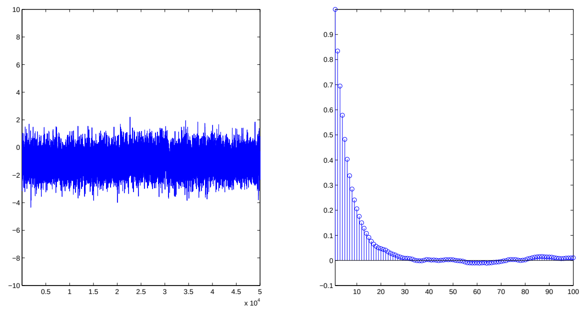


(b) $d = 500$

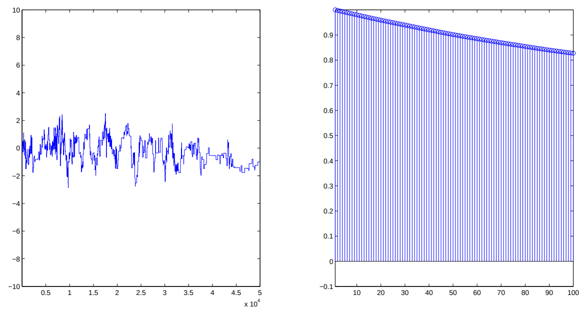


(c) $d = 1000$

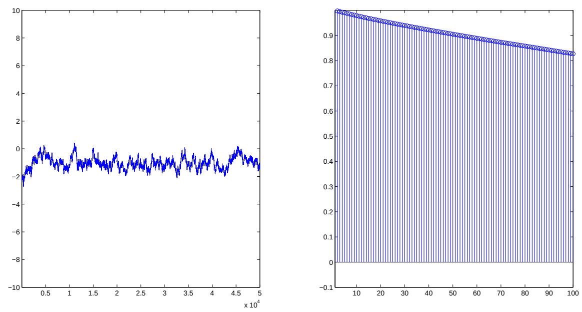
Figure 3.3: Trace plot and auto-correlation in function of the dimension on 10000 iterations with a 10000 burn-in iterations .



(a) EI-MALA



(b) pCN



(c) MALA

Figure 3.5: Trace plot and auto-correlation in dimension 100 for EI-MALA, pCN and MALA on 50000 iterations with a 10000 burn-in iterations.

We have displayed in Figure 3.3 the trace plots for the first coordinate for different dimensions ($d = 10$, $d = 100$ and $d = 1000$). The autocorrelations is monotone decreasing for reversible Markov chain and its integral determines the asymptotic variance of the sample average. We can see that the mixing time is not significantly affected by the dimension, which is consistent with our findings (see Theorem 3.14).

Also, we have compared EI-MALA to the pre conditioned Crank Nicolson algorithm (pCN) proposed by [Cot+13] and to the naive MALA. It can be seen on Figure 3.5 that the mixing time of EI-MALA is significantly better than for the two other algorithms.

3.4 Proofs

Lemma 3.17. *Assume* ***M**1.*

(a) *For all $x, y, z \in \mathbb{R}^d$,*

$$\begin{aligned}
&\langle \nabla\Upsilon(x) - \nabla\Upsilon(y), z\rangle \leq C_\Upsilon \|x-y\|_Q \|z\|_Q \\
&\left\|Q^{-1}\nabla\Upsilon(x)\right\|_Q \leq C_\Upsilon \|x\|_Q + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q \\
&\langle \nabla\Upsilon(z), x-y\rangle \leq C_\Upsilon(C_\Upsilon \|z\|_Q + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q) \|x-y\|_Q \ .
\end{aligned}$$

(b) *For all $x, y \in \mathbb{R}^d$ and $h \leq 4/(C_\Upsilon^2 + 1)$,*

$$\left\|x - (h/2)Q^{-1}\nabla U(x) - \{y - (h/2)Q^{-1}\nabla U(y)\}\right\|_Q \leq \nu \|x-y\|_Q \ , \tag{3.39}$$

where

$$\nu = \left(1 - h(1 - h(1 + C_\Upsilon^2)/4)\right)^{1/2} \ . \tag{3.40}$$

In particular,

$$\left\|x - (h/2)Q^{-1}\nabla U(x)\right\|_Q \leq \nu \|x\|_Q + (h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q \ . \tag{3.41}$$

Proof. (a) is just a consequence of the definition of $\langle\cdot,\cdot\rangle_Q$, **M**1, the Cauchy–Schwarz inequality and the triangle inequality.
(b) Let $h \leq 4/(C_\Upsilon^2+1)$. On **M**1, since Υ is convex and C^1, for all $x, y \in \mathbb{R}^d$,

$$\begin{aligned}
&\left\|x - y - (h/2)Q^{-1}(\nabla U(x) - \nabla U(y))\right\|_Q^2 = (1-h/2)^2 \|x-y\|_Q^2 \\
&+ (h^2/4)\left\|Q^{-1}(\nabla\Upsilon(x) - \nabla\Upsilon(y))\right\|_Q^2 - h(1-h/2)\langle\nabla\Upsilon(x) - \nabla\Upsilon(y), x-y\rangle \\
&\qquad\qquad \leq \left(1 - h(1 - h(C_\Upsilon^2+1)/4)\right)\|x-y\|_Q^2 \ ,
\end{aligned}$$

showing (3.39). Eq. (3.41) follows from (3.39) and the triangle inequality. □

Lemma 3.18. *Assume* ***M**1.*

(a) *For all* $x, y \in \mathbb{R}^d$,

$$\|X_1(x) - Y_1(y)\|_Q \le \nu \|x - y\|_Q$$
$$\|X_1(x)\|_Q \le \nu \|x\|_Q + (h/2) \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \widetilde{h} \left\|Q^{-1/2} Z_0\right\|_Q ,$$

where $(X_1(x), Y_1(y))$ *and* ν *are given by* (3.28) *and* (3.40)*, respectively.*

(b) *For all* $x, y \in \mathbb{R}^d$

$$\left\|Q^{-1}(\nabla\Upsilon(X_1(x)) - \nabla\Upsilon(Y_1(y)))\right\|_Q \le C_\Upsilon \nu \|x - y\|_Q$$
$$\left\|Q^{-1}\nabla\Upsilon(X_1(x))\right\|_Q \le C_\Upsilon \left(\nu \|x\|_Q + (h/2) \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \widetilde{h} \|Z_0\|\right) .$$

Proof. (a) is just a consequence of Lemma 3.17-(b). Then (b) follows from **M** 1 and Lemma 3.17-(a). □

Lemma 3.19. *Assume* ***M*** *1. There exists* C *such that*

$$|G_h(x, X_1(x)) - G_h(y, Y_1(y))| \le C \|x - y\|_Q \times \left\{\|x\|_Q \vee \|y\|_Q + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \|Z_0\|\right\} . \tag{3.42}$$

Proof. Let us write

$$|G_h(x, X_1(x)) - G_h(y, Y_1(y))| \le \sum_{i=1}^{4} I_i .$$

Using **M** 1, Lemma 3.17 and Lemma 3.18 we have the following inequalities for I_i, $i = 1 \cdots 4$.

$$\begin{aligned}
I_1 &\\
= &|\Upsilon(x) - \Upsilon(y)| + |\Upsilon(X_1(x)) - \Upsilon(Y_1(y))| + |\Gamma(x) - \Gamma(y)| + |\Gamma(X_1(x)) - \Gamma(Y_1(y))| \\
&\le \left|\int_0^1 \langle \nabla\Upsilon(tx + (1-t)y), x - y\rangle \,\mathrm{d}t\right| \\
&\quad + \left|\int_0^1 \langle \nabla\Upsilon(tX_1(x) + (1-t)Y_1(y)), X_1(x) - Y_1(y)\rangle \,\mathrm{d}t\right| \\
&\quad + \left|\int_0^1 \langle \nabla\Gamma(tx + (1-t)y), x - y\rangle \,\mathrm{d}t\right| \\
&\quad + \left|\int_0^1 \langle \nabla\Gamma(tX_1(x) + (1-t)Y_1(y)), X_1(x) - Y_1(y)\rangle \,\mathrm{d}t\right| \\
&\le C_1 \|x - y\|_Q \left\{\|x\|_Q \vee \|y\|_Q + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \|Z_0\|\right\} .
\end{aligned}$$

$$
\begin{aligned}
&I_2 \\
&= (1/2) |\langle x - X_1(x), \nabla\Upsilon(x) + \nabla\Upsilon(X_1(x))\rangle - \langle y - Y_1(y), \nabla\Upsilon(y) + \nabla\Upsilon(Y_1(y))\rangle| \\
&= (1/2) |\langle x - y, \nabla\Upsilon(x) + \nabla\Upsilon(X_1(x))\rangle + \langle Y_1(y) - X_1(x), \nabla\Upsilon(x) + \nabla\Upsilon(X_1(x))\rangle \\
&\qquad - \langle y - Y_1(y), \nabla\Upsilon(y) - \nabla\Upsilon(x)\rangle - \langle y - Y_1(y), \nabla\Upsilon(Y_1(y)) - \nabla\Upsilon(X_1(x))\rangle| \\
&\leq C_2 \|x - y\|_Q \left\{ \|x\|_Q \vee \|y\|_Q + \left\| Q^{-1}\nabla\Upsilon(0) \right\|_Q + \|Z_0\| \right\} .
\end{aligned}
$$

Denote $\widehat{h} = h/(8-2h)$. Then,

$$
\begin{aligned}
&I_3 \\
&= \widehat{h} |\langle X_1(x) + x, \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(x)\rangle - \langle Y_1(y) + y, \nabla\Upsilon(Y_1(y)) - \nabla\Upsilon(y)\rangle| \\
&= \widehat{h} |\langle X_1(x) - Y_1(x), \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(x)\rangle + \langle x - y, \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(x)\rangle \\
&\qquad + \langle Y_1(y) + y, \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(Y_1(y))\rangle + \langle Y_1(y) + y, \nabla\Upsilon(y) - \nabla\Upsilon(x)\rangle| \\
&\leq C_3 \widehat{h} \|x - y\|_Q \left\{ \|x\|_Q \vee \|y\|_Q + \left\| Q^{-1}\nabla\Upsilon(0) \right\|_Q + \|Z_0\| \right\} .
\end{aligned}
$$

$$
\begin{aligned}
I_4 &= \widehat{h} \left| \left\| Q^{-1/2}\nabla\Upsilon(y) \right\|_Q^2 - \left\| Q^{-1/2}\nabla\Upsilon(x) \right\|_Q^2 + \left\| Q^{-1/2}\nabla\Upsilon(X_1(x)) \right\|_Q^2 \right. \\
&\qquad\qquad \left. - \left\| Q^{-1/2}\nabla\Upsilon(Y_1(y)) \right\|_Q^2 \right| \\
&= \widehat{h} \left| \left\langle Q^{-1}(\nabla\Upsilon(y) + \nabla\Upsilon(x)), \nabla\Upsilon(y) - \nabla\Upsilon(x) \right\rangle \right. \\
&\qquad + \left\langle Q^{-1}\nabla\Upsilon(X_1(x)), \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(Y_1(y)) \right\rangle \\
&\qquad \left. + \left\langle Q^{-1}\nabla\Upsilon(Y_1(y)), \nabla\Upsilon(X_1(x)) - \nabla\Upsilon(Y_1(y)) \right\rangle \right| \\
&\leq C_4 \widehat{h} \|x - y\|_Q \left\{ \|x\|_Q \vee \|y\|_Q + \left\| Q^{-1}\nabla\Upsilon(0) \right\|_Q + \|Z_0\| \right\} .
\end{aligned}
$$

□

Proof of Proposition 3.12. Set $R = \nu^{-1} \vee R_h$. First, by Lemma 3.18-(a) and **M**1,

$$
\begin{aligned}
\sup_{x \in \mathrm{B}_Q(0,R)} P\mathscr{V}(x) &\leq \sup_{x \in \mathrm{B}_Q(0,R)} \mathbb{E}\left[1 \vee \|x\|_Q \vee \|X_1(x)\|_Q \right] \\
&\leq 1 \vee \left(R + (h/2) \left\| Q^{-1}\nabla\Upsilon(0) \right\|_Q + \mathbb{E}\left[\|Z_0\| \right] \right) < +\infty
\end{aligned} \tag{3.43}
$$

Next, if $x \notin \mathrm{B}_Q(0, R)$, let $X_1(x)$ defined by (3.28) and U be a uniform random variable on $[0, 1]$. Set

$$
\begin{aligned}
\mathscr{A}(x) &= \{\alpha_h(x, X_1(x)) \leq U\} \\
\mathscr{I} &= \left\{ \widetilde{h} \|Z_0\| \leq r_h \right\} .
\end{aligned}
$$

On $\mathscr{A}(x)$, the proposal is accepted and $\mathbf{X}_1 = X_1(x)$. On this complement, $\mathbf{X}_1 = x$. Then by Lemma 3.18-(a) and since $\|x\|_Q \geq \nu \|x\|_Q \geq 1$,

$$
\begin{aligned}
P\mathscr{V}(x) &\leq \mathbb{E}\left[\|X_1(x)\|_Q \mathbb{1}_{\mathscr{A}(x)\cap\mathscr{I}}\right] + \mathbb{E}\left[\|x\|_Q \mathbb{1}_{\mathscr{A}(x)^c\cap\mathscr{I}}\right] + \mathbb{E}\left[\|x\|_Q \vee \|X_1(x)\|_Q \mathbb{1}_{\mathscr{I}^c}\right] \\
&\leq \nu \|x\|_Q \mathbb{P}\left(\mathscr{A}(x)\cap\mathscr{I}\right) + \|x\|_Q \mathbb{P}\left(\mathscr{A}(x)^c\cap\mathscr{I}\right) + \|x\|_Q \mathbb{P}\left(\mathscr{I}^c\right) \\
&\qquad + (h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \tilde{h}\mathbb{E}\left[\|Z_0\|\right] \\
&\leq \lambda \|x\|_Q + (h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \tilde{h}\mathbb{E}\left[\|Z_0\|\right] , \qquad (3.44)
\end{aligned}
$$

with $\lambda = \mathbb{P}(\mathscr{I})(1-(1-\nu)\mathbb{P}(\mathscr{A}(x)\,|\mathscr{I}))+\mathbb{P}(\mathscr{I}^c)$. Since by **M**2, $\mathbb{P}(\mathscr{I})$ and $\mathbb{P}(\mathscr{A}(x)\,|\mathscr{I})$ are positive, $\lambda \in (0,1)$. In addition on **M**1, $(h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \tilde{h}\mathbb{E}\left[\|Z_0\|\right]$ is finite. Therefore, the proof is concluded combining (3.43) and (3.44). □

Lemma 3.20. *Let $x, y \in \mathbb{R}^d$, $T, \epsilon, \eta \in \mathbb{R}_+^*$ and $\psi \in \mathscr{F}_{x,y}^T$. Denote*

$$
\delta_\epsilon(x,y) \overset{def}{=} \epsilon/(\eta(\|x\|_Q \vee \|y\|_Q - \epsilon) \vee 0 + 1) . \qquad (3.45)
$$

(i) $T \geq \|x-y\|_Q$.

(ii) *If* $\eta \int_0^T \|\psi(s)\|_Q \,\mathrm{d}s + T < \epsilon$, *then* $T \leq \delta_\epsilon(x,y) \leq \epsilon$.

(iii) $d_{\eta,\epsilon}(x,y) \leq \epsilon^{-1} \|x-y\|_Q \left(\eta\{\|x\|_Q \vee \|y\|_Q\} + 1\right)$

(iv) *If* $d_{\eta,\epsilon}(x,y) < 1$, *then*

$$
d_{\eta,\epsilon}(x,y) \geq \epsilon^{-1} \|x-y\|_Q \left((\eta\{\|x\|_Q \vee \|y\|_Q\} - \delta_\epsilon(x,y)) \vee 0 + 1\right) .
$$

(v) *If* $d_{\eta,\epsilon}(x,y) < 1$ *then*

$$
\begin{aligned}
d_{\eta,\epsilon}(X_1(x), Y_1(y))/d_{\eta,\epsilon}(x,y) &\leq \Big(\|x-(h/2)\nabla U(x) - (y-(h/2)\nabla U(y))\|_Q \\
&\quad \times \left\{\eta\|x-(h/2)\nabla U(x)\|_Q \vee \|y-(h/2)\nabla U(y)\|_Q + \eta\tilde{h}\|Z_0\|_Q + 1\right\}\Big) \\
&\quad \times \left(\|x-y\|_Q \{\eta\{\|x\| \vee \|y\| - \delta_\epsilon(x,y)\} \vee 0 + 1\}\right)^{-1} ,
\end{aligned}
$$

*and under **M**1 for all $h \leq 4/(C_\Upsilon^2+1)$,*

$$
\begin{aligned}
&d_{\eta,\epsilon}(X_1(x), Y_1(y))/d_{\eta,\epsilon}(x,y) \\
&\leq \left(\nu\left\{\eta\nu\{\|x\|_Q \vee \|y\|_Q\} + \eta(h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \tilde{h}\eta\|Z_0\| + 1\right\}\right) \\
&\quad \times (\eta\{\|x\| \vee \|y\| - \delta_\epsilon(x,y)\} \vee 0 + 1)^{-1} ,
\end{aligned}
$$

with ν given by (3.40).

Proof of Lemma 3.20. (i) By definition of $\mathscr{F}^T_{x,y}$,

$$\|x-y\|_Q = \|\psi(T)-\psi(0)\|_Q \leq \int_0^T \|\psi'(s)\|_Q \, ds = T \ .$$

(ii) First, it is straightforward to see $T<\epsilon$. In addition, for all $s \in [0,T]$,

$$\|\psi(s)\|_Q \geq \Big|\|x\|_Q - s\Big| \vee \Big|\|y\|_Q - (T-s)\Big| \geq \Big|\|x\|_Q \vee \|y\|_Q - \epsilon\Big| \ . \tag{3.46}$$

Then the result follows from integrating the inequality between 0 and T and using the assumption.
(iii) It suffices to consider the particular case, $T = \|y-x\|_Q$ and $\psi \in \mathscr{F}^T_{x,y}$ defined for $s \in [0,T]$ by

$$\psi(s) = x + (y-x)s/\|x-y\|_Q \ .$$

(iv) Let $\{(T_n,\psi_n); n \in \mathbb{N}\}$ such that

$$\lim_{n\to+\infty} \epsilon^{-1}\left(\eta\int_0^{T_n} \|\psi_n(s)\|_Q \, ds + T_n\right) = d_{\eta,\epsilon}(x,y) \tag{3.47}$$

$$\epsilon^{-1}\left(\eta\int_0^{T_n} \|\psi_n(s)\|_Q \, ds + T_n\right) < 1 \quad \forall n$$

Then using (i)-(ii) and (3.46), for all n it holds

$$\epsilon^{-1}\left(\eta\int_0^{T_n} \|\psi_n(s)\|_Q \, ds + T_n\right) \\ \geq \epsilon^{-1}\|x-y\|_Q\Big((\eta\{\|x\|_Q \vee \|y\|_Q\} - \delta_\epsilon(x,y)) \vee 0 + 1\Big) \ .$$

The result now follows from (3.47).
(v) The claimed inequalities come from (iii)-(iv) with the definition of the basic coupling (3.28) for the first one, and using Lemma 3.18-(a) for the second. □

Lemma 3.21. *Assume **M**1 and **M**2.*
Then for all $h \in (0, h_\ell \wedge (4/C_U))$, there exists $\epsilon, \eta, \tau > 0$ such that for all $x,y \in \mathbb{R}^d$, $d_{\eta,\epsilon}(x,y) < 1$,

$$\mathbf{K}_{\mathrm{M}} d_{\eta,\epsilon}(x,y) \leq (1-\tau) d_{\eta,\epsilon}(x,y) \ ,$$

with $d_{\eta,\epsilon}$ given by (3.32). In particular, for all $x,y \in \mathbb{R}^d$,

$$\mathbf{K}_{\mathrm{M}} d_{\eta,\epsilon}(x,y) \leq d_{\eta,\epsilon}(x,y) \ .$$

Proof. For ease of notation, we simply write $\mathbf{K}$ for $\mathbf{K}_{\mathrm{M}}$. Let $x,y \in \mathbb{R}^d$, $d_{\eta,\epsilon}(x,y) < 1$. Then, if $\|y\|_Q \leq R_h$,

$$\|x\|_Q \leq \|x-y\|_Q + \|y\|_Q \leq \epsilon + R_h \ ,$$

where we used Lemma 3.20-(iv). Therefore, as we will choose ϵ small enough we can assume $\epsilon \leq 1$, and we end up with two cases: either $x,y \in \mathrm{B}_Q(0, R_h+2)$ or $x,y \notin$

$\mathrm{B}_Q(0, R_h)$. Let $(\mathbf{X}_1, \mathbf{Y}_1)$ be the basic coupling between $P(x,\cdot)$ and $P(y,\cdot)$; let Z_0, U be resp. the Gaussian variable and the uniform variable used for the basic coupling. Let $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ be the probability and the expectation over Z_1 and U_1. Set

$$\begin{aligned}\mathscr{I} &= \left\{\widetilde{h}\,\|Z_0\| \leq r_h\right\}\\ \mathscr{A}(x,y) &= \{\alpha_h(x, X_1(x)) \wedge \alpha_h(y, Y_1(y)) > U\}\\ \mathscr{R}(x,y) &= \{\alpha_h(x, X_1(x)) \vee \alpha_h(y, Y_1(y)) < U\}\ .\end{aligned}$$

On the event $\mathscr{A}(x,y)$, the moves are both accepted so that $\mathbf{X}_1 = X_1(x)$ and $\mathbf{Y}_1 = Y_1(y)$; On the event $\mathscr{R}(x,y)$, the moves are both rejected so that $\mathbf{X}_1 = x$ and $\mathbf{Y}_1 = y$. Then for all event $\mathscr{I}$, it holds,

$$\begin{aligned}&\mathbf{K}d_{\eta,\epsilon}(x,y) \leq \mathbb{E}\left[d_{\eta,\epsilon}(\mathbf{X}_1, \mathbf{Y}_1)\right]\\ &\leq \mathbb{E}\left[d_{\eta,\epsilon}(X_1(x), Y_1(y))\mathbb{1}_{\mathscr{A}(x,y)\cap\mathscr{I}}\right] + \mathbb{E}\left[d_{\eta,\epsilon}(x,y)\mathbb{1}_{\mathscr{R}(x,y)\cap\mathscr{I}}\right]\\ &+ \mathbb{E}\left[d_{\eta,\epsilon}(x,y) \vee d_{\eta,\epsilon}(X_1(x), Y_1(y))\mathbb{1}_{\mathscr{I}^c}\right] + \mathbb{E}\left[|\alpha_h(x, X_1(x)) - \alpha_h(y, Y_1(y))|\right]\ ,\end{aligned} \tag{3.48}$$

where we have used $d_{\eta,\epsilon}$ is bounded by 1. First, by Lemma 3.20-(v) since $\delta_\epsilon(x,y) \leq \epsilon \leq 1$, there exist $\eta_1, \tau_1 > 0$ such that for all $\eta < \eta_1$

$$\mathbb{E}\left[d_{\eta,\epsilon}(X_1(x), Y_1(y))\mathbb{1}_{\mathscr{A}(x,y)\cap\mathscr{I}}\right] \leq (1-\tau_1)d_{\eta,\epsilon}(x,y)\mathbb{P}\left(\mathscr{A}(x,y)\cap\mathscr{I}\right)\ .$$

Let us define $\tau_2 = (1 - (1 - \mathbb{P}(\mathscr{A}(x,y)\,|\mathscr{I}\,))\tau_1)$. We claim that $\tau_2 > 0$. Indeed, if $x, y \in \mathrm{B}_Q(0, R_h + 2)$, since Υ and $Q^{-1}\nabla\Upsilon$ are continuous by assumption, we have by Lemma 3.18-(a)

$$\begin{aligned}&\mathbb{P}\left(\mathscr{A}(x,y)\,|\mathscr{I}\,\right)\\ &\qquad\geq \inf\left\{\exp(-G(t,z)^+)\ ;\ t \in \mathrm{B}_Q(0, R_h+2)\ ,\ z \in \mathrm{B}_Q(0, \Lambda(R_h))\right\} > 0\,,\end{aligned}$$

where $\Lambda(R_h) = C_\Upsilon((R_h+2)\nu + (h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + r_h)$. If $x, y \notin \mathrm{B}_Q(0, R_h)$, by **M 2**, $\mathbb{P}(\mathscr{A}(x,y)\,|\mathscr{I}\,) > a_l > 0$. Therefore, as $\mathbb{P}(\mathscr{R}(x,y)\cap\mathscr{I}) \leq \mathbb{P}(\mathscr{I}) - \mathbb{P}(\mathscr{I}\cap\mathscr{A}(x,y))$

$$\begin{aligned}&\mathbb{E}\left[d_{\eta,\epsilon}(X_1(x), Y_1(y))\mathbb{1}_{\mathscr{A}(x,y)\cap\mathscr{I}}\right] + \mathbb{E}\left[d_{\eta,\epsilon}(x,y)\mathbb{1}_{\mathscr{R}(x,y)\cap\mathscr{I}}\right]\\ &\qquad\leq \mathbb{P}(\mathscr{I})\,(1-\tau_2)d_{\eta,\epsilon}(x,y)\ .\end{aligned} \tag{3.49}$$

By Lemma 3.20-(ii)-(v), since $\delta_\epsilon(x,y) \leq \epsilon \leq 1$ and for all $a, b, c \in \mathbb{R}_+$, $(a+b)/(a+c) \leq 1 \vee (b/c)$,

$$\frac{d_{\eta,\epsilon}(X_h(x), Y_h(y))}{d_{\eta,\epsilon}(x,y)} \leq \nu\left\{2\eta(1-\nu) + \eta(h/2)\left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \widetilde{h}\eta\,\|Z_0\| + 1\right\}.$$

Also, by the Dominated Convergence theorem and Lemma 3.20-(v), for all $\kappa > 0$ there exists $\eta_2 \in (0, \eta_1]$ such that for all $\eta \in (0, \eta_2]$,

$$\begin{aligned}&\mathbb{E}\left[\{d_{\eta,\epsilon}(x,y) \vee d_{\eta,\epsilon}(X_1(x), Y_1(y))\}\,\mathbb{1}_{\mathscr{I}^c}\right]\\ &\leq d_{\eta,\epsilon}(x,y)\mathbb{E}\left[\left\{1 \vee \frac{d_{\eta,\epsilon}(X_1(x), Y_1(y))}{d_{\eta,\epsilon}(x,y)}\right\}\mathbb{1}_{\mathscr{I}^c}\right] \leq d_{\eta,\epsilon}(x,y)(1+\kappa)\mathbb{P}\left(\mathscr{I}^c\right)\ .\end{aligned} \tag{3.50}$$

Therefore since $\mathbb{P}(\mathscr{I}) > 0$, using (3.49)-(3.50) and choosing κ small enough, there exists $\tau_3, \eta_2 > 0$ such that for all $\eta \in (0, \eta_2]$

$$\mathbb{E}\left[d_{\eta,\epsilon}(X_1(x), Y_1(y))\mathbb{1}_{\mathscr{A}(x,y)\cap\mathscr{I}}\right] + \mathbb{E}\left[d_{\eta,\epsilon}(x,y)\mathbb{1}_{\mathscr{R}(x,y)\cap\mathscr{I}}\right] \\ + \mathbb{E}\left[\{d_{\eta,\epsilon}(x,y) \vee d_{\eta,\epsilon}(X_1(x), Y_1(y))\}\,\mathbb{1}_{\mathscr{I}^c}\right] \leq (1-\tau_3)d_{\eta,\epsilon}(x,y)\ . \tag{3.51}$$

Next, by Lemma 3.19 and Lemma 3.20-(iv), since $\delta_\epsilon(x,y) \leq 1$, there exists C, such that

$$\begin{aligned}
&\mathbb{E}\left[|\alpha_h(x, X_1(x)) - \alpha_h(y, Y_1(y))|\right] \\
&\leq C\,\|x-y\|_Q\left(\|x\|_Q \vee \|y\|_Q - \delta_\epsilon(x,y) + 1 + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \mathbb{E}\left[\|Z_0\|\right]\right) \\
&\leq C((1/\eta) + 1 + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \mathbb{E}\left[\|Z_0\|\right])\epsilon d_{\eta,\epsilon}(x,y)\ .
\end{aligned} \tag{3.52}$$

Set

$$\epsilon_1 \overset{\text{def}}{=} \tau_3/(2C((1/\eta_2) + 1 + \left\|Q^{-1}\nabla\Upsilon(0)\right\|_Q + \mathbb{E}\left[\|Z_0\|\right]))\ . \tag{3.53}$$

Therefore by (3.48)-(3.51) and (3.52), for $\eta \leftarrow \eta_2$ and $\epsilon \leftarrow \epsilon_1$, for all $x, y \in \mathbb{R}^d$,

$$\mathbf{K}d_{\eta,\epsilon}(x,y) \leq (1-\tau_3/2)d_{\eta,\epsilon}(x,y)\ . \tag{3.54}$$

□

Proof of Proposition 3.13. For ease of notation, we simply write $\mathbf{K}$ for $\mathbf{K}_{\mathrm{M}}$. Let $\{(\mathbf{X}_n, \mathbf{Y}_n),\ n \in \mathbb{N}\}$ be a Markov chain with Markov kernel $\mathbf{K}$. We denote for all $n \in \mathbb{N}^*$, Z_n and U_n, respectively the common gaussian variable and uniform variable, sampled to build $(\mathbf{X}_n, \mathbf{Y}_n)$. Let $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ be the probability and the expectation over $\{Z_n, U_n;\ n \in \mathbb{N}\}$. Note that by definition the variables $\{Z_n, U_n;\ n \in \mathbb{N}\}$ are independent. Under **M**1 and **M**2, by definition of Q and Lemma 3.21 the condition (i) and (ii) of Definition 3.1 are satisfied. In addition, there exists $\epsilon, \eta, \tau > 0$ such that for all $x, y \in \mathbb{R}^d$, $d_{\eta,\epsilon}(x,y) < 1$,

$$\mathbf{K}d_{\eta,\epsilon}(x,y) \leq (1-\tau)d_{\eta,\epsilon}(x,y) \tag{3.55}$$

Let $R > 0$, and x, y be in $\mathrm{B}_Q(0, R)$. Assume first $d_{\eta,\epsilon}(x,y) < 1$. Then by (3.55) and Lemma 4.23, for every $n \in \mathbb{N}^*$,

$$\mathbf{K}^n d_{\eta,\epsilon}(x,y) \leq \mathbf{K}^{n-1}d_{\eta,\epsilon}(x,y) \leq \cdots \leq (1-\tau)d_{\eta,\epsilon}(x,y)\ . \tag{3.56}$$

Consider now the case $d_{\eta,\epsilon}(x,y) = 1$. Let $\{(\mathbf{X}_n, \mathbf{Y}_n),\ n \in \mathbb{N}\}$ be the Markov chain with Markov kernel $\mathbf{K}$ starting in (x,y). Let $n \in \mathbb{N}^*$ and denote for all $1 \leq i \leq n$

$$\begin{aligned}
\Psi(\mathbf{X}_{i-1}, \mathbf{Y}_{i-1}, Z_i) &= \alpha_h(\mathbf{X}_{i-1}, \mathscr{O}(\mathbf{X}_{i-1}, Z_i)) \wedge \alpha_h(\mathbf{Y}_{i-1}, \mathscr{O}(\mathbf{Y}_{i-1}, Z_i)) \\
\mathscr{I}_i(n) &= \left\{\left\|(h/2)Q^{-1}\nabla\Upsilon(0) + \tilde{h}Q^{-1/2}Z_i\right\|_Q \leq R/n\right\} \\
\mathscr{A}_i(x,y) &= \{U_i \leq \Psi(\mathbf{X}_{i-1}, \mathbf{Y}_{i-1}, Z_i)\} \\
\widetilde{\mathscr{A}^i}(x,y,n) &= \bigcap_{1\leq j\leq i}\left(\mathscr{I}_j(n) \cap \mathscr{A}_j(x,y)\right)\ ,
\end{aligned}$$

where $\mathscr{O}$ given by (3.26). On the set $\widetilde{\mathscr{A}_i}(x,y,i)$, for all $1 \le j \le i$, $\mathbf{X}_j = \mathscr{O}(\mathbf{X}_{i-1}, Z_i)$, $\mathbf{Y}_j = \mathscr{O}(\mathbf{Y}_{i-1}, Z_i)$ and $\|\mathbf{X}_j\|_Q \vee \|\mathbf{Y}_j\|_Q \le 2R$. Then, since by Lemma 3.20-(iii),

$$d_{\eta,\epsilon}(\mathbf{X}_n, \mathbf{Y}_n) \le \epsilon^{-1} \|\mathbf{X}_n - \mathbf{Y}_n\|_Q \left(\eta \left\{\|\mathbf{X}_n\|_Q \vee \|\mathbf{Y}_n\|_Q\right\} + 1\right) ,$$

by Lemma 3.18-(a) on $\widetilde{\mathscr{A}^n}(x,y,n)$ it holds

$$d_{\eta,\epsilon}(\mathbf{X}_n, \mathbf{Y}_n) \le \epsilon^{-1}\nu^n \|x-y\|_Q (2\eta R + 1) .$$

This inequality and $d_{\eta,\epsilon} \le 1$ yield

$$\begin{aligned}
&\mathbf{K}^n d_{\eta,\epsilon}(x,y) = \widetilde{\mathbb{E}}_{x,y}\left[d_{\eta,\epsilon}(\mathbf{X}_n, \mathbf{Y}_n)(\mathbb{1}_{\widetilde{\mathscr{A}^n}(x,y,n)} + \mathbb{1}_{(\widetilde{\mathscr{A}^n}(x,y,n))^c})\right] \\
&\le \epsilon^{-1}\nu^n \|x-y\|_Q (2\eta R+1)\mathbb{P}\left(\widetilde{\mathscr{A}^n}(x,y,n)\right) + \mathbb{P}\left((\widetilde{\mathscr{A}^n}(x,y,n))^c\right) \\
&\le \epsilon^{-1}\nu^n 2R(2\eta R+1)\mathbb{P}\left(\widetilde{\mathscr{A}^n}(x,y,n)\right) + \mathbb{P}\left((\widetilde{\mathscr{A}^n}(x,y,n))^c\right) \\
&\le 1 + \left(\epsilon^{-1}\nu^n 2R(2\eta R+1) - 1\right) \mathbb{P}\left(\widetilde{\mathscr{A}^n}(x,y,n)\right) .
\end{aligned} \tag{3.57}$$

As $\nu \in (0,1)$, there exists m such that, $\epsilon^{-1}\nu^m 2R(2\eta R+1) < 1$. It remains to lower bound $\mathbb{P}\left(\widetilde{\mathscr{A}^m}(x,y,m)\right)$ by a positive constant to conclude, which is done by the following inequalities, where we use the independence of the random variables $\{Z_i, U_i;\ i \in \mathbb{N}^*\}$.

$$\begin{aligned}
&\mathbb{P}\left(\widetilde{\mathscr{A}^m}(x,y)\right) = \mathbb{P}\left(\widetilde{\mathscr{A}^{m-1}}(x,y,m) \cap \mathscr{I}_m(m)\right) \\
&\times \widetilde{\mathbb{E}}_{x,y}\left[\Psi(\mathbf{X}_{m-1}, \mathbf{Y}_{m-1}, Z_m) \middle| \widetilde{\mathscr{A}}^{m-1}(x,y,m) \cap \mathscr{I}_m(m)\right] .
\end{aligned}$$

For all $1 \le i \le m$, on the event $\bigcap_{j \le i} \mathscr{I}_j(m)$, it holds

$$\Psi(\mathbf{X}_{i-1}, \mathbf{Y}_{i-1}, Z_i) \ge \exp\left(-\sup_{(z,t)\in \mathrm{B}_Q(0,2R)} G(z,t)^+\right) = \delta ,$$

where $\delta \in (0,1)$, since G is continuous by **M** 1. Therefore, since Z_i is independent of $\widetilde{\mathscr{A}^{i-1}}(x,y,m)$, we have

$$\mathbb{P}\left(\widetilde{\mathscr{A}^m}(x,y,m)\right) \ge \delta\, \mathbb{P}\left(\widetilde{\mathscr{A}_{m-1}}(x,y)\right)\, \mathbb{P}\left(\mathscr{I}_m(m)\right) .$$

An immediate induction leads to

$$\mathbb{P}\left(\widetilde{\mathscr{A}^m}(x,y)\right) \ge \mathbb{P}\left(\mathscr{I}_1(m)\right)^m \delta^m .$$

Plugging this result in (3.57) and (3.56) imply there exists $s \in (0,1)$ such that for all $x,y \in \mathrm{B}_Q(0,R)$, $\mathbf{K}^m d_{\eta,\epsilon}(x,y) \le s d_{\eta,\epsilon}(x,y)$. □

Chapter 4

Subgeometric rates of convergence in Wasserstein distance for Markov chains

Alain Durmus[1], Gersende Fort[2], Éric Moulines [3]

Abstract

In this chapter, we provide sufficient conditions for the existence of the invariant distribution and for subgeometric rates of convergence in Wasserstein distance for general state-space Markov chains which are (possibly) not irreducible. Compared to [But14], our approach is based on a purely probabilistic coupling construction which allows to retrieve rates of convergence matching those previously reported for convergence in total variation in [DMS07].

Our results are applied to establish the subgeometric ergodicity in Wasserstein distance of non-linear autoregressive models and of the pre-conditioned Crank-Nicolson Markov chain Monte Carlo algorithm in Hilbert space.

4.1 Introduction

Convergence of general state-space Markov chains in total variation distance (or V-total variation) has been studied by many authors. There is a wealth of contributions establishing explicit rates of convergence under conditions implying geometric ergodicity; see [MT09, Chapter 16], [RR04], [Bax05], [BV13] and the references therein. Subgeometric

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. gersende.fort@telecom-paristech.fr

[3]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

(or Riemanian) convergence has been more scarcely studied; [TT94] characterized subgeometric convergence using a sequence of drift conditions, which proved to be difficult to use in practice. [JR02] have shown that, for polynomial convergence rates, this sequence of drift conditions can be replaced by a single drift condition, which shares some similarities with the classical Foster-Lyapunov approach for the geometric ergodicity. This result was later extended by [FM03] and [Dou+04] to general subgeometric rates of convergence. Explicit convergence rates were obtained in [Ver97; FM03; DGM08] and [AFV14].

The classical proofs of convergence in total variation distance are based either on a regenerative or a pairwise coupling construction, which requires the existence of accessible small sets and additional assumptions to control the moments of the successive return time to these sets. The existence of an accessible small set implies that the chain is irreducible.

In this paper, we establish rates of convergence for general state-space Markov chains which are (possibly) not irreducible. In such cases, Markov chains might not converge in total variation distance, but nevertheless may converge in a weaker sense; see for example [MS10]. We study in this paper the convergence in Wasserstein distance, which also implies the weak convergence. The use of the Wasserstein distance to obtain explicit rates of convergence has been considered by several authors, most often under conditions implying geometric ergodicity. A significant breakthough in this domain has been achieved in [HMS11] . The main motivation of [HMS11] was the convergence of the solutions of stochastic delay differential equations (SDDE) to their invariant measure. Nevertheless, the techniques introduced in [HMS11] laid the foundations of several contributions. [HSV14] used these techniques to prove the convergence of Markov chain Monte Carlo algorithms in infinite dimensional Hilbert spaces. An application for switched and piecewise deterministic Markov processes can be found in [CH14]. The results of [HMS11] were generalized by [But14] which establishes conditions implying the existence and uniqueness of the invariant distribution, and the subgeometric ergodicity of Markov chains (in discrete-time) and Markov processes (in continuous-time). [But14] used this result to establish subgeometric ergodicity of the solutions of SDDE. Nevertheless, when applied to the context of V-total variation, the rates obtained in [But14] in discrete-time do not exactly match the rates established in [Dou+04].

In this paper, we complement and sharpen the results presented in [But14] in the discrete-time setting. The approach developed in this paper is based on a coupling construction, which shares some similarities with the pairwise coupling used to prove geometric convergence in V-total variation. The arguments are therefore mostly probabilistic whereas [But14] heavily relies on functional analysis techniques and methods. We provide a sufficient condition couched in terms of a single drift condition for a coupling kernel outside an appropriately defined coupling set, extending the notion of d-small set of [HMS11]. We then show how this single drift condition implies a sequence of drift inequalities from which we deduce an upper bound of some subgeometric moment of the successive return times to the coupling set. The last step is to show that the Wasserstein distance between the distribution of the chain and the invariant probability measure is controlled by these moments. We apply our results to the convergence of some Markov chain Monte Carlo samplers with heavy tailed target distribution and to nonlinear autoregressive models whose the noise distribution can be singular with the Lebesgue measure. We also study the convergence of the preconditioned Crank-Nicolson algorithm when the target distribution has a density w.r.t. a Gaussian measure on an Hilbert space, under conditions which are weaker than [HSV14].

The paper is organized as follows: in Section 4.2, the main results on the convergence of Markov chains in Wasserstein distance are presented, under different sets of assumptions. Section 4.3 is devoted to the applications of these results. The proofs are given in Section 4.4 and Section 4.5.

Notations

Let (E, d) be a Polish space where d is a distance bounded by 1. We denote by $\mathcal{B}(E)$ the associated Borel σ-algebra and $\mathcal{P}(E)$ the set of probability measures on $(E, \mathcal{B}(E))$. Let $\mu, \nu \in \mathcal{P}(E)$; λ is a transference plan of μ and ν if λ is a probability on the product space $(E \times E, \mathcal{B}(E \times E))$, such that $\lambda(A \times E) = \mu(A)$ and $\lambda(E \times A) = \nu(A)$ for all $A \in \mathcal{B}(E)$. The set of transference plans of $\mu, \nu \in \mathcal{P}(E)$ is denoted $\Pi(\mu, \nu)$. We say that a couple of E-random variables (X, Y) is a coupling of μ and ν if there exists $\lambda \in \Pi(\mu, \nu)$ such that (X, Y) are distributed according to λ. Let P be Markov kernel of $E \times \mathcal{B}(E)$; a Markov kernel Q on $(E \times E, \mathcal{B}(E \times E))$ such that, for every $x, y \in E$, $Q(x, y, \cdot)$ is a transference plan of $P(x, \cdot)$ and $P(y, \cdot)$ is a *coupling kernel* for P.

The Wasserstein metric associated with d, between two probability measures $\mu, \nu \in \mathcal{P}(E)$ is defined by:

$$W_d(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \int_{E \times E} d(x, y) \mathrm{d}\gamma(x, y) \ . \tag{4.1}$$

When d is the trivial metric $d_0(x, y) = \mathbb{1}_{x \neq y}$, the associated Wasserstein metric is the total variation distance $W_{d_0}(\mu, \nu) = \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|$. Since d is bounded, the Monge-Kantorovich duality Theorem implies (see [Vil09, Remark 6.5]) that the lower bound in (4.1) is realized. In addition, W_d is a metric on $\mathcal{P}(E)$ and $\mathcal{P}(E)$ equipped with W_d is a Polish space; see [Vil09, Theorems 6.8 and 6.16]. Finally, the convergence in W_d implies the weak convergence (see e.g. [Vil09, Corollary 6.11]).

Let Λ_0 be the set of measurable functions $r_0 : \mathbb{R}_+ \to [2, +\infty)$, such that r_0 is non-decreasing, $x \mapsto \log(r_0(x))/x$ is non-increasing and $\lim_{x \to \infty} \log(r_0(x))/x = 0$. Denote by Λ the set of positive functions $r : \mathbb{R}_+ \to (0, +\infty)$, such that there exists $r_0 \in \Lambda_0$ satisfying:

$$0 < \liminf_{x \to +\infty} r(x)/r_0(x) \leq \limsup_{x \to +\infty} r(x)/r_0(x) < +\infty \ . \tag{4.2}$$

Finally, let $\mathbb{F}$ be the set of concave increasing functions $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, continuously differentiable on $[1, +\infty)$, and satisfying $\lim_{x \to +\infty} \phi(x) = +\infty$ and $\lim_{x \to +\infty} \phi'(x) = 0$. For $\phi \in \mathbb{F}$, we denote by $\phi^{\leftarrow}$ the inverse of ϕ.

4.2 Main results

The key ingredient for the derivation of a Markov kernel P on (E, d) is the existence of a coupling kernel $Q(x, y, \cdot)$ for P satisfying a strong contraction property when (x, y) belongs to a set Δ, referred to as a *coupling set*. For $\Delta \in \mathcal{B}(E \times E)$, a positive integer ℓ and $\epsilon > 0$, consider the following assumption:

H 2 (Δ, ℓ, ϵ)**.** *(i) Q is a d-weak-contraction: for every $x, y \in E$, $Qd(x, y) \leq d(x, y)$.*

(ii) $Q^\ell d(x, y) \leq (1 - \epsilon)d(x, y)$, for every $x, y \in \Delta$.

A set Δ satisfying **H** 2(Δ, ℓ, ϵ)-(ii) will be referred to as a (ℓ, ϵ, d)-coupling set. Of course the definition of this set also depends on the choice of the coupling kernel Q, but this dependence is implicit in the notation. If $d = d_0$ and Δ is a $(1, \epsilon)$-pseudo small set (with $\epsilon > 0$) in the sense that

$$\inf_{(x,y)\in\Delta} [P(x, \cdot) \wedge P(y, \cdot)](E) \geq \epsilon \ ,$$

then **H**2$(\Delta, 1, \epsilon)$ is satisfied by the pairwise coupling kernel (see [RR01b]).

The following theorem shows that, under **H**2(Δ, ℓ, ϵ) and a condition which essentially claims that the first moment of the hitting time to the coupling set Δ is finite, the Markov kernel P admits a unique invariant distribution.

Theorem 4.1. *Assume that there exist*

*(i) a coupling kernel Q for P, a set $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$ and $\epsilon > 0$ such that **H** 2(Δ, ℓ, ϵ) holds,*

(ii) a measurable function $\mathcal{V} : E^2 \to [1, \infty)$ and a constant $b < \infty$ such that the following drift condition is satisfied.

$$Q\mathcal{V}(x, y) \leq \mathcal{V}(x, y) - 1 + b\mathbb{1}_\Delta(x, y) \ , \qquad \sup_{(x,y)\in\Delta} Q^{\ell-1}\mathcal{V}(x, y) < +\infty \ . \tag{4.3}$$

(iii) an increasing sequence of integers $\{n_k, k \in \mathbb{N}\}$ and a concave function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\lim_{v\to+\infty} \psi(v) = +\infty$ and

$$\sup_{k\in\mathbb{N}} P^{n_k}[\psi \circ \mathcal{V}_{x_0}](x_0) < +\infty \ , \qquad P\mathcal{V}_{x_0}(x_0) < +\infty \quad \textit{for some } x_0 \in E, \tag{4.4}$$

where $\mathcal{V}_{x_0} = \mathcal{V}(x_0, \cdot)$.

Then, P admits a unique invariant distribution.

Proof. See Section 4.4.1. □

If we now combine **H**2(Δ, ℓ, ϵ) with a condition which implies the control of the tail probabilities of the successive return times to the coupling sets (more precisely, of the moments of order larger than one of these return times) then the Wasserstein distance between $P^n(x, \cdot)$ and $P^n(y, \cdot)$ may be shown to decrease at a subgeometric rate. To control these moments, it is quite usual to consider drift conditions. In this paper, we focus on a class of drift conditions which has been first introduced in [Dou+04]. For $\Delta \in \mathcal{B}(E \times E)$, a function $\phi \in \mathbb{F}$, a measurable function $V : E \to [1, +\infty)$, consider the following assumption:

H 3 (Δ, ϕ, V)**.** *(i) There exists a constant $b < \infty$ such that for all $x, y \in E$:*

$$PV(x) + PV(y) \leq V(x) + V(y) - \phi(V(x) + V(y)) + b\mathbb{1}_\Delta(x, y) \; . \tag{4.5}$$

(ii) $\sup_{(x,y)\in\Delta}\{V(x) + V(y)\} < +\infty$.

Not surprisingly, this condition implies that the return time to the coupling set Δ possesses a first moment. This property combined with Theorem 4.1 yields

Corollary 4.2. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, $\epsilon > 0$, $\phi \in \mathbb{F}$ and $V : E \to [1, \infty)$ such that* ***H*** *2(Δ, ℓ, ϵ)-**H** 3(Δ, ϕ, V) are satisfied. Then, P admits a unique invariant probability measure π and $\pi(\phi \circ V) < \infty$.*

Proof. See Section 4.4.2. □

We now derive expressions of the rate of convergence and make explicit the dependence upon the initial condition of the chain. For $\phi \in \mathbb{F}$, set

$$H_\phi(t) = \int_1^t \frac{1}{\phi(s)} \mathrm{d}s \; . \tag{4.6}$$

Since for $t \geq 1$, $\phi(t) \leq \phi(1) + \phi'(1)(t-1)$, the function H_ϕ is monotone increasing to infinity, twice continuously differentiable and concave. Its inverse, denoted $H_\phi^{\leftarrow}$, is well defined on $\mathbb{R}_+$, is twice continuously differentiable and convex (see e.g. [Dou+04, Section 2.1]).

Theorem 4.3. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, $\epsilon > 0$, $\phi \in \mathbb{F}$ and $V : E \to [1, \infty)$ such that* ***H*** *2(Δ, ℓ, ϵ)-**H** 3(Δ, ϕ, V) are satisfied. Let π be the invariant probability of P.*

(i) There exist constants $\{C_i\}_{i=1}^3$ such that for all $x \in E$ and all $n \geq 1$

$$\begin{aligned} W_d(P^n(x, \cdot), \pi) \leq C_1 V(x)/H_\phi^{\leftarrow}(n/2) + C_2/\phi(H_\phi^{\leftarrow}(n/2)) \\ + C_3/H_\phi^{\leftarrow}(-\log(1-\epsilon)\, n/\{2(\log(H_\phi^{\leftarrow}(n)) - \log(1-\epsilon))\}) \; . \end{aligned}$$

(ii) For all $\delta \in (0, 1)$, there exists a constant C_δ such that for all $x \in E$ and all $n \geq 1$

$$W_d(P^n(x, \cdot), \pi) \leq C_\delta \, V(x)/\phi(\{H_\phi^{\leftarrow}(n)\}^\delta) \; .$$

The values of the constants C_i, for $i = 1, 2, 3$, and C_δ are given explicitly in the proof.

Proof. See Section 4.4.3 □

We summarize in Table 4.1 the rates of convergence obtained (for a given $x \in E$) from Theorem 4.3 for usual concave functions ϕ: logarithmic rates $\phi(t) = (1 + \log t))^\kappa$ for some $\kappa > 0$; polynomial rates $\phi(t) = t^\kappa$ for some $\kappa \in (0, 1)$; subexponential rates $\phi(t) = t/(1 + \log t)^\kappa$ for some $\kappa > 0$. Note that since $\phi \in \mathbb{F}$, the first term in the RHS of the bound in (i) is not the leading term (for fixed x, when $n \to \infty$). In the case ϕ is logarithmic or polynomial, the leading term in the RHS is the second one so that the rate of decay is given by $1/\phi(H_\phi^{\leftarrow}(n/2))$. For the logarithmic and polynomial cases, the best rates are given by Theorem 4.3-(i) and for the subexponential case, by Theorem 4.3-(ii).

Order of the rates of convergence in	$\phi(x) = (1 + \log(x))^\kappa$ for $\kappa > 0$	$\phi(x) = x^\kappa$ for $\kappa \in (0, 1)$ set $\varsigma = \kappa/(1 - \kappa)$	$\phi(x) = x/(1 + \log(x))^\kappa$ for $\kappa > 0$ set $\varsigma = 1/(1 + \kappa)$
Theorem 4.3	$1/\log^\kappa(n)$	$1/n^\varsigma$	$\exp(-\delta((1 + \kappa)n)^\varsigma)$ for all $\delta \in (0, 1)$
[Dou+04]	$1/\log^\kappa(n)$	$1/n^\varsigma$	$n^{\kappa\varsigma}\exp(-((1 + \kappa)n)^\varsigma)$
[But14] for all $\delta \in (0, 1)$	$1/\log^{\delta\kappa}(n)$	$1/n^{\delta\varsigma}$	$\exists C > 0$ $\exp(-Cn^\varsigma)$

Table 4.1: Rates of convergence when ϕ increases at a logarithmic rate, a polynomial rate and a subexponential rate, obtained from Theorem 4.3 and from [But14, Theorem 2.1] and [Dou+04, Section 2.3.].

In practice, it is often easier to establish a drift inequality on E rather than on $E \times E$ as in $\mathbf{H}3(\Delta, \phi, V)$. Theorem 4.4 relates the following single drift condition to the drift $\mathbf{H}3$. For a function $\phi \in \mathbb{F}$, a measurable function $V : E \to [1, +\infty)$ and a constant $b \geq 0$, consider the following assumption

H 4 (ϕ, V, b)**.** $\phi(0) = 0$ *and for all* $x \in E$*,*

$$PV(x) \leq V(x) - \phi \circ V(x) + b\,. \tag{4.7}$$

Theorem 4.4. *Let $\phi \in \mathbb{F}$, a measurable function $V : E \to [1, +\infty)$ and a constant $b \geq 0$ such that $\mathbf{H}4(\phi, V, b)$ holds. Then $\mathbf{H}3(\{V \leq \upsilon\}^2, c\phi, V)$ is satisfied for any $\upsilon > \phi^{\leftarrow}(2b)$ and with $c = 1 - 2b/\phi(\upsilon)$.*

The proof is postponed in Section 4.4.4. Note that we can assume without loss of generality that $t \mapsto \phi(t)$ is concave increasing and continuously differentiable only for large t; see Lemma 4.21.

Our assumptions and results can be compared to [But14] which also establish convergence in Wasserstein distance at a subgeometric rate under the single drift condition $\mathbf{H}4(\phi, V, b)$ and the following assumptions

B-(i) For all $x, y \in E$, $W_d(P(x,\cdot), P(y,\cdot)) \leq d(x,y)$.

B-(ii) There exists $\eta > 0$ such that the level set $\Delta = \{(x,y) : V(x) + V(y) \leq \phi^{\leftarrow}(2b) + \eta\}$ is d-small for P: there exists $\epsilon > 0$ such that for any $x, y \in \Delta$, $W_d(P(x,\cdot), P(y,\cdot)) \leq (1-\epsilon)d(x,y)$.

Under these conditions, [But14, Theorem 2.1] shows the existence and uniqueness of the stationary distribution π and provides rates of convergence to stationarity in the Wasserstein distance W_d; expressions for these rates are provided in the last row of Table 4.1 for various choices of functions ϕ. It can be seen that our results always improve the rates of convergence when compared to those of [But14]

Let us compare the assumptions of Theorem 4.3 to (B). It follows from [Vil09, Corollary 5.22] that under B-(i) and B-(ii), there exists a coupling kernel for P (which is the coupling kernel realizing the lower bound in the Monge-Kantorovitch duality theorem) such that $\mathbf{H}2(\Delta, 1, \epsilon)$ holds. Since Theorem 4.4 establishes that a single drift condition of the form $\mathbf{H}4$ implies a drift condition of the form $\mathbf{H}3$, the assumptions of [But14, Theorem 2.1] essentially differ from the assumptions of Theorem 4.3 through the coupling set assumption: [But14, Theorem 2.1] only covers coupling sets of order 1 when our result covers coupling sets of order ℓ, for any $\ell \geq 1$. This is an unnatural and sometimes annoying restriction since in practical examples the order ℓ is most likely to be large (see e.g. the examples in Section 4.3). Note that the strategy consisting in applying a result for a coupling set of order 1 to the ℓ-iterated kernel is not equivalent to applying a result for a coupling set of order ℓ to the one iterated kernel; we provide an illustration of this claim in Section 4.3.1. Checking $\mathbf{H}2(\Delta, \ell, \epsilon)$ is easier than checking (B) since allowing the coupling set to be of any order provides far more flexibility.

Our results can also be compared to the explicit rates in [Dou+04] derived for convergence in total variation distance. In [Dou+04],q it is assumed that P is phi-irreducible, aperiodic, that the drift condition $\mathbf{H}4$ holds and that the level sets $\{V \leq v\}$ are small in the usual sense, i.e. for some $\ell \in \mathbb{N}^*$, $\epsilon \in (0,1)$ and a probability ν that may depend upon the level set, $P^\ell(x, A) \geq \epsilon\nu(A)$, for all $x \in \{V \leq v\}$ and $A \in \mathcal{B}(E)$. Under these assumptions, [Dou+04, Proposition 2.5] shows that for any $x \in E$, $\lim_{n\to\infty} \phi(H_\phi^{\leftarrow}(n))\, W_{d_0}(P^n(x,\cdot), \pi) = 0$, where W_{d_0} is the total variation distance. Table 4.1 displays the rate r_ϕ obtained in [Dou+04] (see penultimate row) and the rates given by Theorem 4.3 (row 2): our results coincide with [Dou+04] for the polynomial and logarithmic cases and the logarithm of the rate differs by a constant (which can be chosen arbitrarily close to one in our case) in the subexponential case. Nevertheless, we would like to stress that our conditions do not require ϕ-irreducibility and therefore apply in more general contexts.

4.3 Application

4.3.1 A symmetric random walk Metropolis algorithm

Let $E \stackrel{\text{def}}{=} \{k/4; k \in \mathbb{Z}\}$ endowed with the trivial distance d_0, thus (E, d_0) is a Polish space. Consider a symmetric random walk Metropolis (SRWM) algorithm on E for an heavy tailed target distribution π given by

$$\pi(x) \propto 1/(1+|x|)^{1+h}\ , \quad \text{for all } x \in E\ , \tag{4.8}$$

where $h \in (0, 1/2)$. Starting at $x \in E$, the Metropolis algorithm proposes at each iteration, a candidate y from a random walk with a symmetric increment distribution q on E. The move is accepted with probability $\alpha(x, y) = 1 \wedge (\pi(y)/\pi(x))$. The Markov kernel associated with the SRWM algorithm is given, for all $x \in E$ and $A \subset E$, by

$$P(x, A) = \sum_{y, x+y \in A} \alpha(x, x+y)q(y) + \delta_x(A) \sum_{y \in E} (1 - \alpha(x, x+y))q(y)\ .$$

Assume that q is the uniform distribution on $\{-1/4, 0, 1/4\}$. It is easily checked that P is irreducible and aperiodic. In the following, we prove that [But14, Theorem 2.1] cannot be applied to this case, contrary to Theorem 4.3.

We first prove that P cannot be geometrically ergodic. The proof essentially follows from [JT03, Theorem 2.2], where the authors established necessary and sufficient conditions for the geometric and the polynomial ergodicity of random walk type Markov chains on $\mathbb{R}$.

Proposition 4.5. *P is not geometrically ergodic.*

Proof. The proof is by contradiction: we assume that P is geometrically ergodic. Since it is also P irreducible and aperiodic, the stationary distribution π is unique and geometrically regular: for any set A such that $\pi(A) > 0$, there exists $L > 1$ such that $\mathbb{E}_\pi[L^{\tau_A}] = \sum_{x \in E} \pi(x)\mathbb{E}_x[L^{\tau_A}] < \infty$, where τ_A is the return time to A. Choose $M > 0$, $A = \{x \in E, |x| \leq M\}$. Since for $|x| \geq M$, $\tau_A \geq 4(|x| - M)$ $\mathbb{P}_x$-a.s. the regularity of π claims that there exists $L > 1$ such that $\sum_{x \in \mathbb{Z}} L^{|x|}\pi(x) < \infty$. This clearly yields to a contradiction. $\square$

We then show that the Markov kernel P satisfies a sub-geometric drift condition. For $s \geq 0$, set $V_s(x) = 1 \vee |x|^s$.

Proposition 4.6. *For all $s \in (2, 2+h)$, there exist $b, c > 0$ such that for all $x \in E$*

$$PV_s(x) \leq V_s(x) - cV_s(x)^{(s-2)/s} + b\ . \tag{4.9}$$

Proof. We have for all $x \geq 5/4$,

$$\begin{aligned} &PV_s(x) - V_s(x) \\ &= (x^s/3)\left(((1-(4x)^{-1})^s - 1) - (1 - 1/(5+4x))^{1+h}(1 - (1+(4x)^{-1})^s)\right) \\ &\underset{x \to +\infty}{=} x^{s-2}s(s-h-2)/48 + o(x^{s-2})\ . \end{aligned}$$

The same expansion remains valid as $x \to -\infty$ upon replacing x by $-x$. $\square$

Using this result, [Dou+04, Proposition 2.5] shows that for any $x \in E$, $P^n(x,\cdot)$ converges to π in total variation norm, at the rates $n^{\tilde{h}}$ for all $\tilde{h} \in (0, h/2)$.

We can also apply Theorem 4.4 and Theorem 4.3-(i). For any $s \in (2, 2+h)$, **H** 4(ϕ_s, V_s, b) is satisfied with $\phi_s(x) = cx^{(s-2)/s}$, $V_s(x) = 1 \vee |x|^s$ and $b < +\infty$. For $x, y \in E$ and $A, B \subset E$, consider the following kernel:

$$Q((x,y),(A \times B)) = P(x,A)P(y,B)\mathbb{1}_{\{x \neq y\}} + P(x, A \cap B)\mathbb{1}_{\{x=y\}} .$$

Clearly, Q is a coupling kernel for P. Let us prove that for any $M > 0$ and any $\ell \geq 4M$, there exists $\epsilon > 0$ such that **H** 2(Δ, ℓ, ϵ) holds with $\Delta = \{|x| \vee |y| \leq M\}$. We have $Qd_0(x,y) \leq d_0(x,y)$ for every $x \neq y \in E$ and by definition of Q, $Qd_0(x,x) = 0$ for every $x \in E$. Let $M > 0$, $\ell \geq 4M$. For any $x, y \in \{|x| \vee |y| \leq M\}$ such that $|x| < |y|$

$$\begin{aligned}\widetilde{\mathbb{P}}_{x,y}\left[X_\ell = Y_\ell\right] &\geq \widetilde{\mathbb{P}}_{x,y}\left[X_{4|y|} = Y_{4|y|}\right] \\ &\geq \widetilde{\mathbb{P}}_{x,y}\left[\tau_0^X = 4|x|, X_{4|x|+1} = 0, \ldots, X_{4|y|} = 0, \tau_0^Y = 4|y|\right] ,\end{aligned}$$

where $\tau_0^X = \inf\{n \geq 1, X_n = 0\}$ and $\tau_0^Y = \inf\{n \geq 1, Y_n = 0\}$. Since

$$\begin{aligned}&\widetilde{\mathbb{P}}_{x,y}\left[\tau_0^X = 4|x|\right] \geq (1/3)^{4|x|} , \qquad \widetilde{\mathbb{P}}_{x,y}\left[\tau_0^Y = 4|y|\right] \geq (1/3)^{4|y|} , \\ &\widetilde{\mathbb{P}}_{x,y}\left[X_{4|x|+1} = 0, \ldots, X_{4|y|} = 0\right] \geq (1/3)^{4(|y|-|x|)} ,\end{aligned}$$

it follows that $Q^\ell d_0(x,y) = 1 - \widetilde{\mathbb{P}}_{x,y}\left[X_\ell = Y_\ell\right] \leq 1 - (1/3)^{8|y|} \leq 1 - (1/3)^{8M} d_0(x,y)$. This inequality remains valid when $x = y$. This concludes the proof of **H** 2(Δ, ℓ, ϵ). By Theorem 4.4, the kernel P is subgeometrically ergodic in total variation distance at the rates $n^{\tilde{h}}$, for $\tilde{h} \in (0, h/2)$.

In this example, [But14, Theorem 2.1] cannot be applied. Indeed, on one hand, for any $M > 0$ the set $\Delta_M = \{|x| \vee |y| \leq M\}$ is a $(1, \epsilon, d_0)$-coupling set for P^ℓ iff $l \geq 4M$. This property is a consequence of the above discussion (for the converse implication) and of the equality $W_{d_0}(P^\ell(x,\cdot), P^\ell(y,\cdot)) = 1$ if $|x - y| > \ell/2$ (for the direct implication). On the other hand, in order to check B-(ii) for some ℓ-iterated kernel P^ℓ, we have to prove that there exits $\eta > 0$ such that $\Delta_\star = \{(x,y) \in E^2; V_s(x) + V_s(y) \leq (2b\ell/c)^{s/(s-2)} + \eta\}$ is a $(1, \epsilon, d_0)$-coupling set for P^ℓ - the constants b, c are given by Proposition 4.6. Unfortunately, since $b/c \geq 1$ (apply the drift inequality 4.9 with $x = 0$), and $1/(s-2) \geq 2$, we get

$$\{(x,y) \in E, |x| \vee |y| \leq 4\ell^2\} \subset \{x, y \in E; |x| \vee |y| \leq (2b\ell/c)^{1/(s-2)}\} \subset \Delta_\star ,$$

Therefore whatever ℓ, $\Delta_\star$ is not a $(1, \epsilon, d_0)$ coupling set for P^ℓ.

4.3.2 Non linear autoregressive model

In this section, we consider the functional autoregressive process $(X_n)_{n \in \mathbb{N}}$ on $E = \mathbb{R}^p$, given by $X_{n+1} = g(X_n) + Z_{n+1}$ Denote by $\|\cdot\|$ the Euclidean norm on E and $\mathrm{B}(x, M)$ the ball of radius $M \geq 0$ and centered at $x \in \mathbb{R}^p$, associated with this norm. Consider the following assumptions:

AR 1. *$(Z_n)_{n\in\mathbb{N}^*}$ is an independent and identically distributed (i.i.d.) zero-mean $\mathbb{R}^p$-valued sequence, independent of X_0, and satisfying $\int \exp\left(\beta_0 \|z\|^{\kappa_0}\right)\mu(\mathrm{d}z) < +\infty$, where μ is the distribution of Z_1 for some $\beta_0 > 0$ and $\kappa_0 \in (0,1]$.*

AR2. *For all $M > 0$, $g : \mathbb{R}^p \to \mathbb{R}^p$ is C_M-Lipschitz on $\mathrm{B}(0,M)$ with respect to $\|\cdot\|$ where $C_M \in (0,1)$. Furthermore, there exist positive constants r, M_0, and $\rho \in [0,2)$, such that $\|g(x)\| \leq \|x\|\left(1 - r\|x\|^{-\rho}\right) \quad$ if $\|x\| \geq M_0$.*

A simple example of function g satisfying **AR2** is $x \mapsto x \cdot \max\left(1/2, 1 - 1/\|x\|^{\rho}\right)$ with $\rho \in [0,2)$. Denote by P the Markov kernel defined by the process $(X_n)_n$. Proposition 4.7 establishes **H4**(ϕ, V, b) in the case where $\rho > \kappa_0$, and a geometric drift condition in the other case.

Proposition 4.7. *[Dou+04, Theorem 3.3] Assume **AR1** and **AR2**.*

*(i) If $\rho > \kappa_0$, there exist $\beta \in (0,\beta_0)$ and $b, c > 0$ such that **H4**(ϕ, V, b) holds with $\phi(x) := cx(1 + \log(x))^{1-\rho/(\kappa_0\wedge(2-\rho))}$ and $V(x) := \exp(\beta\|x\|^{\kappa_0\wedge(2-\rho)})$.*

(ii) It $\rho \leq \kappa_0$, then there exist $b < +\infty$ and $\zeta \in (0,1)$ such that for all $x \in \mathbb{R}^p$, $PV(x) \leq \zeta V(x) + b$ where $V(x) = \exp(\beta\|x\|^{\kappa_0})$ with $\beta \in (0,\beta_0)$.

Proof. The proof of Proposition 4.7 is along the same lines as [Dou+04, Theorem 3.3] and is omitted [1] □

Consider the coupling kernel Q defined for all $x, y \in E$ and $A \in \mathcal{B}(E \times E)$ by

$$Q((x,y), A) = \int \mathbb{1}_A(g(x) + z, g(y) + z)\mu(\mathrm{d}z)\,. \tag{4.10}$$

For $\eta > 0$, define $d_\eta(x,y) \stackrel{\text{def}}{=} 1 \wedge \eta^{-1}\|x - y\|$.

Proposition 4.8. *Assume **AR1**-**AR2**. For any $M > 0$, there exists $\epsilon > 0$ such that **H2**$(\Delta_M, 1, \epsilon)$ is satisfied with $\Delta_M = \mathrm{B}(0,M) \times \mathrm{B}(0,M)$.*

Proof. Since $d_\eta(x,y) = \|x - y\|/\eta$ for any $x, y \in \mathrm{B}(0,M)$ and $\eta = 2M$, we get under **AR2**,

$$\mathbb{E}[d_\eta(g(x) + \epsilon_1, g(y) + \epsilon_1)] \leq \eta^{-1}\|g(x) - g(y)\| \wedge 1 \leq C_M\eta^{-1}\|x - y\| \leq C_M d_\eta(x,y)\,. \tag{4.11}$$

Finally, since **AR2** implies that g is 1-Lipschitz on $\mathbb{R}^p$, (4.11) shows that $\mathbb{E}[d_\eta(g(x) + \epsilon_1, g(y) + \epsilon_1)] \leq d_\eta(x,y)$ for all $x, y \in \mathbb{R}^p$. □

For all $\eta, \eta' > 0$, d_η and $d_{\eta'}$ are Lipschitz equivalent, i.e., there exists $C > 0$ such that for all $x, y \in \mathbb{R}^p$, $C^{-1}d_\eta(x,y) \leq d_{\eta'}(x,y) \leq Cd_\eta(x,y)$, which implies (see (4.1)) that W_{d_η} and $W_{d_{\eta'}}$ are Lipschitz equivalent.

[1] We point out that in [Dou+04], it is additionally required that the distribution of Z_1 has a nontrivial absolutely continuous component which is bounded away from zero in a neighborhood of the origin. However, this condition is only required to establish the ϕ-irreducibility of the Markov chain, which is not needed here.

Theorem 4.9. *Assume* **AR1** *and* **AR2** *hold. Then P admits a unique invariant distribution π.*

(i) If $\rho > \kappa_0$, there exist two constants C_1 and C_2 such that for all $x \in \mathbb{R}^p$ and $n \in \mathbb{N}^*$

$$W_{d_1}(P^n(x, \cdot), \pi) \leq C_1 V(x) \exp\left(-C_2 n^{\varsigma}\right) ,$$

where $\varsigma = (\kappa_0 \wedge (2 - \rho))/\rho$.

(ii) If $\rho \leq \kappa_0$, then there exist $\tilde{\zeta} \in (0, 1)$ and a constant C such that for all $x \in \mathbb{R}^p$ and $n \in \mathbb{N}^*$

$$W_{d_1}(P^n(x, \cdot), \pi) \leq C_1 V(x) \tilde{\zeta}^n .$$

Proof. By application of Corollary 4.2, Theorem 4.3 and Theorem 4.4, we deduce (i) from Proposition 4.7-(i) and Proposition 4.8. By an application of [HMS11, Theorem 4.8, Corollary 4.11], we deduce (ii) from Proposition 4.7-(ii) and Proposition 4.8. □

Perhaps surprisingly, we cannot relax the condition $\kappa_0 \in (0, 1]$, to obtain geometric convergence for $1 < \rho \leq \kappa_0$. Indeed, [RT96a, Theorem 3.2(a)] provides an example where **AR1** and **AR2** are satisfied for $\kappa_0 = 2$ and $\rho \in (1, 2)$, but the chain fails to be geometrically ergodic (for the total variation distance).

4.3.3 The preconditioned Crank-Nicolson algorithm

In this section, we consider the preconditioned Crank-Nicolson algorithm introduced in [Bes+08] and analyzed in [HSV14] for sampling in a separable Hilbert space $(\mathsf{H}, \|\cdot\|)$ a distribution with density $\pi \propto \exp(-g)$ with respect to a zero-mean Gaussian measure γ with covariance operator C; see [Bog98]. This algorithm is studied in [HSV14] under conditions which imply the geometric convergence in Wasserstein distance. We consider

Algorithm 2: Preconditioned Crank-Nicolson Algorithm

```
Data: ρ ∈ [0, 1)
Result: (X_n)_{n∈ℕ}
begin
    Initialize X_0
    for n ≥ 0 do
        Generate Z_{n+1} ~ γ.
        Generate U_{n+1} ~ 𝒰([0, 1])
        if U_{n+1} ≤ α(X_n, ρX_n + √(1 − ρ²) Z_{n+1}) =
        1 ∧ exp(g(X_n) − g(ρX_n + √(1 − ρ²) Z_{n+1})) then
            X_{n+1} = ρX_n + √(1 − ρ²) Z_{n+1}
        else
            X_{n+1} = X_n
```

the convergence of the Crank-Nicolson algorithm under the weaker condition **CN**1 below for which the results in [HSV14] cannot be applied. We will show that subgeometric convergence can nevertheless be obtained.

CN1. *The function $g : \mathsf{H} \to \mathbb{R}$ is β-Hölder for some $\beta \in (0, 1]$ i.e., there exists C_g, such that for all $x, y \in \mathsf{H}$, $|g(x) - g(y)| \leq C_g \|x - y\|^\beta$.*

Examples of densities satisfying **CN** 1 are $g(x) = -\|x\|^\beta$ with $\beta \in (0, 1]$. The following Theorem implies that under **CN**1, $\exp(-g)$ is γ-integrable (see [Bog98, Theorem 2.8.5]).

Theorem 4.10 (Fernique's theorem)**.** *There exist $\theta \in \mathbb{R}_+^*$ and a constant C_θ such that $\int_{\mathsf{H}} \exp(\theta \|\xi\|^2) \mathrm{d}\gamma(\xi) \leq C_\theta$.*

The Crank-Nicolson kernel P_{cn} has been shown to be geometrically ergodic by [HSV14] under the assumptions that g is globally Lipschitz and that there exist positive constants C, M_1, M_2 such that for $x \in \mathsf{H}$ with $\|x\| \geq M_1$, $\inf_{z \in \overline{B}(\rho x, M_2)} \exp(g(x) - g(z)) \geq C$ (see [HSV14, Assumption 2.10-2.11]), where we denote by $\mathrm{B}(x, M)$ the open ball centered at $x \in \mathsf{H}$ and of radius $M > 0$ associated with $\|\cdot\|$, and by $\overline{B}(x, M)$ its closure. Such an assumption implies that the acceptance ratio $\alpha(x, \rho x + \sqrt{1-\rho^2}\xi)$ is bounded from below as $\|x\| \to \infty$ uniformly on $\xi \in \overline{B}(0, M_2/\sqrt{1-\rho^2})$. In **CN** 1, this condition is weakened in order to address situations in which the acceptance-rejection ratio vanishes when $\|x\| \to \infty$: this happens when $\lim_{\|x\| \to +\infty} \{g(\rho x) - g(x)\} = +\infty$. We first check that **H**4(ϕ, V, b) is satisfied with

$$V(x) = \exp(s \|x\|^2) \;, \tag{4.12}$$

where $s = (1-\rho)^2\theta/16$ and θ is given by Theorem 4.10.

Proposition 4.11. *Assume **CN** 1, and let $\rho \in [0, 1)$. Then there exist $b \in \mathbb{R}_+$ and $c \in (0, 1)$ such that for all $x \in \mathsf{H}$*

$$P_{\mathrm{cn}} V(x) \leq V(x) - \phi \circ V(x) + b \;,$$

where $\phi \in \mathbb{F}$ and $\phi(t) \sim_{t \to \infty} ct \exp(-\{\log(t)/\kappa\}^{\beta/2})$, with $\kappa = \theta C_g^{-2/\beta}/36$.

Proof. The proof is postponed to Section 4.5.1. □

We now deal with showing **H**2. To that goal, we introduce the distance $d_\eta(x, y) = 1 \wedge \eta^{-1} \|x - y\|^\beta$, for any $\eta > 0$, and for $x, y \in E$ the basic coupling Q_{cn} between $P_{\mathrm{cn}}(x, \cdot)$ and $P_{\mathrm{cn}}(y, \cdot)$: the same Gaussian variable Ξ and the same uniform variable U are generated to build X_1 and Y_1, with initial conditions x, y. Define $\Lambda_{(x,y)}(z) = (\rho x + \sqrt{1-\rho^2}z, \rho y + \sqrt{1-\rho^2}z)$ and $\tilde{\gamma}_{(x,y)}$ the pushforward of γ by $\Lambda_{(x,y)}$. Then an explicit form of Q_{cn} is given, for $A \in \mathcal{B}(\mathsf{H} \times \mathsf{H})$, by:

$$\begin{aligned} Q_{\mathrm{cn}}((x,y),A) = \\ \int_A \alpha(x,v) \wedge \alpha(y,t) \mathrm{d}\tilde{\gamma}_{(x,y)}(v,t) + \int_{\mathsf{H}\times\mathsf{H}} (\alpha(y,t) - \alpha(x,v))_+ \mathbb{1}_A(x,t) \mathrm{d}\tilde{\gamma}_{(x,y)}(v,t) \\ + \int_{\mathsf{H}\times\mathsf{H}} (\alpha(x,v) - \alpha(y,t))_+ \mathbb{1}_A(v,y) \mathrm{d}\tilde{\gamma}_{(x,y)}(v,t) \\ + \delta_{(x,y)}(A) \int_{\mathsf{H}\times\mathsf{H}} (1 - \alpha(x,v) \vee \alpha(y,t)) \mathrm{d}\tilde{\gamma}_{(x,y)}(v,t) \end{aligned} \tag{4.13}$$

where for $u \in \mathbb{R}$, $(u)_+ = \max(u, 0)$. The following Proposition shows that **H**2 is satisfied.

Proposition 4.12. *Assume* ***CN*** *1. There exists $\eta > 0$ such that, Q_{cn} is a d_η-weak contraction and for every $u > 1$, there exist $\ell \geq 1$ and $\epsilon > 0$ such that $\{V \leq u\}^2$ is a (ℓ, ϵ, d_η)-coupling set.*

Proof. See Section 4.5.2 □

Note that for all $\eta > 0$, d_η is Lipschitz equivalent to d_1, therefore W_{d_η} and W_{d_1} are Lipschitz equivalent. As a consequence of Proposition 4.11, Proposition 4.12, Theorem 4.3 and Theorem 4.4, we have

Theorem 4.13. *Let P_{cn} be the kernel of the preconditioned Crank-Nicolson algorithm with target density $\mathrm{d}\pi \propto \exp(-g)\mathrm{d}\gamma$ and design parameter $\rho \in [0,1)$. Assume* ***CN*** *1. Then P_{cn} admits π as a unique invariant probability measure and there exist C_1, C_2 such that for all $n \in \mathbb{N}^*$ and $x \in \mathsf{H}$*

$$W_{d_1}(P_{\mathrm{cn}}^n(x,\cdot), \pi) \leq C_1 V(x) \exp\left(-\kappa(\log(n) - C_2 \log(\log(n)))^{2/\beta}\right) ,$$

where V is given by (4.12), $d_1(x,y) = \|x-y\|^\beta \wedge 1$ and $\kappa = \theta C_g^{-2/\beta}/36$ for θ given by Theorem 4.10.

4.4 Proofs of Section 4.2

In this section, C is a constant which may take different values upon each appearance.

For $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$ and a canonical Markov chain on the space $((E \times E)^{\mathbb{N}}, (\mathcal{B}(E) \otimes \mathcal{B}(E))^{\otimes \mathbb{N}})$, denote by $T_0 = \inf\{n \geq \ell, (X_n, Y_n) \in \Delta\}$ the first return time to Δ after $\ell - 1$ steps. Then, define recursively for $j \geq 1$,

$$T_j = T_0 \circ \theta^{T_{j-1}} + T_{j-1} = T_0 + \sum_{k=0}^{j-1} T_0 \circ \theta^{T_k} , \tag{4.14}$$

where θ is the shift operator.

Let Q be a coupling kernel for P. Hereafter, $\{(X_n, Y_n), n \in \mathbb{N}\}$ is the canonical Markov chain on the space $((E\times E)^{\mathbb{N}}, (\mathcal{B}(E)\otimes\mathcal{B}(E))^{\otimes\mathbb{N}})$ with Markov kernel Q. We denote by $\widetilde{\mathbb{P}}_{x,y}$ and $\widetilde{\mathbb{E}}_{x,y}$ the associated canonical probability and expectation, respectively, when the initial distribution of the Markov chain is the Dirac mass at (x,y).

For any $n \in \mathbb{N}^\star$ and $x, y \in E$, the n-iterated kernel $Q^n((x,y),\cdot)$ is a coupling of $(P^n(x,\cdot), P^n(y,\cdot))$; hence $W_d(P^n(x,\cdot),P^n(y,\cdot)) \le \widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right]$. Define the filtration $\{\widetilde{\mathcal{F}}_n, n\ge 0\}$ by $\widetilde{\mathcal{F}}_n = \sigma((X_k,Y_k), k\le n)$.

Proposition 4.14. *Assume that there exists a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, $\ell \in \mathbb{N}^*$ and $\epsilon > 0$ such that* **H2**(Δ,ℓ,ϵ). *Then, for all $x, y \in E$, and $n \ge 0$, $m \ge 0$:*

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right] \le (1-\epsilon)^m + \widetilde{\mathbb{P}}_{x,y}\left[T_m \ge n\right] \ .$$

Proof. Set $Z_n = d(X_n,Y_n)$; under **H2**(Δ,ℓ,ϵ), $\{(Z_n,\widetilde{\mathcal{F}}_n)\}_{n\ge 0}$ is a bounded non-negative supermartingale and for all $(x,y)\in\Delta$, $\widetilde{\mathbb{E}}_{x,y}\left[Z_\ell\right] \le (1-\epsilon)d(x,y)$. Denote by Z_∞ its $\widetilde{\mathbb{P}}_{x,y}$-a.s limit. By the optional stopping theorem, we have for every $m \ge 0$: $\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_{m+1}} \middle| \widetilde{\mathcal{F}}_{T_m+\ell}\right] \le Z_{T_m+\ell}$. On the other hand, by the strong Markov, $\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_m+\ell}\middle|\widetilde{\mathcal{F}}_{T_m}\right] \le (1-\epsilon)Z_{T_m}$. By combining these two relations, we get: $\widetilde{\mathbb{E}}_{x,y}\left[Z_{T_{m+1}}\middle|\widetilde{\mathcal{F}}_{T_m}\right] \le (1-\epsilon)Z_{T_m}$. Since Z_n is upper bounded by 1, the proof follows from [JT01b, lemma 3.1]. $\square$

4.4.1 Proof of Theorem 4.1

By Proposition 4.14 and the Markov inequality for all $m \ge 0$, we get

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right] \le (1-\epsilon)^m + n^{-1}\widetilde{\mathbb{E}}_{x,y}\left[T_m\right] \ . \tag{4.15}$$

Using (4.14) and the strong Markov property, we obtain $\widetilde{\mathbb{E}}_{x,y}\left[T_m\right] = \widetilde{\mathbb{E}}_{x,y}\left[T_0\right] + \widetilde{\mathbb{E}}_{x,y}\left[\sum_{k=0}^{m-1}\widetilde{\mathbb{E}}_{X_{T_k},Y_{T_k}}\left[T_0\right]\right]$. Using [MT09, Proposition 11.3.3] and the Markov property we have that

$$\widetilde{\mathbb{E}}_{x,y}\left[T_0\right] \le Q^{\ell-1}\mathcal{V}(x,y) + b + \ell - 1 \ ,$$

which implies that $\widetilde{\mathbb{E}}_{x,y}\left[T_m\right] \le m \sup_{(x,y)\in\Delta} Q^{\ell-1}\mathcal{V}(x,y) + Q^{\ell-1}\mathcal{V}(x,y) + (m+1)\left(b+\ell-1\right)$, where the constant b is defined in (4.3). Plugging this inequality into (4.15) and taking $m = \lceil -\log(n)/\log(1-\epsilon)\rceil$ implies that there exists $C < \infty$ satisfying

$$Q^n d(x,y) = \widetilde{\mathbb{E}}_{x,y}\left[d(X_n,Y_n)\right] \le C(\log(n)/n)Q^{\ell-1}\mathcal{V}(x,y) \le C(\log(n)/n)\mathcal{V}(x,y) \ , \tag{4.16}$$

where we have used that $Q^{\ell-1}\mathcal{V}(x,y) \le \mathcal{V}(x,y) + b(\ell-1)$ (the constant C takes different values upon each appearance).

Uniqueness of the invariant probability

The proof is by contradiction. Assume that there exist two invariant distributions π and ν, and let $\lambda \in \Pi(\pi,\nu)$. According to Lemma 4.23-(i), we have for every integer n,

$$W_d(\pi,\nu) = W_d(\pi P^n, \nu P^n) \le \int_{E\times E} Q^n d(x,y)\,\lambda(\mathrm{d}x,\mathrm{d}y) \ .$$

We prove that the RHS converges to zero by application of the dominated convergence theorem. It follows from (4.16) that for all $x, y \in E$ and $n \geq 0$, $g_n(x,y) \stackrel{\text{def}}{=} Q^n d(x,y) \leq C\mathcal{V}(x,y) \ \log(n)/n$ for some $C < \infty$. Therefore, the sequence of functions $(g_n)_{n\in\mathbb{N}}$ converges pointwise to 0. Since $d \leq 1$, $g_n(x,y) \leq 1$. Hence, by the Lebesgue theorem, $\int_{E\times E} g_n(x,y)\,\lambda(\mathrm{d}x,\mathrm{d}y) \underset{n\to+\infty}{\longrightarrow} 0$ showing that $W_d(\pi,\nu) = 0$, or equivalently $\nu = \pi$ since W_d is a distance on $\mathcal{P}(E)$.

Existence of an invariant measure

Let $x_0 \in E$. We first show that there exists $\{m_k, k \in \mathbb{N}\}$ such that $\{P^{m_k}(x_0,\cdot), k \in \mathbb{N}\}$ is a Cauchy sequence for W_d. Let $n, k \in \mathbb{N}^*$ and choose $M \geq 1$. By Lemma 4.23-(i):

$$W_d(P^n(x_0,\cdot), P^{n+n_k}(x_0,\cdot)) \leq \inf_{\lambda\in\ \Pi(\delta_{x_0}, P^{n_k}(x_0,\cdot))} \left\{ \int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)\geq M\}} Q^n\, d(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) + \int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)<M\}} Q^n d(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) \right\} . \tag{4.17}$$

We consider separately the two terms. Set $M_\psi = \sup_k P^{n_k}[\psi \circ \mathcal{V}_{x_0}](x_0)$. Let $\lambda \in \Pi(\delta_{x_0}, P^{n_k}(x_0,\cdot))$. Since d is bounded by 1, we get

$$\begin{aligned}
&\int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)\geq M\}} Q^n d(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) \\
&\leq \int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)\geq M\}} \lambda(\mathrm{d}z,\mathrm{d}t) \leq P^{n_k}(x_0, \{\mathcal{V}_{x_0} \geq M\}) \\
&\leq P^{n_k}(x_0, \{\psi\circ\mathcal{V}_{x_0} \geq \psi(M)\}) \leq P^{n_k}[\psi\circ\mathcal{V}_{x_0}](x_0)/\psi(M) \leq M_\psi/\psi(M) ,
\end{aligned} \tag{4.18}$$

where we have used (4.4) and the Markov inequality. In addition by (4.16), there exists $C > 0$ such that:

$$\int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)<M\}} Q^n d(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) \leq C(\log(n)/n) \int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)<M\}} \mathcal{V}(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) .$$

Furthermore, $x \mapsto \psi(x)/x$ is non-increasing so that $\mathcal{V}(z,t) \leq M\psi(\mathcal{V}(z,t))/\psi(M)$ on $\{\mathcal{V}(z,t) \leq M\}$. This inequality and (4.4) imply

$$\int_{E\times E} \mathbb{1}_{\{\mathcal{V}(z,t)<M\}} Q^n d(z,t)\lambda(\mathrm{d}z,\mathrm{d}t) \leq C(\log(n)/n) M_\psi M/\psi(M) . \tag{4.19}$$

Plugging (4.18) and (4.19) in (4.17), we have for every $M > 0$, $n, k \in \mathbb{N}^*$

$$W_d(P^n(x_0,\cdot), P^{n+n_k}(x_0,\cdot)) \leq \frac{M_\psi}{\psi(M)} + C(\log(n)/n)\left(M_\psi M/\psi(M)\right) .$$

Setting $M = n/\log(n)$, we get that for all $n, k \in \mathbb{N}^*$

$$W_d(P^n(x_0,\cdot), P^{n+n_k}(x_0,\cdot)) \leq C/\psi(n/\log(n)) . \tag{4.20}$$

Since $\lim_{x\to+\infty}\psi(x)=+\infty$ and $\lim_{k\to+\infty}n_k=+\infty$ there exists $\{u_k, k\in\mathbb{N}\}$ such that $u_0=1$ and for $k\geq 1$, $u_k=\inf\{n_l \mid l\in\mathbb{N};\psi(n_l/\log(n_l))\geq 2^k\}$. Set $m_k=\sum_{i=0}^k u_i$. Since for all $k\in\mathbb{N}$, $m_{k+1}=m_k+u_{k+1}$, by (4.20),

$$W_d(P^{m_k}(x_0,\cdot),P^{m_{k+1}}(x_0,\cdot))\leq C2^{-k}\ ,$$

which implies that the series $\sum_k W_d(P^{m_k}(x_0,\cdot),P^{m_{k+1}}(x_0,\cdot))$ converges and $(P^{m_k}(x_0,\cdot))_{k\in\mathbb{N}}$ is a Cauchy sequence in $(\mathcal{P}(E),W_d)$.

Since $(\mathcal{P}(E),W_d)$ is Polish, there exists $\pi\in\mathcal{P}(E)$ such that

$$\lim_{k\to+\infty}W_d(P^{m_k}(x_0,\cdot),\pi)=0\ .$$

The second step is to prove that π is invariant. Since $\lim_{k\to+\infty}W_d(P^{m_k}(x_0,\cdot),\pi)=0$, by the triangular inequality it holds

$$W_d(\pi,\pi P)\leq\lim_{k\to+\infty}W_d(P^{m_k}(x_0,\cdot),\delta_{x_0}PP^{m_k})+\lim_{k\to+\infty}W_d(\delta_{x_0}P^{m_k}P,\pi P)\ . \tag{4.21}$$

By Lemma 4.23-(i) and (4.16) , there exists C such that for any $k\geq 1$,

$$\begin{aligned}W_d(P^{m_k}(x_0,\cdot),\delta_{x_0}P^{m_k+1})&\leq\inf_{\lambda\in\Pi(\delta_{x_0},\delta_{x_0}P)}\int_{E\times E}Q^{m_k}d(z,t)\mathrm{d}\lambda(z,t)\\&\leq C(\log(m_k)/m_k)\inf_{\lambda\in\Pi(\delta_{x_0},\delta_{x_0}P)}\int_{E\times E}\mathcal{V}(z,t)\lambda(\mathrm{d}z,\mathrm{d}t)\leq C(\log(m_k)/m_k)P\mathcal{V}_{x_0}(x_0)\ .\end{aligned}$$

By definition, $\lim_k m_k=+\infty$ so that by (4.4), the RHS converges to 0 when $k\to+\infty$. In addition, by Lemma 4.23-(ii), $W_d(\delta_{x_0}P^{m_k}P,\pi P)\leq W_d(P^{m_k}(x_0,\cdot),\pi)$, and this RHS converges to 0 by definition of π. Plugging these results in (4.21) yields $W_d(\pi,\pi P)=0$, and therefore $\pi P=\pi$.

4.4.2 Proof of Corollary 4.2

We prove that the assumptions of Theorem 4.1 are satisfied. Set $\mathcal{V}(x,y)=1+(V(x)+V(y))/\phi(2)$. Since Q is a coupling kernel for P, it holds

$$Q\mathcal{V}(x,y)=1+\frac{1}{\phi(2)}\ ,(PV(x)+PV(y))\leq\mathcal{V}(x,y)-\frac{\phi(V(x)+V(y))}{\phi(2)}+b\mathbb{1}_\Delta(x,y)\ .$$

This yields the drift inequality (4.3) upon noting that ϕ is increasing and $V\geq 1$ so that $\phi(V(x)+V(y))/\phi(2)\geq 1$. By iterating this inequality, we have for any ℓ,

$$\sup_{(x,y)\in\Delta}Q^{\ell-1}\mathcal{V}(x,y)\leq\ell\sup_{(x,y)\in\Delta}\mathcal{V}(x,y)+b\ell\ ,$$

and the RHS is finite since by assumption, $\sup_{(x,y)\in\Delta}(V(x)+V(y))<\infty$.

Under **H** 3(Δ,ϕ,V), $PV(x)\leq PV(x)+PV(x_0)\leq V(x)-\phi\circ V(x)+b+V(x_0)$ where we have used that $\phi(V(x)+V(x_0))\geq\phi(V(x))$. This implies that for every

$n \in \mathbb{N}^*$, $n^{-1}\sum_{k=0}^{n-1} P^k(\phi \circ V)(x) \le b + V(x_0) + V(x)/n$ and $\pi(\phi \circ V) < \infty$; see [But14, lemma 4.1]. For any x, we have $P\mathcal{V}_x(x) < \infty$. Finally, since $\phi \in \mathbb{F}$, we can set $\psi = \phi$. Let us define the increasing sequence $\{n_k, k \in \mathbb{N}\}$. Set $M_\phi > b + V(x_0)$; there exists an increasing sequence $\{n_k, k \in \mathbb{N}\}$ such that $\lim_k n_k = +\infty$ and

$$P^{n_k}(\phi \circ V)(x_0) \le M_\phi \ , \ \text{for all } k \in \mathbb{N} \ . \tag{4.22}$$

4.4.3 Proof of Theorem 4.3

We preface the proof by some preliminary technical results. By using Proposition 4.14, for every $x, y \in E$ and $m \ge 0$, $\widetilde{\mathbb{E}}_{x,y}[d(X_n, Y_n)] \le (1-\epsilon)^m + \widetilde{\mathbb{P}}_{x,y}[T_m > n]$. The crux of the proof is to obtain estimates of tails of the successive return times to Δ. Following [TT94], we start by considering a sequence of drift conditions on the product space $E \times E$. For $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathcal{V}_n, n \in \mathbb{N}\}$, $\mathcal{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and a constant $b < \infty$, let us consider the following assumption:

A *$(\Delta, \ell, \mathcal{V}_n, r, b)$ For all $x, y \in E$:*

$$Q\mathcal{V}_{n+1}(x,y) \le \mathcal{V}_n(x,y) - r(n) + br(n)\mathbb{1}_\Delta(x,y) \ , \quad \textit{and} \qquad \sup_{(x,y)\in\Delta} Q^{\ell-1}\mathcal{V}_0(x,y) < \infty \ .$$

Under $\mathbf{A}(\Delta, \ell, \mathcal{V}_n, r, b)$, we first obtain bounds on the moments $\widetilde{\mathbb{E}}_{x,y}[R(T_0)]$ for $x, y \in E$ (see Proposition 4.15), where

$$R(t) = 1 + \int_0^t r(s)\mathrm{d}s \ , t \ge 0 \ . \tag{4.23}$$

We will then deduce bounds for $\widetilde{\mathbb{P}}_{x,y}[T_m \ge n]$ (see Lemma 4.17). Set

$$c_{1,r} = \sup_{k\in\mathbb{N}^*} R(k)/\sum_{i=0}^{k-1} r(i) \ , \quad c_{2,r} = \sup_{m,n\in\mathbb{N}} R(m+n)/\{R(m)R(n)\} \ . \tag{4.24}$$

It follows from Lemma 4.24 that these constants are finite.

Proposition 4.15. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathcal{V}_n, n \in \mathbb{N}\}$, $\mathcal{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and a constant $b < \infty$ such that $\boldsymbol{A}(\Delta, \ell, \mathcal{V}_n, r, b)$ is satisfied. Then, for any $x, y \in E$,*

$$\widetilde{\mathbb{E}}_{x,y}[R(T_0)] \le c_{1,r}c_{2,r}R(\ell-1)\{Q^{\ell-1}\mathcal{V}_0(x,y) + br(0)\} \ , \tag{4.25}$$

and $\sup_{(z,t)\in\Delta} \widetilde{\mathbb{E}}_{z,t}[R(T_0)]$ is finite.

Proof. By [MT09, Proposition 11.3.2], $\widetilde{\mathbb{E}}_{x,y}\left[\sum_{k=0}^{\tau_\Delta - 1} r(k)\right] \le \mathcal{V}_0(x,y) + br(0)$, where τ_Δ is the return time to Δ. Since $R(k) \le c_{1,r}\sum_{p=0}^{k-1} r(p)$, the previous inequality provides a bound on $\widetilde{\mathbb{E}}_{x,y}[R(\tau_\Delta)]$. The conclusion follows from the Markov property upon noting that $R(T_0) \le c_{2,r}R(\ell-1)R(\tau_\Delta) \circ \theta^{\ell-1}$. $\square$

Combining the strong Markov property, (4.14) and Proposition 4.15, it is easily seen that $\widetilde{\mathbb{E}}_{x,y}[T_m] < \infty$ for any $m \geq 0$ and $x, y \in E$. This yields the following result.

Corollary 4.16. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathcal{V}_n, n \in \mathbb{N}\}$, $\mathcal{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and a constant $b < \infty$ such that $\boldsymbol{A}(\Delta, \ell, \mathcal{V}_n, r, b)$ is satisfied. Then, for all $j \geq 0$ and $(x, y) \in E \times E$, $\widetilde{\mathbb{P}}_{x,y}[T_j < \infty] = 1$.*

For $r \in \Lambda$, there exists $r_0 \in \Lambda_0$ such that $c_{3,r} = 1 \wedge \sup_{t\geq 0} r(t)/r_0(t) < \infty$ and $c_{4,r} = 1 \vee \sup_{t \geq 0} r_0(t)/r(t) < \infty$. Denote $c_{5,r} = \sup_{t,u\in\mathbb{R}_+} r(t+u)/\{r(t)r(u)\}$ and define for $\kappa > 0$, the real M_κ such that for all $t \geq M_\kappa$, $r(t) \leq \kappa R(t)$. M_κ is well defined by Lemma 4.24-(iii).

Lemma 4.17. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathcal{V}_n, n \in \mathbb{N}\}$, $\mathcal{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and constants $\epsilon > 0$, $b < \infty$ such that $\boldsymbol{H}2(\Delta, \ell, \epsilon)$ and $\boldsymbol{A}(\Delta, \ell, \mathcal{V}_n, r, b)$ are satisfied. Then,*

(i) For all $x, y \in E$ and for all $n \in \mathbb{N}$,$m \in \mathbb{N}^*$,

$$\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq \{a_1 Q^{\ell-1}\mathcal{V}_0(x, y) + a_2\}/R(n/2) + a_3/R(n/(2m)) .$$

(ii) For all $\kappa > 0$, for all $x, y \in E$ and for all $n, m \in \mathbb{N}$,

$$\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq (1 + b_1\kappa)^m \{\kappa^{-1} r(M_\kappa) + a_1 Q^{\ell-1}\mathcal{V}_0(x, y) + b_2\}/R(n) ,$$

The constants $\{a_i, b_i\}_{i=1}^3$ can be directly obtained from the proof

Proof. Since $r \in \Lambda$, there exists $r_0 \in \Lambda_0$ such that $c_{3,r} + c_{4,r} < \infty$. Denote by R_0 the function (4.23) associated with r_0.

$$\begin{aligned}
&\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq \widetilde{\mathbb{P}}_{x,y}[T_0 \geq n/2] + \widetilde{\mathbb{P}}_{x,y}[T_m - T_0 \geq n/2] \\
&\leq \widetilde{\mathbb{E}}_{x,y}[R(T_0)]/R(n/2) + \widetilde{\mathbb{E}}_{x,y}[R_0((T_m - T_0)/m)]/R_0(n/(2m)) \\
&\leq \{a_1 Q^{\ell-1}\mathcal{V}_0(x, y) + a_2\}/R(n/2) + c_{4,r}\widetilde{\mathbb{E}}_{x,y}[R_0((T_m - T_0)/m)]/R(n/(2m)) ,
\end{aligned} \tag{4.26}$$

where we used Proposition 4.15 in the last inequality, and $a_1 = c_{1,r}c_{2,r}R(\ell - 1)$; $a_2 = a_1 b r(0)$. Since R_0 is convex (see Lemma 4.24), we have by (4.14):

$$\widetilde{\mathbb{E}}_{x,y}[R_0((T_m - T_0)/m)] \leq (c_{3,r}m)^{-1}\widetilde{\mathbb{E}}_{x,y}\left[\sum_{k=0}^{m-1} R(T_0 \circ \theta^{T_k})\right] .$$

Using Corollary 4.16 and the strong Markov property, for any $x, y \in E$ and $m \geq 1$,

$$\widetilde{\mathbb{E}}_{x,y}[R_0((T_m - T_0)/m)] \leq C_\Delta/c_{3,r} , \qquad \text{with } C_\Delta = \sup_{(x,y)\in\Delta} \widetilde{\mathbb{E}}_{x,y}[R(T_0)] . \tag{4.27}$$

Plugging (4.27) in (4.26) implies (i) with $a_3 = c_{4,r}C_\Delta/c_{3,r}$. We now consider (ii). Again by the Markov inequality, since R is increasing,

$$\widetilde{\mathbb{P}}_{x,y}[T_m \geq n] \leq R^{-1}(n)\widetilde{\mathbb{E}}_{x,y}[R(T_m)] . \tag{4.28}$$

If $m = 0$, the result follows from Proposition 4.15. If $m \geq 1$, using the definitions of T_m and R, given respectively in (4.14) and (4.23), and since for all $t, u \in \mathbb{R}_+$, $R(t+u) \leq R(t) + c_{5,r} R(u) r(t)$, we get

$$\widetilde{\mathbb{E}}_{x,y}\left[R(T_m)\right] \leq \widetilde{\mathbb{E}}_{x,y}\left[R(T_{m-1})\right] + c_{5,r} \widetilde{\mathbb{E}}_{x,y}\left[r(T_{m-1}) R(T_0 \circ \theta^{T_{m-1}})\right] .$$

Thus, by the strong Markov property

$$\widetilde{\mathbb{E}}_{x,y}\left[R(T_m)\right] \leq \widetilde{\mathbb{E}}_{x,y}\left[R(T_{m-1})\right] + c_{5,r} C_\Delta \widetilde{\mathbb{E}}_{x,y}\left[r(T_{m-1})\right] . \tag{4.29}$$

Let $\kappa > 0$. Since by definition, for all $t \geq M_\kappa$, $r(t) \leq \kappa R(t)$, $\widetilde{\mathbb{E}}_{x,y}\left[r(T_{m-1})\right] \leq r(M_\kappa) + \kappa \widetilde{\mathbb{E}}_{x,y}\left[R(T_{m-1})\right]$, so that (4.29) becomes

$$\widetilde{\mathbb{E}}_{x,y}\left[R(T_m)\right] \leq (1 + c_{5,r} C_\Delta \kappa) \widetilde{\mathbb{E}}_{x,y}\left[R(T_{m-1})\right] + c_{5,r} C_\Delta r(M_\kappa) .$$

By a straightforward induction we get,

$$\widetilde{\mathbb{E}}_{x,y}\left[R(T_m)\right] \leq (1 + c_{5,r} C_\Delta \kappa)^m (\widetilde{\mathbb{E}}_{x,y}\left[R(T_0)\right] + r(M_\kappa)/\kappa) .$$

Plugging this result in (4.28), using Proposition 4.15 and setting $b_1 = c_{5,r} C_\Delta$ and $b_2 = a_1(C_\Delta + br(0))$ conclude the proof. □

Lemma 4.18. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, $\ell \in \mathbb{N}^*$, a sequence of measurable functions $\{\mathcal{V}_n, n \in \mathbb{N}\}$, $\mathcal{V}_n : E \times E \to \mathbb{R}_+$, a function $r \in \Lambda$ and constants $\epsilon > 0$, $b < \infty$ such that $\boldsymbol{H}2(\Delta, \ell, \epsilon)$ and $\boldsymbol{A}(\Delta, \ell, \mathcal{V}_n, r, b)$ are satisfied. Then,*

(i) For all $x, y \in E$ and $n \in \mathbb{N}$,

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n, Y_n)\right] \leq 1/R(n) + \{a_1 Q^{\ell-1} \mathcal{V}_0(x,y) + a_2\}/R(n/2) + a_3 v_n^{-1} ,$$

where $v_n \stackrel{def}{=} R(-n \log(1-\epsilon)/\{2(\log(R(n)) - \log(1-\epsilon))\})$.

(ii) For all $\delta \in (0,1)$, $x, y \in E$ and $n \in \mathbb{N}$,

$$\widetilde{\mathbb{E}}_{x,y}\left[d(X_n, Y_n)\right] \leq \left(1 + (1 + b_1\kappa)\{\kappa^{-1} r(M_\kappa) + a_1 Q^{\ell-1}\mathcal{V}_0(x,y) + b_2\}\right)/R^\delta(n) ,$$

where $\kappa = ((1-\epsilon)^{-(1-\delta)/\delta} - 1)/b_1$.

The constants a_i, b_j are given by Lemma 4.17.

Proof. By Proposition 4.14 and Lemma 4.17-(i), there exists C such that for all x, y in E and for all $n \geq 0$ and $m \geq 0$

$$\begin{aligned}\widetilde{\mathbb{E}}_{x,y}\left[d(X_n, Y_n)\right] &\leq (1-\epsilon)^m + \widetilde{\mathbb{P}}_{x,y}\left[T_m \geq n\right] \\ &\leq (1-\epsilon)^m + \{a_1 Q^{\ell-1}\mathcal{V}_0(x,y) + a_2\}/R(n/2) + a_3/R(n/(2m)) .\end{aligned}$$

We get the first inequality by choosing $m = \lceil -\log(R(n))/\log(1-\epsilon) \rceil$. Let us prove (ii). Fix $\delta \in (0,1)$ and choose the smallest integer m such that $(1-\epsilon)^m \leq R(n)^{-\delta}$

(i.e. $m = \lceil -\delta \log R(n) / \log(1-\epsilon) \rceil$). Apply Lemma 4.17-(ii), with $\kappa > 0$ such that $(1+b_1\kappa) = (1-\epsilon)^{-((1-\delta)/\delta)}$; hence, upon noting that $R(n)^{-\delta} \leq (1-\epsilon)^{m-1}$, it holds

$$(1+b_1\kappa)^m = (1+b_1\kappa)\left\{(1-\epsilon)^{m-1}\right\}^{-((1-\delta)/\delta)} \leq (1+b_1\kappa)\left\{R(n)^{-\delta}\right\}^{-((1-\delta)/\delta)} = (1+b_1\kappa)R(n)^{1-\delta} .$$

□

We now prove that **H**3(Δ, ϕ, V) implies **A**. For a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1, \infty)$, set

$$r_\phi(t) = (H_\phi^{\leftarrow})'(t) = \phi(H_\phi^{\leftarrow}(t)) , \tag{4.30}$$

where H_ϕ is defined in (4.6) and $H_\phi^{\leftarrow}$ denotes its inverse; and define for $k \geq 0$, $H_k : [1, \infty) \to \mathbb{R}_+$ and $\mathcal{V}_k : E \times E \to \mathbb{R}_+$ by

$$H_k(u) = \int_0^{H_\phi(u)} r_\phi(t+k)\mathrm{d}t = H_\phi^{-1}(H_\phi(u)+k) - H_\phi^{-1}(k) , \tag{4.31}$$

$$\mathcal{V}_k(x,y) = H_k\left(V(x)+V(y)\right) . \tag{4.32}$$

Note that $\mathcal{V}_k$ is measurable, H_k is twice continuously differentiable on $[1, \infty)$ and that $H_0(x) \leq x$ so $\mathcal{V}_0(x,y) \leq V(x)+V(y)$. The proof of the following lemma is adapted from [Dou+04, Proposition 2.1].

Lemma 4.19. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E \times E)$, a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1, \infty)$ such that* ***H****3(Δ, ϕ, V) is satisfied. For any $x, y \in E$ and any transference plan $\lambda \in \Pi(P(x, \cdot), P(y, \cdot))$ we have:*

$$\int_{E\times E} \mathcal{V}_{k+1}(z,t)\mathrm{d}\lambda(z,t) \leq \mathcal{V}_k(x,y) - r_\phi(k) + \frac{b}{r_\phi(0)} r_\phi(k+1)\mathbb{1}_\Delta(x,y) ,$$

where r_ϕ and $\mathcal{V}_k$ are defined in (4.30) and (4.32) respectively.

Proof. Set $\mathcal{V}(x,y) = V(x)+V(y)$. This implies that for all $u \geq 1$ and $t \in \mathbb{R}$ such that $t+u \geq 1$, we have

$$H_{k+1}(t+u) - H_{k+1}(u) \leq H'_{k+1}(u)t . \tag{4.33}$$

In addition, according to [Dou+04, Proposition 2.1] and its proof, H_{k+1} is concave and for every $u \geq 1$

$$H_{k+1}(u) - \phi(u)H'_{k+1}(u) \leq H_k(u) - r_\phi(k) . \tag{4.34}$$

Therefore, the Jensen inequality and (4.5) imply

$$\int_{E\times E} \mathcal{V}_{k+1}(z,t)\mathrm{d}\lambda(z,t) \leq H_{k+1}\left(\int_{E\times E} \mathcal{V}(z,t)\mathrm{d}\lambda(z,t)\right) \leq H_{k+1}\left(\mathcal{V}(x,y) - \phi\circ\mathcal{V}(x,y) + b\mathbb{1}_\Delta(x,y)\right) .$$

Using (4.33), (4.34) and the inequality $H'_{k+1}(\mathcal{V}(x,y)) \leq H'_{k+1}(1)$ we get that

$$\begin{aligned}
&\int_{E\times E} \mathcal{V}_{k+1}(z,t)\mathrm{d}\lambda(z,t) \\
&\leq H_{k+1}\left(\mathcal{V}(x,y)\right) - \phi\circ\mathcal{V}(x,y)\, H'_{k+1}(\mathcal{V}(x,y)) + bH'_{k+1}(1)\mathbb{1}_\Delta(x,y) \\
&\leq H_k\left(\mathcal{V}(x,y)\right) - r_\phi(k) + bH'_{k+1}(1)\mathbb{1}_\Delta(x,y) \; .
\end{aligned}$$

The proof is concluded upon noting that $H'_{k+1}(1) = r_\phi(k+1)/r_\phi(0)$. □

Proposition 4.20. *Assume that there exist a coupling kernel Q for P, $\Delta \in \mathcal{B}(E\times E)$, a function $\phi \in \mathbb{F}$ and a measurable function $V : E \to [1,\infty)$ such that $\boldsymbol{H}3(\Delta,\phi,V)$ is satisfied. Then for any $\ell \geq 0$, $\boldsymbol{A}(\Delta,\ell,\mathcal{V}_n,r_\phi,\{\sup_{p\geq 0} r_\phi(p+1)/r_\phi(p)\}\, b/r_\phi(0))$ holds with $\mathcal{V}_n(x,y) = H_n(V(x)+V(y))$ where r_ϕ and H_n are given by (4.30) and (4.31) respectively.*

Proof. By [Dou+04, Lemma 2.3], $r_\phi \in \Lambda$. Then, it follows from Lemma 4.19 and Lemma 4.24-(i) that

$$Q\mathcal{V}_{k+1}(x,t) \leq \mathcal{V}_k(x,y) - r_\phi(k) + b\left\{\sup_{p\geq 0} r_\phi(p+1)/r_\phi(p)\right\} r_\phi(k)\mathbb{1}_\Delta(x,y)/r_\phi(0) \; .$$

Finally, since Q^ℓ is a coupling kernel for P^ℓ, we have by iterating the inequality (4.5)

$$Q^\ell\mathcal{V}_0(x,y) \leq P^\ell V(x) + P^\ell V(y) \leq V(x) + V(y) + \ell b \; .$$

Therefore under $\mathbf{H}3(\Delta,\phi,V)$, $\sup_{(x,y)\in\Delta} Q^{\ell-1}\mathcal{V}_0(x,y) < +\infty$. □

Proof of Theorem 4.3-(i). Using Proposition 4.20, Lemma 4.18 applies with $R(t) = 1 + \int_0^t r_\phi(s)\mathrm{d}s$ for $t\in\mathbb{R}_+$. Note that we have $R = H_\phi^{\leftarrow}$.

Set $M_V > 0$ such that $\pi(V \leq M_V) \geq 1/2$; such a constant exists since $\pi(E) = 1$ and $E = \bigcup_{k\in\mathbb{N}}\{V \leq k\}$. Set $M > M_V$ and define the probability π_M by $\pi_M(\cdot) = \pi(\cdot \cap \{V \leq M\})/\pi(\{V \leq M\})$. Since π is invariant for P, $W_d(P^n(x,\cdot),\pi) = W_d(P^n(x,\cdot),\pi P^n)$ and the triangle inequality implies:

$$W_d(P^n(x,\cdot),\pi) \leq W_d(P^n(x,\cdot),\pi_M P^n) + W_d(\pi_M P^n,\pi P^n)\;, \quad \text{for all } n\geq 1. \tag{4.35}$$

Consider the first term in the RHS of (4.35). By Lemma 4.23-(i), for all $x\in E$ and $n \geq 1$:

$$W_d(P^n(x,\cdot),\pi_M P^n) \leq \inf_{\lambda\in\ \Pi(\delta_x,\pi_M)} \int_{E\times E} Q^n d(z,t)\,\mathrm{d}\lambda(z,t) \; .$$

Let $v_n = R(-n\log(1-\epsilon)/\{2(\log(R(n)) - \log(1-\epsilon))\})$. By Lemma 4.18-(i) and since $R = H_\phi^{\leftarrow}$ is increasing, for all $x \in E$ and $n\geq 1$

$$\begin{aligned}
&R(n/2)\, W_d(P^n(x,\cdot),\pi_M P^n) \\
&\quad\leq R(n/2)/R(n) + a_2 + a_3 R(n/2)/v_n \\
&\qquad\quad + a_1 \inf_{\lambda\in\ \Pi(\delta_x,\pi_M)} \int_{E\times E} (P^{\ell-1}V(z) + P^{\ell-1}V(t))\,\mathrm{d}\lambda(z,t) \\
&\quad\leq a_1\left(V(x) + \int_E V(t)\mathrm{d}\pi_M(t) + b(\ell-1)\right) + a_2 + 1 + a_3 R(n/2)/v_n \;,
\end{aligned} \tag{4.36}$$

where in the last inequality, we used

$$P^k V(x) \le V(x) + bk/2 \ . \tag{4.37}$$

which is obtained by iterating the drift inequality (4.5) and applying it with $x = y$. Since $x \mapsto \phi(x)/x$ is non-increasing, $V(t) \le M\phi(V(t))/\phi(M)$ on $\{V \le M\}$, we have

$$\int_E V(t) \mathrm{d}\pi_M(t) \le 2\pi(\phi \circ V) M/\phi(M) \ . \tag{4.38}$$

Note that by Corollary 4.2, $M_\phi = \pi(\phi \circ V) < \infty$. Combining (4.36) and (4.38) yield

$$\begin{aligned} &W_d(P^n(x,\cdot), \pi_M P^n) \\ &\qquad \le \{a_1 (V(x) + 2M_\phi M/\phi(M) + b(\ell - 1)) + a_2 + 1\}/R(n/2) + a_3/v_n \ . \end{aligned} \tag{4.39}$$

Consider the second term in the RHS of (4.35). Since d is bounded by 1, $W_d(\mu,\nu) \le W_{d_0}(\mu,\nu)$ (where W_{d_0} is the total variation distance) and Lemma 4.23-(ii) implies $W_d(\pi_M P^n, \pi P^n) \le W_d(\pi_M, \pi) \le W_{d_0}(\pi_M, \pi)$. For every $A \in \mathcal{B}(E)$, we get

$$\begin{aligned} |\pi_M(A) - \pi(A)| &= |\pi_M(A)(1 - \pi(\{V \le M\})) + \pi_M(A)\pi(V \le M) - \pi(A)| \\ &\le 2\pi(\{V > M\}) \ , \end{aligned}$$

showing that

$$W_d(\pi_M P^n, \pi P^n) \le 2\pi(\{V > M\}) = 2\pi\left(\{\phi(V) > \phi(M)\}\right) \le 2M_\phi/\phi(M) \ . \tag{4.40}$$

Since $R(n/2) > M_V$ for all n large enough, we can now choose $M = R(n/2)$ in (4.39) and (4.40). This yields

$$\begin{aligned} &W_d(P^n(x,\cdot), \pi) \\ &\quad \le \{a_1 (V(x) + b(\ell - 1)) + a_2 + 1\}/H_\phi^{\leftarrow}(n/2) + 2M_\phi(a_1 + 1)/\phi(R(n/2)) + a_3/v_n \ . \end{aligned}$$

(ii) The proof is along the same lines, using Lemma 4.18-(ii) instead of Lemma 4.18-(i). Finally, we end up with the following inequality for n large enough:

$$\begin{aligned} W_d(P^n(x,\cdot), \pi) \le (1 + (1 + b_1\kappa)\{\kappa^{-1} r_\phi(M_\kappa) + a_1(V(x) + b(\ell - 1)) + b_2\})/\{R^\delta(n)\} \\ + 2M_\phi((1 + b_1\kappa)a_1 + 1)/\{\phi(R^\delta(n))\} \ , \end{aligned}$$

where $\kappa = ((1 - \epsilon)^{-(1-\delta)/\delta} - 1)/b_1$. □

4.4.4 Proof of Theorem 4.4

Note that since $c = 1 - 2b/\phi(v)$ and $v > \phi^{\leftarrow}(2b)$, we get $c \in (0,1)$. By (4.7),

$$PV(x) + PV(y) \le V(x) + V(y) - c\phi\left(V(x) + V(y)\right) + 2b\mathbb{1}_{\mathtt{C}\times\mathtt{C}}(x,y) + \Omega(x,y)$$

where $\Omega(x,y) = c\phi\left(V(x) + V(y)\right) - \phi(V(x)) - \phi(V(y)) + 2b\mathbb{1}_{(\mathtt{C}\times\mathtt{C})^c}(x,y)$. We show that for every $x, y \in E$, $\Omega(x,y) \le 0$. Since ϕ is sub-additive (note that $\phi(0) = 0$), for all $x, y \in E$

$$\Omega(x,y) \le -(1 - c)\left(\phi(V(x)) + \phi(V(y))\right) + 2b\mathbb{1}_{(\mathtt{C}\times\mathtt{C})^c}(x,y) \ .$$

On $(\mathtt{C} \times \mathtt{C})^c$, $\phi(V(x)) + \phi(V(y)) \ge \phi(v)$. The definition of c implies that $\Omega(x,y) \le 0$.

4.5 Proofs of Section 4.3.3

Lemma 4.21. *Let $M > 0$. Assume that there exists an increasing continuoulsy differentiable concave function $\phi : [M, \infty) \to \mathbb{R}_+$, such that $\lim_{x\to\infty} \phi'(x) = 0$ and satisfying, on $\{V \geq M\}$, $PV(x) \leq V(x) - \phi \circ V(x) + b$. Then, there exist $\tilde{\phi} \in \mathbb{F}$ and $\tilde{b}$ such that, $PV \leq V - \tilde{\phi} \circ V + \tilde{b}$ on E, $\phi(v) = \tilde{\phi}(v)$ for all v large enough, and $\tilde{\phi}(0) = 0$.*

Proof. Observe indeed that the function $\tilde{\phi}$ defined by

$$\tilde{\phi}(t) = \begin{cases} (2\phi'(M) - \frac{\phi(M)}{M})t + \frac{2(\phi(M) - M\phi'(M))}{\sqrt{M}}\sqrt{t} & \text{for } 0 \leq t < M \\ \phi(t) & \text{for } t \geq M \ , \end{cases}$$

is concave increasing and continuously differentiable on $[1, +\infty)$, $\tilde{\phi}(0) = 0$, $\lim_{v\to\infty} \tilde{\phi}(v) = \infty$ and $\lim_{v\to\infty} \tilde{\phi}'(v) = 0$. The drift inequality (4.5) implies that for all $x \in E$

$$PV(x) \leq V(x) - \tilde{\phi}\,(V(x)) + \tilde{b} \ ,$$

with $\tilde{b} = b + \sup_{\{t \leq M\}} \left\{\tilde{\phi}\,(t) - \phi\,(t)\right\}$. $\square$

4.5.1 Proof of Proposition 4.11

For notational simplicity, let $P = P_{\text{cn}}$. By definition of P, $V(X_1) \leq V(X_0) \vee V(\rho X_0 + \sqrt{1-\rho^2} Z_1)$. Since $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$, we get

$$\sup_{x \in \mathrm{B}(0,1)} PV(x) \leq \sup_{x \in \mathrm{B}(0,1)} \int_{\mathsf{H}} \exp\left(2s\left(\|x\|^2 + (1-\rho^2)\|z\|^2\right)\right) \mathrm{d}\gamma(z) \ , \tag{4.41}$$

and Theorem 4.10 implies that the RHS is finite.
Now, let $x \notin \mathrm{B}(0,1)$ and set $w(x) = (1-\rho)\|x\|/2$. Define the events $\mathscr{I} = \{\|Z_1\| \leq w(X_0)/\sqrt{1-\rho^2}\}$, $\mathscr{A} = \{\alpha(X_0, \rho X_0 + \sqrt{1-\rho^2} Z_1) \geq U\}$, and $\mathscr{R} = \{\alpha(X_0, \rho X_0 + \sqrt{1-\rho^2} Z_1) < U\}$, where $U \sim \mathcal{U}([0,1])$, $Z_1 \sim \gamma$, and U and Z_1 are independent. With these definitions, we get,

$$PV(x) = \mathbb{E}_x\left[V(X_1)\mathbb{1}_{\mathscr{I}^c}\right] + \mathbb{E}_x\left[V(X_1)\mathbb{1}_{\mathscr{I}}(\mathbb{1}_{\mathscr{A}} + \mathbb{1}_{\mathscr{R}})\right] \ . \tag{4.42}$$

For the first term in the RHS, using again $V(X_1) \leq V(X_0) \vee V(\rho X_0 + \sqrt{1-\rho^2} Z_1)$ and $\|x+y\|^2 \leq 2\|x\|^2 + 2\|y\|^2$, we get

$$\begin{aligned} \mathbb{E}_x\left[V(X_1)\mathbb{1}_{\mathscr{I}^c}\right] &\leq \exp\left(2s\|x\|^2\right) \int_{\sqrt{1-\rho^2}\|z\| \geq w(x)} \exp\left(2s(1-\rho^2)\|z\|^2\right) \mathrm{d}\gamma(z) \\ &\leq \exp\left(2s\|x\|^2 - (\theta/2) w(x)^2\right) \int_{\mathsf{H}} \exp\left((\theta/2 + 2s)(1-\rho^2)\|z\|^2\right) \mathrm{d}\gamma(z) \\ &\leq \int_{\mathsf{H}} \exp((5/8)(1-\rho^2)\theta\|z\|^2) \mathrm{d}\gamma(z) \ , \end{aligned}$$

where the definition of s and w are used for the last inequality. Hence by Theorem 4.10, there exists a constant $b < \infty$ such that

$$\sup_{x \in \mathsf{H}} \mathbb{E}_x \left[V(X_1) \mathbb{1}_{\mathscr{I}^c} \right] \le b \,. \tag{4.43}$$

Consider the second term in the RHS of (4.42). On the event $\mathscr{A} \cap \mathscr{I}$, the move is accepted and $\|X_1 - \rho X_0\| \le w(X_0)$. On $\mathscr{R}$, the move is rejected and $X_1 = X_0$. Hence,

$$\mathbb{E}_x \left[V(X_1) \mathbb{1}_{\mathscr{I}} (\mathbb{1}_{\mathscr{A}} + \mathbb{1}_{\mathscr{R}}) \right] \le \left\{ \sup_{z \in \mathrm{B}(\rho x, w(x))} V(z) \right\} \mathbb{P}_x \left[\mathscr{I} \cap \mathscr{A} \right] + V(x) \mathbb{P}_x \left[\mathscr{I} \cap \mathscr{R} \right] \,.$$

For $z \in \mathrm{B}(\rho x, w(x))$, by the triangle inequality, $V(z) \le \exp(s(1+\rho)^2 \left\|x\right\|^2 /4)$. Therefore for any $x \notin \mathrm{B}(0,1)$ since $\rho \in [0,1)$, $\sup_{z \in \mathrm{B}(\rho x, w(x))} V(z) \le \zeta V(x)$, with $\zeta = \exp\{((1+\rho)^2/4 - 1)s\} < 1$. This yields

$$\begin{aligned} \mathbb{E}_x \left[V(X_1) \mathbb{1}_{\mathscr{I}} (\mathbb{1}_{\mathscr{A}} + \mathbb{1}_{\mathscr{R}}) \right] &\le \zeta V(x) \mathbb{P}_x \left[\mathscr{I} \cap \mathscr{A} \right] + V(x) \mathbb{P}_x \left[\mathscr{I} \cap \mathscr{R} \right] \\ &\le V(x) \mathbb{P}_x \left[\mathscr{I} \right] - (1-\zeta) V(x) \mathbb{P}_x \left[\mathscr{A} \cap \mathscr{I} \right] \,. \end{aligned}$$

Since U_1 and Z_1 are independent, we get

$$\mathbb{P}_x \left[\mathscr{A} \cap \mathscr{I} \right] = \mathbb{E}_x \left[\left(1 \wedge \mathrm{e}^{g(x) - g(\rho x + \sqrt{1-\rho^2} Z_1)} \right) \mathbb{1}_{\mathscr{I}} \right] \,.$$

By definition of the set $\mathscr{I}$ and using the inequality $\inf_{z \in \overline{B}(\rho x, w(x))} \exp(g(x) - g(z)) \ge \exp(-C_g (1-\rho)^\beta (3/2)^\beta \left\|x\right\|^\beta)$, we get $\mathbb{P}_x \left[\mathscr{A} \cap \mathscr{I} \right] \ge \exp(-\{\ln V(x)/\kappa\}^{\beta/2}) \mathbb{P}_x \left[\mathscr{I} \right]$, with $\kappa = \theta C_g^{-2/\beta}/36$. Hence, for any $x \notin \mathrm{B}(0,1)$,

$$\mathbb{E}_x \left[V(X_1) \mathbb{1}_{\mathscr{I}} (\mathbb{1}_{\mathscr{A}} + \mathbb{1}_{\mathscr{R}}) \right] \le V(x) - (1-\zeta) \, V(x) \, \exp(-\kappa^{-\beta/2} \log^{\beta/2} V(x)) \,. \tag{4.44}$$

Combining (4.41), (4.43) and (4.44) in (4.42), it follows that there exists $\tilde{b} > 0$ such that, for every $x \in \mathsf{H}$,

$$PV(x) \le V(x) - (1-\zeta) \, V(x) \, \exp(-\kappa^{-\beta/2} \log^{\beta/2} V(x)) + \tilde{b} \,.$$

The proof follows from Lemma 4.21.

4.5.2 Proof of Proposition 4.12

We preface the proof of Proposition 4.12 by a Lemma.

Lemma 4.22. *Assume* ***CN1****. There exists $\eta \in (0,1)$ satisfying the following assertions*

(i) For all $L > 0$, there exists $k(Q_{\mathrm{cn}}, L, \eta) < 1$ such that, for all $x, y \in \mathrm{B}(0, L)$ satisfying $d_\eta(x,y) < 1$, $Q_{\mathrm{cn}} d_\eta(x,y) \le k(Q_{\mathrm{cn}}, L, \eta) d_\eta(x,y)$.

(ii) For all $x, y \in \mathsf{H}$, $Q_{\mathrm{cn}} d_\eta(x,y) \le d_\eta(x,y)$.

Proof. Let $\eta \in (0,1)$; for ease of notation, we simply write Q for Q_{cn}. Let $L > 0$ and choose $x, y \in \mathrm{B}(0, L)$ satisfying $d_\eta(x,y) < 1$. Let (X_1, Y_1) be the basic coupling between $P(x,\cdot)$ and $P(y,\cdot)$; let Z_1, U_1 be the Gaussian variable and the uniform variable used for the basic coupling. Set $\mathscr{I} = \left\{\sqrt{1-\rho^2}\,\|Z_1\| \leq 1\right\}$, $\mathscr{A} = \{\Psi_\wedge(X_0, Y_0, Z_1) > U_1\}$, $\mathscr{R} = \{\Psi_\vee(X_0, Y_0, Z_1) < U_1\}$, where

$$\Psi_\wedge(x,y,z) = \alpha(x, \rho x + \sqrt{1-\rho^2} z) \wedge \alpha(y, \rho y + \sqrt{1-\rho^2} z) \tag{4.45}$$

$$\Psi_\vee(x,y,z) = \alpha(x, \rho x + \sqrt{1-\rho^2} z) \vee \alpha(y, \rho y + \sqrt{1-\rho^2} z) \; . \tag{4.46}$$

On the event $\mathscr{A}$, the moves are both accepted so that $X_1 = \rho X_0 + \sqrt{1-\rho^2} Z_1$ and $Y_1 = \rho X_0 + \sqrt{1-\rho^2} Z_1$; On the event $\mathscr{R}$, the moves are both rejected so that $X_1 = X_0$ and $Y_1 = Y_0$. It holds,

$$Q d_\eta(x,y) \leq \widetilde{\mathbb{E}}_{x,y}\left[d_\eta(X_1,Y_1)\right] \leq \widetilde{\mathbb{E}}_{x,y}\left[d_\eta(X_1,Y_1)\mathbb{1}_{\mathscr{A}\cup\mathscr{R}}\right] + \widetilde{\mathbb{P}}_{x,y}\left[(\mathscr{A}\cup\mathscr{R})^c\right] \; , \tag{4.47}$$

where we have used d_η is bounded by 1. Since $d_\eta(X_1, Y_1) = \rho^\beta d_\eta(X_0, Y_0)$, on $\mathscr{A}$, and $d_\eta(X_1, Y_1) = d_\eta(X_0, Y_0)$, on $\mathscr{R}$, we get

$$\widetilde{\mathbb{E}}_{x,y}\left[d_\eta(X_1,Y_1)(\mathbb{1}_{\mathscr{A}\cup\mathscr{R}})\right] \leq \rho^\beta d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\right] + d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{R}\right] \; .$$

Since $\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\right] + \widetilde{\mathbb{P}}_{x,y}\left[\mathscr{R}\right] \leq 1$, we have

$$\begin{aligned}\widetilde{\mathbb{E}}_{x,y}\left[d_\eta(X_1,Y_1)(\mathbb{1}_{\mathscr{A}\cup\mathscr{R}})\right] &\leq d_\eta(x,y) - (1-\rho^\beta)\, d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\right] \\ &\leq d_\eta(x,y) - (1-\rho^\beta)\, d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\cap\mathscr{I}\right] \; .\end{aligned} \tag{4.48}$$

Set $\Delta(x,y,z) = \left|\alpha(x, \rho x + \sqrt{1-\rho^2} z) - \alpha(y, \rho y + \sqrt{1-\rho^2} z)\right|$. Since Z_1 and U_1 are independent, it follows that $\widetilde{\mathbb{P}}_{x,y}\left[(\mathscr{A}\cup\mathscr{R})^c\right] \leq \int_{\mathsf{H}} \Delta(x,y,z)\mathrm{d}\gamma(z)$ Plugging this identity and (4.48) in (4.47) yields

$$Q d_\eta(x,y) \leq d_\eta(x,y) - (1-\rho^\beta) d_\eta(x,y)\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\cap\mathscr{I}\right] + \int_{\mathsf{H}} \Delta(x,y,z)\mathrm{d}\gamma(z) \; . \tag{4.49}$$

Let us now define $h : \mathsf{H} \to \mathbb{R}$ by

$$h(z) = g(z) - g(\rho z) \; . \tag{4.50}$$

We bound from below $\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\cap\mathscr{I}\right]$. Since U_1 is independent of Z_1, it follows that

$$\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\cap\mathscr{I}\right] \geq \widetilde{\mathbb{E}}_{x,y}\left[\Psi_\wedge(X_0,Y_0,Z_1)\mathbb{1}_{\mathscr{I}}\right] \; .$$

By **CN1**, for all z such that $\sqrt{1-\rho^2}\,\|z\| \leq 1$, it holds for $z \in \mathsf{H}$, $g(z) - g(\rho z + \sqrt{1-\rho^2} z) \geq h(z) - C_g$. Then,

$$\Psi_\wedge(x,y,z) \geq 1 \wedge (\mathrm{e}^{-C_g}\mathrm{e}^{h(x)}) \wedge (\mathrm{e}^{-C_g}\mathrm{e}^{h(y)}) \geq \mathrm{e}^{-C_g}\left[1 \wedge \mathrm{e}^{h(x)\wedge h(y)}\right] \; .$$

Therefore,

$$\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{A}\cap\mathscr{I}\right]\geq \mathrm{e}^{-C_g}\left[1\wedge \mathrm{e}^{h(x)\wedge h(y)}\right]\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{I}\right]\ . \tag{4.51}$$

We now upper bound the integral term in (4.49). For $x,y\in\mathsf{H}$, define the partition of H,

$$\begin{aligned}
\mathscr{K}_1(x,y) &= \{z\in\mathsf{H} : \alpha(x,\rho x+\sqrt{1-\rho^2}z)=\alpha(y,\rho y+\sqrt{1-\rho^2}z)=1\}\\
\mathscr{K}_2(x,y) &= \{z\in\mathsf{H} : \alpha(x,\rho x+\sqrt{1-\rho^2}z)=1>\alpha(y,\rho y+\sqrt{1-\rho^2}z)\}\\
\mathscr{K}_3(x,y) &= \{z\in\mathsf{H} : \alpha(y,\rho y+\sqrt{1-\rho^2}z)=1>\alpha(x,\rho x+\sqrt{1-\rho^2}z)\}\\
\mathscr{K}_4(x,y) &= \{z\in\mathsf{H} : \alpha(y,\rho y+\sqrt{1-\rho^2}z)<1 \text{ and } \alpha(x,\rho x+\sqrt{1-\rho^2}z)<1\}\ .
\end{aligned}$$

Since on $\mathscr{K}_1(x,y)$, $\Theta(x,y,z)=0$,

$$\int_{\mathsf{H}}\Theta(x,y,z)\mathrm{d}\gamma(z)=\sum_{j=2}^{4}\int_{\mathscr{K}_j(x,y)}\Theta(x,y,z)\mathrm{d}\gamma(z)\ . \tag{4.52}$$

For any $a,b>0$, we have $|a-b|=(a\vee b)\left[1-((a/b)\wedge(b/a))\right]$. Upon noting that $1-\mathrm{e}^{-t}\leq t$ for any $t\geq 0$, we have

$$\begin{aligned}
&\Theta(x,y,z)\\
&\leq \Psi_{\vee}(x,y,z)\left|g(y)-g(x)-g(\rho y+\sqrt{1-\rho^2}z)+g(\rho x+\sqrt{1-\rho^2}z)\right|\mathbb{1}_{\cup_{i=2}^4\mathscr{K}_i(x,y)}(z)\ .
\end{aligned}$$

By **CN**1, this yields, for $x,y\in\mathsf{H}$ such that $d_\eta(x,y)<1$,

$$\Theta(x,y,z)\leq 2C_g\left\|y-x\right\|^{\beta}\Psi_{\vee}(x,y,z)\leq 2C_g\eta d_\eta(x,y)\Psi_{\vee}(x,y,z)\ . \tag{4.53}$$

On $\mathscr{K}_2(x,y)$, $g(x)>g(\rho x+\sqrt{1-\rho^2}z)$ and, together with the definition (4.50), this implies that $h(x)\geq g(\rho x+\sqrt{1-\rho^2}z)-g(\rho x)$. Therefore, since under **CN** 1, $h(x)\geq -C_g(1-\rho^2)^{\beta/2}\left\|z\right\|^{\beta}$ we get

$$\begin{aligned}
&\int_{\mathscr{K}_2(x,y)}\Theta(x,y,z)\mathrm{d}\gamma(z)\leq 2C_g\eta d_\eta(x,y)\int_{\mathscr{K}_2(x,y)}\mathrm{d}\gamma(z)\\
&\leq 2C_g\eta d_\eta(x,y)\left\{\left[\mathrm{e}^{h(x)}\int_{\mathscr{K}_2(x,y)}\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^{\beta}}\mathrm{d}\gamma(z)\right]\wedge 1\right\}\leq C_I\eta d_\eta(x,y)\left\{\mathrm{e}^{h(x)}\wedge 1\right\}\ ,
\end{aligned} \tag{4.54}$$

for a constant C_I, which is finite according to Theorem 4.10. By symmetry, on $\mathscr{K}_3(x,y)$,

$$\int_{\mathscr{K}_3(x,y)}\Theta(x,y,z)\mathrm{d}\gamma(z)\leq C_I\eta d_\eta(x,y)\left\{\mathrm{e}^{h(y)}\wedge 1\right\}\ . \tag{4.55}$$

On $\mathscr{K}_4(x,y)$, using **CN**1,

$$\alpha(x,\rho x+\sqrt{1-\rho^2}z)=\mathrm{e}^{g(x)-g(\rho x+\sqrt{1-\rho^2}z)}\wedge 1\leq\left(\mathrm{e}^{h(x)}\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^{\beta}}\right)\wedge 1\ ;$$

and by symmetry, we obtain a similar upper bound for $\alpha(y, \rho y + \sqrt{1-\rho^2}z)$. Since $\mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta} \geq 1$, these two inequalities imply $\Psi_\vee(x,y,z) \leq \mathrm{e}^{C_g(1-\rho^2)^{\beta/2}\|z\|^\beta}(\mathrm{e}^{h(x)\vee h(y)} \wedge 1)$. Hence, using again (4.53) and Theorem 4.10, there exists $C_I < +\infty$ such that

$$\int_{\mathscr{K}_4(x,y)} \Theta(x,y,z)\mathrm{d}\gamma(z) \leq C_I \eta d_\eta(x,y)\left[\mathrm{e}^{h(x)\vee h(y)} \wedge 1\right] . \tag{4.56}$$

Plugging (4.54), (4.55), (4.56) into (4.52), we finally obtain

$$\int_{\mathsf{H}} \Theta(x,y,z)\mathrm{d}\gamma(z) \leq 3C_I \eta d_\eta(x,y)\left[\mathrm{e}^{h(x)\vee h(y)} \wedge 1\right] .$$

Finally, under **CN**1, for every $x, y \in \mathsf{H}$ such that $d_\eta(x,y) < 1$, $|h(x) - h(y)| \leq 2C_g \|x-y\|^\beta \leq 2C_g\eta^\beta$. Therefore $\mathrm{e}^{h(x)\vee h(y)} \wedge 1 \leq \mathrm{e}^{2C_g\eta^\beta}\left[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\right]$ and

$$\int_{\mathsf{H}} \Theta(x,y,z)\mathrm{d}\gamma(z) \leq 3C_I \mathrm{e}^{2C_g\eta^\beta}\eta d_\eta(x,y)\left[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\right] . \tag{4.57}$$

Plugging (4.51) and (4.57) in (4.49) yields

$$Qd_\eta(x,y) \leq d_\eta(x,y)\left(1 - \left\{(1-\rho^\beta)\mathrm{e}^{-C_g}\widetilde{\mathbb{P}}_{x,y}\left[\mathscr{I}\right] - 3C_I\mathrm{e}^{2C_g\eta^\beta}\eta\right\}\left[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\right]\right) .$$

Note that $M = \widetilde{\mathbb{P}}_{x,y}\left[\mathscr{I}\right]$ is a positive quantity that does not depend on x, y. Therefore, we may choose η sufficiently small so that, for every $x, y \in \mathsf{H}$ satisfying $d_\eta(x,y) < 1$,

$$Qd_\eta(x,y) \leq d_\eta(x,y)\left(1 - (1/2)(1-\rho^\beta)\mathrm{e}^{-C_g}M\left[\mathrm{e}^{h(x)\wedge h(y)} \wedge 1\right]\right) , \tag{4.58}$$

which implies Lemma 4.22-(i) upon noting that, under the stated assumptions, $\inf_{\mathrm{B}(0,L)} h > -\infty$.

We now consider (ii). For every $x, y \in \mathsf{H}$, $d_\eta(x,y) \leq 1$, which implies that $Qd_\eta(x,y) \leq 1$. For every $x, y \in \mathsf{H}$ such that $d_\eta(x,y) = 1$, $Qd_\eta(x,y) \leq 1 = d_\eta(x,y)$. If $d_\eta(x,y) < 1$, (4.58) shows that $Qd_\eta(x,y) \leq d_\eta(x,y)$. □

Proof of Proposition 4.12. Let $\{(X_n, Y_n),\ n \in \mathbb{N}\}$ be a Markov chain with Markov kernel Q given by (4.13). We denote for all $n \in \mathbb{N}^*$, Z_n and U_n, respectively the common Gaussian variable and uniform variable, used in the definition (X_n, Y_n). Note that by definition the variables $\{Z_n, U_n;\ n \in \mathbb{N}\}$ are independent.

Since $\{x : V(x) \leq u\} = \{x : \|x\| \leq \log(u)\}$, for $u \geq 1$, we only prove that for all $L > 0$, there exist $\ell \in \mathbb{N}^*$ and $\epsilon > 0$ such that $\overline{B}(0,L)^2$ is a (ℓ, ϵ, d_η)-coupling set. By Lemma 4.22-(i), for any $L > 0$, there exists $k(Q,L,\eta) \in (0,1)$ such that for any $x, y \in \overline{B}(0,L)$ satisfying $d_\eta(x,y) < 1$, $Qd_\eta(x,y) \leq k(Q,L,\eta)d_\eta(x,y)$. Then by Lemma 4.22-(ii) , for every $n \in \mathbb{N}^*$,

$$Q^n d_\eta(x,y) \leq Q^{n-1}d_\eta(x,y) \leq \cdots \leq k(Q,L,\eta)d_\eta(x,y) . \tag{4.59}$$

Consider now the case $d_\eta(x,y) = 1$. Let $n \in \mathbb{N}^*$ and denote for all $1 \leq i \leq n$ $\mathscr{A}_i = \{U_i \leq \Psi_\wedge(X_{i-1}, Y_{i-1}, Z_i)\}$ and $\widetilde{\mathscr{A}_i}(n) = \bigcap_{1\leq j\leq i}\left(\{\sqrt{1-\rho^2}\,\|Z_j\| \leq L/n\} \cap \mathscr{A}_j\right)$ where $\Psi_\wedge$ is defined in (4.45)

On the event $\widetilde{\mathscr{A}_i}(n)$, $X_j = \rho X_{j-1} + \sqrt{1-\rho^2} Z_j$ and $Y_j = \rho Y_{j-1} + \sqrt{1-\rho^2} Z_j$ for all $1 \le j \le i$. Then, since $d_\eta(x,y) \le \eta^{-1} \|x-y\|^\beta$, on $\widetilde{\mathscr{A}_n}(n)$ it holds $d_\eta(X_n, Y_n) \le \eta^{-1}\rho^{\beta n} \|X_0 - Y_0\|^\beta$. This inequality and $d_\eta(x,y) \le 1$ yield

$$\begin{aligned} Q^n d_\eta(x,y) &= \widetilde{\mathbb{E}}_{x,y}\left[d_\eta(X_n,Y_n)(\mathbb{1}_{\widetilde{\mathscr{A}_n}(n)} + \mathbb{1}_{(\widetilde{\mathscr{A}_n}(n))^c})\right] \\ &\le \rho^{\beta n} \|x-y\|^\beta \, \widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_n}(n)\right] + \widetilde{\mathbb{P}}_{x,y}\left[(\widetilde{\mathscr{A}_n}(n))^c\right] \\ &\le \rho^{\beta n}(2L)^\beta \widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_n}(n)\right] + \widetilde{\mathbb{P}}_{x,y}\left[(\widetilde{\mathscr{A}_n}(n))^c\right] \le 1 + \left(\rho^{\beta n}(2L)^\beta - 1\right) \widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_n}(n)\right] . \end{aligned} \tag{4.60}$$

As $\rho \in [0,1)$, there exists ℓ such that, $\rho^{\beta\ell}(2L)^\beta < 1$. It remains to lower bound $\widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_\ell}(\ell)\right]$ by a positive constant to conclude. Since the random variables $\{(Z_i, U_i);\ i \in \mathbb{N}^*\}$ are independent, we get

$$\begin{aligned} \widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_\ell}(\ell)\right] = \widetilde{\mathbb{P}}_{x,y}&\left[\widetilde{\mathscr{A}_{\ell-1}}(\ell) \cap \{\sqrt{1-\rho^2}\,\|Z_\ell\| \le L/\ell\}\right] \\ &\times \widetilde{\mathbb{E}}_{x,y}\left[\Psi_\wedge(X_{\ell-1},Y_{\ell-1},Z_\ell)\,\middle|\,\widetilde{\mathscr{A}}_{\ell-1}(\ell) \cap \{\sqrt{1-\rho^2}\,\|Z_\ell\| \le L/\ell\}\right] . \end{aligned}$$

For all $1 \le i \le \ell$, on the event $\bigcap_{j\le i}\left\{\sqrt{1-\rho^2}\,\|Z_j\| \le L/\ell\right\}$, it holds

$$\Psi_\wedge(X_{i-1},Y_{i-1},Z_i) \ge \exp\left(-\sup_{z\in \mathrm{B}(0,2L)} g(z) + \inf_{z\in\mathrm{B}(0,2L)} g(z)\right) = \delta ,$$

where $\delta \in (0,1)$. Therefore, since Z_ℓ is independent of $\widetilde{\mathscr{A}_{\ell-1}}(\ell)$, we have

$$\widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_\ell}(\ell)\right] \ge \delta \widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_{\ell-1}}(\ell)\right]\ \widetilde{\mathbb{P}}_{x,y}\left[\sqrt{1-\rho^2}\,\|Z_\ell\| \le L/\ell\right] .$$

An immediate induction leads to $\widetilde{\mathbb{P}}_{x,y}\left[\widetilde{\mathscr{A}_\ell}(\ell)\right] \ge \left(\widetilde{\mathbb{P}}_{x,y}\left[\sqrt{1-\rho^2}\,\|Z_1\| \le L/\ell\right]\right)^\ell \delta^\ell$. Plugging this result in (4.60) and (4.59) implies there exists $\zeta \in (0,1)$ such that for all $x,y \in \overline{B}(0,L)$, $Q^\ell d_\eta(x,y) \le \zeta d_\eta(x,y)$. ☐

4.6 Wasserstein distance: some useful properties

Let (E,d) be a Polish space, with d bounded be 1. Then, for all $\mu,\nu \in \mathcal{P}(E)$: $W_d(\mu,\nu) \le W_{d_0}(\mu,\nu)$ since for all $x,y \in E$, $d(x,y) \le d_0(x,y)$. Hence when d is bounded by 1, the convergence in total variation distance implies the convergence in the Wasserstein metric W_d.

Lemma 4.23. *Let (E,d) be a Polish space, with d bounded by 1, and let P be a Markov kernel on $(E,\mathcal{B}(E))$. Let Q be a coupling kernel for P.*

(i) Then, for all probability measures $\mu, \nu \in \mathcal{P}(E)$ and $n \in \mathbb{N}^*$,

$$W_d(\mu P^n, \nu P^n) \leq \inf_{\lambda \in \Pi(\mu,\nu)} \int_{E \times E} Q^n d(z,t) \mathrm{d}\lambda(z,t) \ .$$

(ii) If in addition Q is a d-weak-contraction, then for all probability measures $\mu, \nu \in \mathcal{P}(E)$,

$$W_d(\mu P, \nu P) \leq W_d(\mu, \nu) \ .$$

In particular, for all $x, y \in E$, $W_d(P(x,\cdot), P(y,\cdot)) \leq d(x,y)$.

Proof. (i) For every $\lambda \in \Pi(\mu,\nu)$, λQ^n is a transference plan of μP^n and νP^n. This yields the result. Consider now (ii). Using (i), we get

$$\begin{aligned} W_d(\mu P, \nu P) &\leq \inf_{\lambda \in \Pi(\mu,\nu)} \int_{E \times E} Q d(z,t) \mathrm{d}\lambda(z,t) \\ &\leq \inf_{\lambda \in \Pi(\mu,\nu)} \int_{E \times E} d(z,t) \mathrm{d}\lambda(z,t) \leq W_d(\mu,\nu) \ . \end{aligned}$$

□

4.7 Subgeometric functions and sequences

Lemma 4.24. *Let $r \in \Lambda_0$ and R be given by (4.23).*

(i) For all $t, v \in \mathbb{R}_+$, $r(t+v) \leq r(t) r(v)$.

(ii) R is differentiable, convex and increasing to $+\infty$.

(iii) $\lim_{t\to\infty} r(t)/R(t) = 0$.

(iv) There exists a constant C such that for any $t, v \in \mathbb{R}_+$, $R(t+v) \leq C R(t) R(v)$.

(v) $\sup_k R(k) / \sum_{i=0}^{k-1} r(i) < \infty$.

Proof. (i) follows from [SW67, Lemma 1]. Consider now (ii). By definition, r is non-decreasing, thus is bounded on every compact set; then, R is continuous. Moreover, it is differentiable and its derivative is r, which is non-decreasing. Then R is convex. In addition $r(0) \geq 2$, thus R is increasing to $+\infty$. (iii). Set $u(t) \stackrel{\text{def}}{=} \log(r(t))/t$. Since $r \in \Lambda_0$, the function u is non increasing, which implies that, for every $h \in (0,1)$,

$$\begin{aligned} \log\left(1 + \{r(t+h) - r(t)\}/r(t)\right) &= \log\left(r(t+h)/r(t)\right) \\ &= t(u(t+h) - u(t)) + h u(t+h) \leq h u(t+h) \ . \end{aligned}$$

Since $\lim_{t\to+\infty} u(t) = 0$, for all $\epsilon > 0$, there exists $T \in \mathbb{R}_+$ such that for all $t \geq T$ and $h \in (0,1)$, $(r(t+h) - r(t)) \leq \epsilon h r(t)$. Therefore for all $t \geq T$ and $h \in (0,1)$, $(R(t+h) - R(t))/(hR(t)) \leq \epsilon + r(T+1)/R(t)$. Taking $h \to 0$ it follows $r(t)/R(t) \leq \epsilon + r(T+1)/R(t)$, for all $t \geq T$. The proof is concluded by (ii). (iv) follows from (i) and (iii). Finally, for (v), the upper bound follows from (iv) and $R(k-1) \leq 1 + \sum_{i=0}^{k-1} r(i)$. □

Part II

Study of ULA

Chapter 5

Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm

Alain Durmus[1], Éric Moulines [2]

Abstract

In this chapter, we study a method to sample from a target distribution π over $\mathbb{R}^d$ having a positive density with respect to the Lebesgue measure, known up to a normalisation factor. This method is based on the Euler discretization of the overdamped Langevin stochastic differential equation associated with π. For both constant and decreasing step sizes in the Euler discretization, we obtain non-asymptotic bounds for the convergence to the target distribution π in total variation distance. A particular attention is paid to the dependency on the dimension d, to demonstrate the applicability of this method in the high dimensional setting. These bounds improve and extend the results of [Dal16].

5.1 Introduction

Sampling distributions over high-dimensional state-spaces is a problem which has recently attracted a lot of research efforts in computational statistics and machine learning (see [Cot+13] and [And+03] for details); applications include Bayesian non-parametrics, Bayesian inverse problems and aggregation of estimators. All these problems boil down to sample a target distribution π having a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalisation factor $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$ where U is continuously

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

differentiable. We consider a sampling method based on the Euler discretization of the overdamped Langevin stochastic differential equation (SDE)

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t^d \,, \tag{5.1}$$

where $(B_t^d)_{t\geq 0}$ is a d-dimensional Brownian motion. It is well-known that the Markov semi-group associated with the Langevin diffusion $(Y_t)_{t\geq 0}$ is reversible w.r.t. π. Under suitable conditions, the convergence to π takes place at geometric rate. Precise quantitative estimates of the rate of convergence with explicit dependency on the dimension d of the state space have been recently obtained using either functional inequalities such as Poincaré and log-Sobolev inequalities (see [BCG08; CG09] [BGL14]) or by coupling techniques (see [Ebe15]). The Euler-Maruyama discretization scheme associated to the Langevin diffusion yields the discrete time-Markov chain given by

$$X_{k+1} = X_k - \gamma_{k+1}\nabla U(X_k) + \sqrt{2\gamma_{k+1}}Z_{k+1} \tag{5.2}$$

where $(Z_k)_{k\geq 1}$ is an i.i.d. sequence of standard Gaussian d-dimensional random vectors and $(\gamma_k)_{k\geq 1}$ is a sequence of step sizes, which can either be held constant or be chosen to decrease to 0. The idea of using the Markov chain $(X_k)_{k\geq 0}$ to sample approximately from the target π has been first introduced in the physics literature by [Par81] and popularised in the computational statistics community by [Gre83] and [GM94]. It has been studied in depth by [RT96a], which proposed to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π leading to the Metropolis Adjusted Langevin Algorithm (MALA). They coin the term *unadjusted* Langevin algorithm (ULA) when the Metropolis-Hastings step is skipped.

The purpose of this paper is to study the convergence of the ULA algorithm. The emphasis is put on non-asymptotic computable bounds; we pay a particular attention to the way these bounds scale with the dimension d and constants characterizing the smoothness and curvature of the potential U. Our study covers both constant and decreasing step sizes and we analyse both the "finite horizon" (where the total number of simulations is specified before running the algorithm) and "any-time" settings (where the algorithm can be stopped after any iteration).

When the step size $\gamma_k = \gamma$ is constant, under appropriate conditions (see [RT96a]), the Markov chain $(X_n)_{n\geq 0}$ is V-uniformly geometrically ergodic with a stationary distribution π_γ. With few exceptions, the stationary distribution π_γ is different from the target π. If the step size γ is small enough, then the stationary distribution of this chain is in some sense close to π. We provide non-asymptotic bounds of the V-total variation distance between π_γ and π, with explicit dependence on the step size γ and the dimension d. Our results complete and extend the recent works by [DT12] and [Dal16].

When $(\gamma_k)_{k\geq 1}$ decreases to zero, then $(X_k)_{k\geq 0}$ is a non-homogeneous Markov chain. If in addition $\sum_{k=1}^{\infty}\gamma_k = \infty$, we show that the marginal distribution of this non-homogeneous chain converges, under some mild additional conditions, to the target distribution π, and provide explicit bounds for the convergence. Compared to the related works by [LP02], [LP03], [Lem05] and [LM10], we establish not only the weak convergence of the weighted empirical measure of the path to the target distribution but a much stronger convergence in total variation, similarly to [Dal16], where the strongly log-concave case is considered.

The paper is organized as follows. In Section 5.2, the main convergence results are stated under abstract assumptions. We then specialize in Section 5.3 these results to different classes of densities. The proofs are gathered in Section 6.8. Some general convergence results for diffusions based on reflection coupling, which are of independent interest, are stated in Section 5.5.

Notations and conventions

$\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-field of $\mathbb{R}^d$ and $\mathbb{F}(\mathbb{R}^d)$ the set of all Borel measurable functions on $\mathbb{R}^d$. For $f \in \mathbb{F}(\mathbb{R}^d)$ set $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. Denote by $\mathbb{M}(\mathbb{R}^d)$ the space of finite signed measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\mathbb{M}_0(\mathbb{R}^d) = \{\mu \in \mathbb{M}(\mathbb{R}^d) \mid \mu(\mathbb{R}^d) = 0\}$. For $\mu \in \mathbb{M}(\mathbb{R}^d)$ and $f \in \mathbb{F}(\mathbb{R}^d)$ a μ-integrable function, denote by $\mu(f)$ the integral of f w.r.t. μ. Let $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. For $f \in \mathbb{F}(\mathbb{R}^d)$, the V-norm of f is given by $\|f\|_V = \sup_{x\in\mathbb{R}^d} |f(x)|/V(x)$. For $\mu \in \mathbb{M}(\mathbb{R}^d)$, the V-total variation distance of μ is defined as

$$\|\mu\|_V = \sup_{f\in\mathbb{F}(\mathbb{R}^d), \|f\|_V\leq 1} \left| \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x) \right| .$$

If $V \equiv 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{\mathrm{TV}}$.

For $p \geq 1$, denote by $\mathrm{L}^p(\pi)$ the set of measurable functions such that $\pi(|f|^p) < \infty$. For $f \in \mathrm{L}^2(\pi)$, the variance of f under π is denoted by $\mathrm{Var}_\pi\{f\}$. For all functions f such that $f\log(f) \in \mathrm{L}^1(\pi)$, the entropy of f with respect to π is defined by

$$\mathrm{Ent}_\pi(f) = \int_{\mathbb{R}^d} f(x)\log(f(x))\mathrm{d}\pi(x) .$$

Let μ and ν be two probability measures on $\mathbb{R}^d$. If $\mu \ll \nu$, we denote by $\mathrm{d}\mu/\mathrm{d}\nu$ the Radon-Nikodym derivative of μ w.r.t. ν. Denote for all $x, y \in \mathbb{R}^d$ by $\langle x, y\rangle$ the scalar product of x and y and $\|x\|$ the Euclidean norm of x. For $k \geq 0$, denote by $C^k(\mathbb{R}^d)$, the set of k-times continuously differentiable functions $f : \mathbb{R}^d \to \mathbb{R}$. For $f \in C^2(\mathbb{R}^d)$, denote by ∇f the gradient of f and Δf the Laplacian of f. For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $\mathrm{B}(x, M)$, the ball centered at x of radius M. Denote for $K \geq 0$, the oscillation of a function $f \in C^0(\mathbb{R}^d)$ in the ball $\mathrm{B}(0, K)$ by $\mathrm{osc}_K(f) = \sup_{\mathrm{B}(0,K)}(f) - \inf_{\mathrm{B}(0,K)}(f)$. Denote the oscillation of a bounded function $f \in C^0(\mathbb{R}^d)$ on $\mathbb{R}^d$ by $\mathrm{osc}_{\mathbb{R}^d}(f) = \sup_{\mathbb{R}^d}(f) - \inf_{\mathbb{R}^d}(f)$. In the sequel, we take the convention that $\sum_p^n = 0$ and $\prod_p^n = 1$, for $n, p \in \mathbb{N}$, $n < p$.

5.2 General conditions for the convergence of ULA

In this section, we derive a bound on the convergence of the ULA to the target distribution π when the Langevin diffusion is geometrically ergodic and the Markov kernel associated with the EM discretization satisfies a Foster-Lyapunov drift inequality.

Consider the following assumption on the potential U:

L1. *The function U is continuously differentiable on $\mathbb{R}^d$ and gradient Lipschitz, i.e. there exists $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U(x) - \nabla U(y)\| \leq L \|x - y\| .$$

Under **L**1, by [IW89, Theorem 2.4-3.1] for every initial point $x \in \mathbb{R}^d$, there exists a unique strong solution $(Y_t(x))_{t\geq 0}$ to the Langevin SDE (5.1). Define for all $t \geq 0$, $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$, $P_t(x, \mathrm{A}) = \mathbb{P}(Y_t(x) \in \mathrm{A})$. The semigroup $(P_t)_{t\geq 0}$ is reversible w.r.t. π, and hence admits π as its (unique) invariant distribution. In this section, we consider the case where $(P_t)_{t\geq 0}$ is geometrically ergodic, i.e. there exists $\kappa \in [0, 1)$ such that for any initial distribution μ_0 and $t > 0$,

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq C(\mu_0)\kappa^t , \tag{5.3}$$

for some constant $C(\mu_0) \in [0, +\infty]$. Denote by $\mathscr{A}^{\mathrm{L}}$ the generator associated with the semigroup $(P_t)_{t\geq 0}$, given for all $f \in C^2(\mathbb{R}^d)$ by

$$\mathscr{A}^{\mathrm{L}} f = -\langle \nabla U, \nabla f \rangle + \Delta f .$$

A twice continuously differentiable function $V : \mathbb{R}^d \to [1, \infty)$ is a *Lyapunov function* for the generator $\mathscr{A}^{\mathrm{L}}$ if there exist $\theta > 0$, $\beta \geq 0$ and $\mathcal{E} \subset \mathcal{B}$ such that,

$$\mathscr{A}^{\mathrm{L}} V \leq -\theta V + \beta \mathbb{1}_{\mathcal{E}} . \tag{5.4}$$

By [RT96a, Theorem 2.2], if $\mathcal{E}$ in (5.4) is a non-empty compact set, then the Langevin diffusion is geometrically ergodic.

Consider now the EM discretization of the diffusion (5.2). Let $(\gamma_k)_{k\geq 1}$ be a sequence of positive and nonincreasing step sizes and for $0 \leq n \leq p$, denote by

$$\Gamma_{n,p} = \sum_{k=n}^{p} \gamma_k , \qquad \Gamma_n = \Gamma_{1,n} . \tag{5.5}$$

For $\gamma > 0$, consider the Markov kernel R_γ given for all $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$R_\gamma(x, \mathrm{A}) = \int_{\mathrm{A}} (4\pi\gamma)^{-d/2} \exp\left(-(4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2\right) \mathrm{d}y .$$

The discretized Langevin diffusion $(X_n)_{n\geq 0}$ given in (5.2) is a time-inhomogeneous Markov chain, for $p \geq n \geq 1$ and $f \in \mathbb{F}_+(\mathbb{R}^d)$, $\mathbb{E}^{\mathcal{F}_n}[f(X_p)] = Q_\gamma^{n,p} f(X_n)$ where $\mathcal{F}_n = \sigma(X_\ell, 0 \leq \ell \leq n)$ and

$$Q_\gamma^{n,p} = R_{\gamma_n} \cdots R_{\gamma_p} , \qquad Q_\gamma^n = Q_\gamma^{1,n} ,$$

with the convention that for $n, p \geq 0$, $n < p$, $Q_\gamma^{p,n}$ is the identity operator. Under **L** 1, the Markov kernel R_γ is strongly Feller, irreducible and strongly aperiodic. We will say that a function $V : \mathbb{R}^d \to [1, \infty)$ satisfies a Foster-Lyapunov drift condition for R_γ if there exist constants $\bar{\gamma} > 0$, $\lambda \in [0, 1)$ and $c > 0$ such that, for all $\gamma \in (0, \bar{\gamma}]$

$$R_\gamma V \leq \lambda^\gamma V + \gamma c . \tag{5.6}$$

The particular form of (5.6) reflects how the mixing rate of the Markov chain depends upon the step size $\gamma > 0$. If $\gamma = 0$, then $R_0(x, \mathrm{A}) = \delta_x(\mathrm{A})$ for $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$. A Markov chain with transition kernel R_0 is not mixing. Intuitively, as γ gets larger, then it is expected that the mixing of R_γ increases. If for some $\gamma > 0$, R_γ satisfies (5.6), then R_γ admits a unique stationary distribution π_γ. We use (5.6) to control quantitatively the moments of the time-inhomogeneous chain. The types of bounds which are needed, are summarised in the following elementary Lemma.

Lemma 5.1. *Let $\bar{\gamma} > 0$. Assume that for all $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$, (5.6) holds for some constants $\lambda \in (0,1)$ and $c > 0$. Let $(\gamma_k)_{k \geq 1}$ be a sequence of nonincreasing step sizes such that $\gamma_k \in (0, \bar{\gamma}]$ for all $k \in \mathbb{N}^*$. Then for all $n \geq 0$ and $x \in \mathbb{R}^d$, $Q_\gamma^n V(x) \leq F(\lambda, \Gamma_n, c, \gamma_1, V(x))$ where*

$$F(\lambda, a, c, \gamma, w) = \lambda^a w + c(-\lambda^\gamma \log(\lambda))^{-1} \,. \tag{5.7}$$

Proof. The proof is postponed to Section 5.4.1. □

Note that Lemma 5.1 implies that $\sup_{k \geq 0}\{Q_\gamma^k V(x)\} \leq G(\lambda, c, \gamma_1, V(x))$ where

$$G(\lambda, c, \gamma, w) = w + c(-\lambda^\gamma \log(\lambda))^{-1} \,. \tag{5.8}$$

We give below the main ingredients which are needed to obtain a quantitative bound for $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ for all $x \in \mathbb{R}^d$. This quantity is decomposed as follows: for all $0 \leq n < p$,

$$\begin{aligned} &\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \\ &\qquad \leq \|\delta_x Q_\gamma^n Q_\gamma^{n+1,p} - \delta_x Q_\gamma^n P_{\Gamma_{n+1,p}}\|_{\mathrm{TV}} + \|\delta_x Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\mathrm{TV}} \,. \end{aligned} \tag{5.9}$$

To control the first term on the right hand side, we use a method introduced in [DT12] and elaborated in [Dal16]. The second term is bounded using the convergence of the semi-group to π, see (5.3).

Proposition 5.2. *Assume that **L** 1 and (5.3) hold. Let $(\gamma_k)_{k \geq 0}$ be a sequence of non-negative step sizes. Then for all $x \in \mathbb{R}^d$, $n \geq 0$, $p \geq 1$, $n < p$,*

$$\begin{aligned} &\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \\ &\qquad \leq 2^{-1/2} L \left(\sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) A(\gamma, x) + d\gamma_{k+1}^2 \right\} \right)^{1/2} + C(\delta_x Q_\gamma^n) \kappa^{\Gamma_{n+1,p}} \,, \end{aligned} \tag{5.10}$$

where $\kappa, C(\delta_x Q_\gamma^n)$ are defined in (5.3) and

$$A(\gamma, x) = \sup_{k \geq 0} \int_{\mathbb{R}^d} \|\nabla U(z)\|^2 \, Q_\gamma^k(y, \mathrm{d}z) \,. \tag{5.11}$$

Proof. The proof follows the same lines as [Dal16, Lemma 2] but is given for completeness. For $0 \leq s \leq t$, let $\mathrm{C}([s,t], \mathbb{R}^d)$ be the space of continuous functions on $[s,t]$ taking values in $\mathbb{R}^d$. For all $y \in \mathbb{R}^d$, denote by $\mu^y_{n,p}$ and $\bar{\mu}^y_{n,p}$ the laws on $\mathrm{C}([\Gamma_n, \Gamma_p], \mathbb{R}^d)$ of the Langevin diffusion $(Y_t(y))_{\Gamma_n \leq t \leq \Gamma_p}$ and of the continuously-interpolated Euler discretization $(\bar{Y}_t(y))_{\Gamma_n \leq t \leq \Gamma_p}$, both started at y at time Γ_n. Denote by $(Y_t(y), \overline{Y}_t(y))_{t \geq \Gamma_n}$ the unique strong solution started at (y,y) at time $t = \Gamma_n$ of the time-inhomogeneous diffusion defined for $t \geq \Gamma_n$, by

$$\begin{cases} \mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B^d_t \\ \mathrm{d}\bar{Y}_t = -\overline{\nabla U}(\bar{Y}, t)\mathrm{d}t + \sqrt{2}\mathrm{d}B^d_t \, , \end{cases} \tag{5.12}$$

where for any continuous function $\mathrm{w} : \mathbb{R}_+ \to \mathbb{R}^d$ and $t \geq \Gamma_n$

$$\overline{\nabla U}(\mathrm{w}, t) = \sum_{k=n}^{\infty} \nabla U(\mathrm{w}_{\Gamma_k}) \mathbb{1}_{[\Gamma_k, \Gamma_{k+1})}(t) \, . \tag{5.13}$$

Girsanov's Theorem [KS91, Theorem 5.1, Corollary 5.16, Chapter 3] shows that $\mu^y_{n,p}$ and $\bar{\mu}^y_{n,p}$ are mutually absolutely continuous and in addition, $\bar{\mu}^y_{n,p}$-almost surely

$$\begin{aligned} \frac{\mathrm{d}\mu^y_{n,p}}{\mathrm{d}\bar{\mu}^y_{n,p}} = \exp\Bigg(\frac{1}{2} \int_{\Gamma_n}^{\Gamma_p} \left\langle \nabla U(\bar{Y}_s(y)) - \overline{\nabla U}(\bar{Y}(y), s), \mathrm{d}\bar{Y}_s(y) \right\rangle \\ - \frac{1}{4} \int_{\Gamma_n}^{\Gamma_p} \left\{ \left\| \nabla U(\bar{Y}_s(y)) \right\|^2 - \left\| \overline{\nabla U}(\bar{Y}(y), s) \right\|^2 \right\} \mathrm{d}s \Bigg) \, . \end{aligned} \tag{5.14}$$

Under **L**1, (5.14) implies for all $y \in \mathbb{R}^d$:

$$\begin{aligned} \mathrm{KL}(\mu^y_{n,p} | \bar{\mu}^y_{n,p}) &\leq 4^{-1} \int_{\Gamma_n}^{\Gamma_p} \mathbb{E}\left[\left\| \nabla U(\bar{Y}_s(y)) - \overline{\nabla U}(\bar{Y}(y), s) \right\|^2 \right] \mathrm{d}s \\ &\leq 4^{-1} \sum_{k=n}^{p-1} \int_{\Gamma_k}^{\Gamma_{k+1}} \mathbb{E}\left[\left\| \nabla U(\bar{Y}_s(y)) - \nabla U(\bar{Y}_{\Gamma_k}(y)) \right\|^2 \right] \mathrm{d}s \\ &\leq 4^{-1} L^2 \sum_{k=n}^{p-1} \left\{ (\gamma^3_{k+1}/3) \int_{\mathbb{R}^d} \left\| \nabla U(z) \right\|^2 Q^{n+1,k}_\gamma(y, \mathrm{d}z) + d\gamma^2_{k+1} \right\} \, . \end{aligned} \tag{5.15}$$

By the Pinsker inequality, $\|\delta_y Q^{n+1,p}_\gamma - \delta_y P_{\Gamma_{n+1,p}}\|_{\mathrm{TV}} \leq \sqrt{2}\{\mathrm{KL}(\mu^y_{n,p} | \bar{\mu}^y_{n,p})\}^{1/2}$. The proof is concluded by combining this inequality, (5.15) and (5.3) in (5.9). $\square$

In the sequel, depending on the conditions on the potential U and the techniques of proof, for any given $x \in \mathbb{R}^d$, $C(\delta_x Q^n_\gamma)$ can have two kinds of upper bounds, either of the form $-\log(\gamma_n) W(x)$, or $\exp(a\Gamma_n) W(x)$, for some function $W : \mathbb{R}^d \to \mathbb{R}$ and $a > 0$. In both cases, as shown in Proposition 5.3, it is possible to choose n as a function of p, so that $\lim_{p \to +\infty} \|\delta_x Q^p_\gamma - \pi\|_{\mathrm{TV}} = 0$ under appropriate conditions on the sequence of step sizes $(\gamma_k)_{k \geq 1}$.

Proposition 5.3. *Assume that* **L** *1 and* (5.3) *hold. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence satisfying* $\lim_{k\to+\infty}\Gamma_k = +\infty$ *and* $\lim_{k\to\infty}\gamma_k = 0$*. Then,* $\lim_{n\to\infty}\|\delta_x Q_\gamma^n - \pi\|_{\mathrm{TV}} = 0$ *for any* $x \in \mathbb{R}^d$ *for which one of the two following conditions holds:*

(i) $A(\gamma, x) < \infty$ *and* $\limsup_{n\to+\infty} C(\delta_x Q_\gamma^n)/(-\log(\gamma_n)) < +\infty$*, where* $A(\gamma, x)$ *is defined in* (5.11)*.*

(ii) $\sum_{k=1}^{\infty}\gamma_k^2 < +\infty$*,* $A(\gamma, x) < \infty$ *and* $\limsup_{n\to+\infty}\log\{C(\delta_x Q_\gamma^n)\}/\Gamma_n < +\infty$*.*

Proof. (i) There exists $p_0 \geq 1$ such that for all $p \geq p_0$, $\kappa^{\gamma_p} > \gamma_p$ and $\kappa^{\Gamma_p} \leq \gamma_1$. Therefore, we can define for all $p \geq p_0$,

$$n(p) \stackrel{\text{def}}{=} \min\left\{k \in \{0, \cdots, p-1\} \,|\, \kappa^{\Gamma_{k+1,p}} > \gamma_{k+1}\right\} . \tag{5.16}$$

and $n(p) \geq 1$. We first show that $\liminf_{p\to\infty} n(p) = \infty$. The proof goes by contradiction. If $\liminf_{p\to\infty} n(p) < \infty$ we could extract a bounded subsequence $(n(p_k))_{k\geq 1}$. For such sequence, $(\gamma_{n(p_k)+1})_{k\geq 1}$ is bounded away from 0, but $\lim_{k\to+\infty}\kappa^{\Gamma_{n(p_k)+1,p_k}} = 0$ which yields to a contradiction. The definition of $n(p)$ implies that $\kappa^{\Gamma_{n(p),p}} \leq \gamma_{n(p)}$, showing that

$$\limsup_{p\to+\infty} C(\delta_x Q_\gamma^{n(p)})\kappa^{\Gamma_{n(p),p}} \leq \limsup_{p\to+\infty}\frac{C(\delta_x Q_\gamma^{n(p)})}{-\log(\gamma_{n(p)})}\limsup_{p\to+\infty}\left\{\gamma_{n(p)}(-\log(\gamma_{n(p)}))\right\} = 0 .$$

On the other hand, since $(\gamma_k)_{k\geq 1}$ is nonincreasing, for any $\ell \geq 2$,

$$\sum_{k=n(p)+1}^{p}\gamma_k^\ell \leq \gamma_{n(p)+1}^{\ell-1}\Gamma_{n(p)+1,p} \leq \gamma_{n(p)+1}^{\ell-1}\log(\gamma_{n(p)+1})/\log(\kappa) .$$

The proof follows from (5.10) using $\lim_{p\to\infty}\gamma_{n(p)} = 0$.

(ii) For all $p \geq 1$, define $n(p) = \max(0, \lfloor\log(\Gamma_p)\rfloor)$. Note that since $\lim_{k\to+\infty}\Gamma_k = +\infty$, we have $\lim_{p\to+\infty} n(p) = +\infty$. Using $\sum_{k=1}^{+\infty}\gamma_k^2 < +\infty$ and $(\gamma_k)_{k\geq 1}$ is a nonincreasing sequence, we get for all $\ell \geq 2$,

$$\lim_{p\to+\infty}\sum_{k=n(p)}^{p}\gamma_k^\ell = 0 ,$$

which shows that the first term in the right side of (5.10) goes to 0 as p goes to infinity. As for the second term, since $\limsup_{n\to+\infty}\log\{C(\delta_x Q_\gamma^n)\}/\Gamma_n < +\infty$, we get using that $(\gamma_k)_{k\geq 1}$ is nonincreasing and $n(p) \leq \log(\Gamma_p)$,

$$\begin{aligned}
&C(\delta_x Q_\gamma^{n(p)})\kappa^{\Gamma_{n(p),p}} \\
&\leq \exp\left(\log(\kappa)\Gamma_p + \left[\{\log(C(\delta_x Q_\gamma^{n(p)}))/\Gamma_{n(p)}\}_+ - \log(\kappa)\right]\Gamma_{n(p)}\right) \\
&\leq \exp\left(\log(\kappa)\Gamma_p + \left[\sup_{k\geq 1}\{\log(C(\delta_x Q_\gamma^k))/\Gamma_k\}_+ - \log(\kappa)\right]\gamma_1\log(\Gamma_p)\right) .
\end{aligned}$$

Using $\kappa < 1$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$, we have $\lim_{p\to+\infty} C(\delta_x Q_\gamma^{n(p)})\kappa^{\Gamma_{n(p),p}} = 0$, which concludes the proof.

□

Using (5.10), we can also assess the convergence of the algorithm for constant step sizes $\gamma_k = \gamma$ for all $k \geq 1$. Two different kinds of results can be derived. First, for a given precision $\varepsilon > 0$, we can try to optimize the step size γ to minimize the number of iterations p required to achieve $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$. Second if the total number of iterations is fixed $p \geq 1$, we may determine the step size $\gamma > 0$ which minimizes $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$.

Lemma 5.4. *Assume that (5.10) holds. Assume that there exists $\bar{\gamma} > 0$ such that $\bar{C}(x) = \sup_{\gamma\in(0,\bar{\gamma}]} \sup_{n\geq 1} C(\delta_x R_\gamma^n) < +\infty$ and $\sup_{\gamma\in(0,\bar{\gamma}]} A(\gamma, x) \leq \bar{A}(x)$, where $C(\delta_x R_\gamma^n)$ and $A(\gamma, x)$ are defined in (5.3) and (5.11) respectively. Then for all $\varepsilon > 0$, we get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ if*

$$p > T\gamma^{-1} \quad \text{and} \quad \gamma \leq \frac{-d + \sqrt{d^2 + (2/3)\bar{A}(x)\varepsilon^2 (L^2 T)^{-1}}}{2\bar{A}(x)/3} \wedge \bar{\gamma}\,, \tag{5.17}$$

where

$$T = \left(\log\{\bar{C}(x)\} - \log(\varepsilon/2)\right) \Big/ (-\log(\kappa))\,.$$

Proof. For $p > T\gamma^{-1}$, set $n = p - \lfloor T\gamma^{-1} \rfloor$. Then using the stated expressions of γ and T in (5.10) concludes the proof. □

Note that an upper bound for γ defined in (5.17) is $\epsilon^2 (L^2 T d)^{-1}$. The dependency of T on the dimension d will be addressed in Section 5.3.

Lemma 5.5. *Assume that* **L***1 and (5.3) hold. In addition, assume that there exist $\bar{\gamma} > 0$ and $n \in \mathbb{N}$, $n > 0$, such that $\bar{C}_n(x) = \sup_{\gamma\in(0,\bar{\gamma}]} C(\delta_x R_\gamma^n) < +\infty$ and $\sup_{\gamma\in(0,\bar{\gamma}]} A(\gamma, x) \leq \bar{A}(x)$. For all $p > n$ and all $x \in \mathbb{R}^d$, if $\gamma = \log(p-n)\{(p-n)(-\log(\kappa))\}^{-1} \leq \bar{\gamma}$, then*

$$\begin{aligned}
&\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \\
&\quad \leq (p-n)^{-1/2}\{\bar{C}_n(x)(p-n)^{-1/2} + \log(p-n)(d + \bar{A}(x)\log(p-n)(p-n)^{-1})^{1/2}\}\,.
\end{aligned}$$

Proof. The proof is a straightforward calculation using (5.10). □

To get quantitative bounds for the total variation distance $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ it is therefore required to get bounds on κ, $A(\gamma, x)$ and to control $C(\delta_x Q_\gamma^n)$. We will consider in the sequel two different approaches to get (5.3), one based on functional inequalities, the other on coupling techniques. We will consider also increasingly stringent assumptions for the potential U. Whereas we will always obtain the same type of exponential bounds, the dependency of the constants on the dimension will be markedly different. In the worst case, the dependency is exponential. It is polynomial when U is convex.

5.3 Practical conditions for geometric ergodicity of the Langevin diffusion and their consequences for ULA

5.3.1 Superexponential densities

Assume first that the potential is superexponential outside a ball. This is a rather weak assumption (we do not assume convexity here).

H5. *The potential U is twice continuously differentiable and there exist $\rho > 0$, $\alpha \in (1, 2]$ and $M_\rho \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\rho$, $\langle \nabla U(x), x - x^\star \rangle \geq \rho \|x - x^\star\|^\alpha$.*

The price to pay will be constants which are exponential in the dimension. Under **H** 5, the potential U is unbounded off compact set. Since U is continuous, it has a global minimizer $x^\star$, which is a point at which $\nabla U(x^\star) = 0$. Without loss of generality, it is assumed that $U(x^\star) = 0$.

Lemma 5.6. *Assume **L** 1 and **H** 5. Then for all $x \in \mathbb{R}^d$,*

$$U(x) \geq \rho \|x - x^\star\|^\alpha / (\alpha + 1) - a_\alpha \quad \text{with} \quad a_\alpha = \rho M_\rho^\alpha / (\alpha + 1) + M_\rho^2 L/2 . \tag{5.18}$$

Proof. The elementary proof is postponed to Section 5.4.2. □

Following [RT96a, Theorem 2.3], we first establish a drift condition for the diffusion.

Proposition 5.7. *Assume **L** 1 and **H** 5. For any $\varsigma \in (0, 1)$, the drift condition (5.4) is satisfied with the Lyapunov function $V_\varsigma(x) = \exp(\varsigma U(x))$, $\theta_\varsigma = \varsigma dL$, $\mathcal{E}_\varsigma = \mathrm{B}(x^\star, K_\varsigma)$, $K_\varsigma = \max(\{2dL/(\rho(1 - \varsigma))\}^{1/(2(\alpha-1))}, M_\rho)$ and $\beta_\varsigma = \varsigma dL \sup_{\{y \in \mathcal{E}_\varsigma\}}\{V_\varsigma(y)\}$. Moreover, there exist constants $C_\varsigma < \infty$ and $\upsilon_\varsigma > 0$ such that for all $t \in \mathbb{R}_+$ and probability measures μ_0 and ν_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, satisfying $\mu_0(V_\varsigma) + \nu_0(V_\varsigma) < +\infty$,*

$$\|\mu_0 P_t - \nu_0 P_t\|_{V_\varsigma} \leq C_\varsigma \mathrm{e}^{-\upsilon_\varsigma t} \|\mu_0 - \nu_0\|_{V_\varsigma} , \; \|\mu_0 P_t - \pi\|_{V_\varsigma} \leq C_\varsigma \mathrm{e}^{-\upsilon_\varsigma t} \mu_0(V_\varsigma) .$$

Proof. The proof, adapted from [RT96a, Theorem 2.3] and [MT93a, Theorem 6.1], is postponed to Section 5.4.3. □

Under **H** 5, explicit expressions for C_ς and υ_ς have been developed in the literature but these estimates are in general very conservative. We now turn to establish (5.6) for the Euler discretization.

Proposition 5.8. *Assume **L** 1 and **H** 5. Let $\bar{\gamma} \in (0, L^{-1})$. For all $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$, R_γ satisfies the drift condition (5.6) with $V(x) = \exp(U(x)/2)$, $K = \max(M_\rho, (8\log(\lambda)/\rho^2)^{1/(2(\alpha-1))})$, $c = -2\log(\lambda)\lambda^{-\bar{\gamma}} \sup_{\{y \in \mathrm{B}(x^\star, K)\}} V(y)$ and $\lambda = \mathrm{e}^{-dL/\{2(1-L\bar{\gamma})\}}$.*

Proof. The proof is postponed to Section 5.4.4. □

Theorem 5.9. *Assume* **L** *1 and* **H** *5. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 < \bar{\gamma}$, $\bar{\gamma} \in (0, L^{-1})$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (5.10) holds with $\log(\kappa) = -\upsilon_{1/2}$ and*

$$A(\gamma, x) \leq L^2 \left(\frac{\alpha+1}{\rho}\left[a_\alpha + \frac{4(2-\alpha)(\alpha+1)}{\alpha\rho} + 2\log\left\{G(\lambda, c, \gamma_1, V(x))\right\}\right]\right)^{2/\alpha}$$

$$C(\delta_x Q_\gamma^n) \leq C_{1/2} F(\lambda, \Gamma_{1,n}, c, \gamma_1, V(x)) , \tag{5.19}$$

where $C_{1/2}$, $\upsilon_{1/2}$ are given by Proposition 5.7, F by (5.7), V, λ, c in Proposition 5.8, G by (5.8), a_α in (5.18).

Proof. The proof is postponed to Section 5.4.5. □

Equation (5.19) implies that for all $x \in \mathbb{R}^d$, we have $\sup_{n\geq 0} C(\delta_x Q_\gamma^n) \leq G(\lambda, c, \gamma_1, V(x))$, so Proposition 5.3-(i) shows that $\lim_{p\to+\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ for all $x \in \mathbb{R}^d$ provided that $\lim_{k\to+\infty} \gamma_k = 0$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$. In addition, for the case of constant step size $\gamma_k = \gamma$ for all $k \geq 1$, Lemma 5.4 and Lemma 5.5 can be applied.

Let $V : \mathbb{R}^d \to \mathbb{R}$, defined for all $x \in \mathbb{R}^d$ by $V(x) = \exp(U(x)/2)$. By Proposition 5.7, $(P_t)_{t\geq 0}$ is a contraction operator on the space of finite signed measure $\mu \in \mathbb{M}_0$, $\mu(V^{1/2}) < +\infty$, endowed with the norm $\|\cdot\|_{V^{1/2}}$. It is therefore possible to control $\|\delta_x Q_\gamma^p - \pi\|_{V^{1/2}}$. To simplify the notations, we limit our discussion to constant step sizes.

Theorem 5.10. *Assume* **L** *1 and* **H** *5. Then, for all $p \geq 1$, $x \in \mathbb{R}^d$ and $\gamma \in (0, L^{-1})$, we have*

$$\|\delta_x R_\gamma^p - \pi\|_{V^{1/2}} \leq C_{1/4} \kappa^{\gamma p} V^{1/2}(x) + B(\gamma, V(x)) , \tag{5.20}$$

where $\log(\kappa) = -\upsilon_{1/4}$, $C_{1/4}, \upsilon_{1/4}, \theta_{1/2}, \beta_{1/2}$ are defined in Proposition 5.7, V, λ, c in Proposition 5.8, G in (5.8) and

$$B^2(\gamma, v) = L^2 \max(1, C_{1/4}^2)(1+\gamma)(1-\kappa)^{-2}\left(2G(\lambda, c, \gamma, v) + \beta_{1/2}/\theta_{1/2}\right) \\ \times \left(\gamma d + 3^{-1}\gamma^2 \|\nabla U\|_{V^{1/2}}^2 G(\lambda, c, \gamma, v)\right) .$$

Moreover, R_γ has a unique invariant distribution π_γ and

$$\|\pi - \pi_\gamma\|_{V^{1/2}} \leq B(\gamma, 1) .$$

Proof. The proof of (5.20) is postponed to Section 5.4.6. The bound for $\|\pi - \pi_\gamma\|_{V^{1/2}}$ is an easy consequence of (5.20): by Proposition 5.13 and [MT09, Theorem 16.0.1], R_γ is $V^{1/2}$-uniformly ergodic: $\lim_{p\to+\infty} \|\delta_x R_\gamma^p - \pi_\gamma\|_{V^{1/2}} = 0$ for all $x \in \mathbb{R}^d$. Finally, (5.20) shows that for all $x \in \mathbb{R}^d$,

$$\|\pi - \pi_\gamma\|_{V^{1/2}} \leq \lim_{p\to+\infty}\left\{\|\delta_x R_\gamma^p - \pi\|_{V^{1/2}} + \|\delta_x R_\gamma^p - \pi_\gamma\|_{V^{1/2}}\right\} \leq B(\gamma, V(x)) .$$

Taking the minimum over $x \in \mathbb{R}^d$ concludes the proof. □

Note that Theorem 5.10 implies that there exists a constant $C \geq 0$ which does not depend on γ such that $\|\pi - \pi_\gamma\|_{V^{1/2}} \leq C\gamma^{1/2}$.

Remark 5.11. *It is shown in [TT90, Theorem 4] that for $\phi \in C^\infty(\mathbb{R}^d)$ with polynomial growth, $\pi_\gamma(\phi) - \pi(\phi) = b(\phi)\gamma + \mathcal{O}(\gamma^2)$, for some constant $b(\phi) \in \mathbb{R}$, provided that $U \in C^\infty(\mathbb{R}^d)$ satisfies **L**1 and **H**5. Our result does not match this bound since $B(\gamma, 1) = \mathcal{O}(\gamma^{1/2})$. However the bound $B(\gamma, 1)$ is uniform over the class of measurable functions ϕ satisfying for all $x \in \mathbb{R}^d$, $|\phi(x)| \leq V^{1/2}(x)$. Obtaining such uniform bounds in total variation is important in Bayesian inference, for example to compute high posterior density credible regions. Our result also strengthens and completes [MSH02, Corollary 7.5], which states that under **H**5 with $\alpha = 2$, for any measurable functions $\phi : \mathbb{R}^d \to \mathbb{R}$ satisfying for all $x, y \in \mathbb{R}^d$,*

$$|\phi(x) - \phi(y)| \leq C\|x - y\|\left\{1 + \|x\|^k + \|y\|^k\right\},$$

for some $C \geq 0$, $k \geq 1$, $|\pi_\gamma(\phi) - \pi(\phi)| \leq C\gamma^\chi$ for some constants $C \geq 0$ and $\chi \in (0, 1/2)$, which does not depend on ϕ.

The bounds in Theorem 5.9 and Theorem 5.10 depend upon the constants appearing in Proposition 5.7 which are computable but are known to be pessimistic in general; see [RT00]. More explicit rates of convergence for the semigroup can be obtained using Poincaré inequality; see [BCG08], [CG09] and [BGL14, Chapter 4] and the references therein. The probability measure π is said to satisfy a Poincaré inequality with the constant C_{P} if for every locally Lipschitz function h,

$$\mathrm{Var}_\pi\{h\} \leq C_{\mathrm{P}} \int_{\mathbb{R}^d} \|\nabla h(x)\|^2 \pi(\mathrm{d}x). \tag{5.21}$$

This inequality implies by [CG09, Theorem 2.1] that for all $t \geq 0$ and any initial distribution μ_0, such that $\mu_0 \ll \pi$,

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq \exp(-t/C_{\mathrm{P}}) \left(\mathrm{Var}_\pi\{\mathrm{d}\mu_0/\mathrm{d}\pi\}\right)^{1/2}. \tag{5.22}$$

[Bak+08, Theorem 1.4] shows that if the Lyapunov condition (5.4) is satisfied, then the Poincaré inequality (5.21) holds with an explicit constant. Denote by

$$D_n(\gamma) \overset{\text{def}}{=} \left(4\pi\left\{\prod_{k=1}^n (1 - L\gamma_k)\right\}^2 \sum_{i=1}^n \gamma_i (1 - L\gamma_i)^{-1}\right)^{-d/2}. \tag{5.23}$$

Theorem 5.12. *Assume **L**1 and **H**5. Let $(\gamma_k)_{k\geq 1}$ be a non increasing sequence. Then for all $n \geq 1$ and $x \in \mathbb{R}^d$, Equation (5.3) holds with*

$$\log(\kappa) = \left(-\theta_{1/2}^{-1}\left\{1 + (4\beta_{1/2}K_{1/2}^2/\pi^2)\mathrm{e}^{\mathrm{osc}_{K_{1/2}}(U)}\right\}\right)^{-1},$$
$$C(\delta_x Q_\gamma^n) \leq \frac{(\alpha+1)^d (2\pi)^{(d+1)/2}(d-1)!}{\rho^d \mathbf{\Gamma}((d+1)/2)} D_n(\gamma)\mathrm{e}^{a_\alpha}\,\mathrm{e}^{U(x)},$$

where $\mathbf{\Gamma}$ is the Gamma function and the constants $\beta_{1/2}, \theta_{1/2}, K_{1/2}, a_\alpha$ are given in Proposition 5.7 and (5.18) respectively.

Proof. The proof is postponed to Section 5.4.7. □

Note that for all $x \in \mathbb{R}^d$, $C(\delta_x Q_\gamma^n)$ satisfies the conditions of Proposition 5.3-(ii). Therefore using in addition the bound on $A(\gamma, x)$ for all $x \in \mathbb{R}^d$ and $\gamma \in (0, L^{-1})$ given in Theorem 5.9, we get $\lim_{k\to+\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ if $\lim_{n\to+\infty} \Gamma_n = +\infty$ and $\lim_{n\to+\infty} \sum_{k=1}^n \gamma_k^2 < +\infty$.

5.3.2 Log-concave densities

We now consider the following additional assumption.

H 6. *U is convex and admits a minimizer $x^\star$ for U. Moreover there exist $\eta > 0$ and $M_\eta \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\eta$,*

$$U(x) - U(x^\star) \geq \eta \|x - x^\star\| \ . \tag{5.24}$$

It is shown in [Bak+08, Lemma 2.2] that if U satisfies **L** 1 and is convex, then (5.24) holds for some constants η, M_η which depend in an intricate way on U. Since the constants η, M_η appear explicitly in the bounds we derive, we must assume that these constants are explicitly computable. We still assume in this section that $U(x^\star) = 0$. Define the function $W_{\mathrm{c}} : \mathbb{R}^d \to [1, +\infty)$ for all $x \in \mathbb{R}^d$ by

$$W_{\mathrm{c}}(x) = \exp((\eta/4)(\|x - x^\star\|^2 + 1)^{1/2}) \ . \tag{5.25}$$

We now derive a drift inequality for R_γ under **H**6.

Proposition 5.13. *Assume **L** 1 and **H** 6. Let $\bar{\gamma} \in (0, L^{-1}]$. Then for all $\gamma \in (0, \bar{\gamma}]$, W_{c} satisfies (5.6) with $\lambda = \mathrm{e}^{-2^{-4}\eta^2(2^{1/2}-1)}$, $R_{\mathrm{c}} = \max(1, 2d/\eta, M_\eta)$,*

$$c = \{(\eta/4)(d + (\eta\bar{\gamma}/4)) - \log(\lambda)\} \, \mathrm{e}^{\eta(R_{\mathrm{c}}^2+1)^{1/2}/4 + (\eta\bar{\gamma}/4)(d + (\eta\bar{\gamma}/4))} \ . \tag{5.26}$$

Proof. The proof is postponed to Section 5.4.8 □

Corollary 5.14. *Assume **L** 1 and **H** 6. Let $(\gamma_k)_{k\geq1}$ be a nonincreasing sequence with $\gamma_1 \leq \bar{\gamma}$, $\bar{\gamma} \in (0, L^{-1}]$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$,*

$$A(\gamma, x) = L^2 \left(4\eta^{-1} \left[1 + \log\left\{G(\lambda, c, \gamma_1, W_{\mathrm{c}}(x))\right\}\right]\right)^2 \ , \tag{5.27}$$

where $A(\gamma, x)$ is defined by (5.11) and G, W_{c}, λ, c, are given in (5.8), (5.25), Proposition 5.13 respectively.

Proof. The proof is postponed to Section 5.4.9. □

If U is convex, [Bob99, Theorem 1.2] shows that π satisfies a Poincaré inequality with a constant depending only on the variance of π.

Theorem 5.15. *Assume* **L**1 *and* **H**6*. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq \bar{\gamma}$, $\bar{\gamma} \in (0, L^{-1}]$*. Then, for all* $n \geq 0$, $p \geq 1$, $n < p$*, and* $x \in \mathbb{R}^d$, (5.10) *holds with* $A(\gamma, x)$ *given in* (5.27),

$$\log(\kappa) = \left(-432 \int_{\mathbb{R}^d} \left\| x - \int_{\mathbb{R}^d} y \pi(\mathrm{d}y) \right\|^2 \pi(\mathrm{d}x)\right)^{-1} \tag{5.28a}$$

$$C(\delta_x Q_\gamma^n) = \left(\frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d \mathbf{\Gamma}((d+1)/2)} + \frac{\pi^{d/2} M_\eta^d}{\mathbf{\Gamma}(d/2+1)}\right) D_n(\gamma) \exp(U(x)) \;, \tag{5.28b}$$

where $D_n(\gamma)$ *is given in* (5.23).

Proof. The proof is postponed to Section 5.4.10. □

For all $x \in \mathbb{R}^d$, $C(\delta_x Q_\gamma^n)$ satisfies the conditions of Proposition 5.3-(ii). Therefore, if $\lim_{n\to+\infty} \Gamma_n = +\infty$ and $\lim_{n\to+\infty} \sum_{k=1}^n \gamma_k^2 < +\infty$, we get $\lim_{k\to+\infty} \|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$.

There are two difficulties when applying Theorem 5.15. First the Poincaré constant (5.28a) is in closed form but is not computable, although it can be bounded by a $\mathcal{O}(d^{-2})$. Second, the bound of $\mathrm{Var}_\pi\{\mathrm{d}\delta_x Q_\gamma^n / \mathrm{d}\pi\}$ is likely to be suboptimal. To circumvent these two issues, we now give new quantitative results on the convergence of $(P_t)_{t\geq 0}$ to π in total variation. Instead of using functional inequality, we use in the proof the coupling by reflection, introduced in [LR86]. Define the function $\omega : (0,1) \times \mathbb{R}_+^* \to \mathbb{R}_+$ for all $\epsilon \in (0,1)$ and $R \geq 0$, by

$$\omega(\epsilon, R) = R^2 / \left\{2\mathbf{\Phi}^{-1}(1-\epsilon/2)\right\}^2 \;, \tag{5.29}$$

where $\mathbf{\Phi}$ is the cumulative distribution function of the standard Gaussian distribution and $\mathbf{\Phi}^{-1}$ is the associated quantile function. Before stating the theorem, we first show that (5.4) holds and provide explicit expressions for the constants which come into play. These constants will be used to obtain the explicit convergence rate of the semigroup $(P_t)_{t\geq 0}$ to π which is derived in Theorem 5.17.

Proposition 5.16. *Assume* **L**1 *and* **H**6*. Then* W_{c} *satisfies the drift condition* (5.4) *with* $\theta = \eta^2/8$, $\mathcal{E} = \mathrm{B}(x^\star, K)$, $K = \max(1, M_\eta, 4d/\eta)$ *and*

$$\beta = (\eta/4)\left((\eta/4)K + d\right) \max\left\{1, (K^2+1)^{-1/2} \exp(\eta(K^2+1)^{1/2}/4)\right\} \,.$$

Proof. The proof is adapted from [Bak+08, Corollary 1.6] and is postponed to Section 5.4.11. □

Theorem 5.17. *Assume* **L**1 *and* **H**6*. Then for all* $x \in \mathbb{R}^d$,

$$\|\delta_x P_t - \pi\|_{\mathrm{TV}} \leq \Lambda(x) \mathrm{e}^{-\theta t/4} + 2\varpi^t \;,$$

where

$$\log(\varpi) = -\log(2)(\theta/4) \times \left[\log\left\{\theta^{-1}\beta\left(3+4\mathrm{e}^{2\theta^{-1}\omega\left(2^{-1},(8/\eta)\log(4\theta^{-1}\beta)\right)}\right)\right\}+\log(2)\right]^{-1} , \tag{5.30a}$$

$$\Lambda(x) = (1/2)(W_{\mathrm{c}}(x)+\theta^{-1}\beta) + 2\theta^{-1}\beta\mathrm{e}^{4\theta^{-1}\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} , \tag{5.30b}$$

the function W_{c} is defined in (5.25), the constants θ, β in Proposition 5.16.

Proof. The proof is postponed to Section 5.5.1. ☐

Note that the bound, we obtain is a little different from (5.3). The initial condition is isolated on purpose to get a better bound. A consequence of this result is the following bound on the convergence of the sequence $(\delta_x Q_\gamma^n)_{n\geq 0}$ to π.

Corollary 5.18. *Assume **L** 1 and **H** 6. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $x \in \mathbb{R}^d$, $n \geq 0$, $p \geq 1$, $n < p$,*

$$\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq 2^{-1/2} L \left(\sum_{k=n}^{p-1}\left\{(\gamma_{k+1}^3/3)A(\gamma,x) + d\gamma_{k+1}^2\right\}\right)^{1/2} + \Lambda(\delta_x Q_\gamma^n)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 2\varpi^{\Gamma_{n+1,p}} ,$$

where $A(\gamma,x)$, ϖ are given by (5.27) and (5.30a) respectively and

$$\Lambda(\delta_x Q_\gamma^n) = (1/2)(F(\lambda,\Gamma_n,\gamma_1,c,W_{\mathrm{c}}(x)) + \theta^{-1}\beta) + 2\theta^{-1}\beta\mathrm{e}^{4\theta^{-1}\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} , \tag{5.31}$$

the functions F and W_{c} are defined in (5.7) and (5.25), the constants $\lambda, c, \theta, \beta$ in Proposition 5.13 and Proposition 5.16 respectively.

Proof. By Theorem 5.17, we have for all $x \in \mathbb{R}^d$,

$$\|\delta_x P_{\Gamma_{n+1,p}} - \pi\|_{\mathrm{TV}} \leq \Lambda(\delta_x Q_\gamma^n)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 2\varpi^{\Gamma_{n+1,p}}\Lambda(x)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 2\varpi^{\Gamma_{n+1,p}} .$$

By Proposition 5.13 and Lemma 5.1.

$$\|\delta_x Q_\gamma^n P_{\Gamma_{n+1,p}} - \pi\|_{\mathrm{TV}} \leq \Lambda(\delta_x Q_\gamma^n)\mathrm{e}^{-\theta\Gamma_{n+1,p}/4} + 2\varpi^{\Gamma_{n+1,p}} .$$

Finally the proof follows the same line as the one of Proposition 5.2. ☐

Contrary to (5.28b), (5.31) is uniformly bounded in n. By Corollary 5.18 and (5.27), we can apply Proposition 5.3-(i), which implies the convergence to 0 of $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}}$ as p goes to infinity, if $\lim_{k\to+\infty}\gamma_k = 0$ and $\lim_{k\to+\infty}\Gamma_k = +\infty$. Since $\log(\beta)$ in Proposition 5.16 is of order d, we get that the rate of convergence $\log(\kappa)$ is of order d^{-2} as d goes to infinity (note indeed that the leading term when d is large is $2\theta^{-1}\omega\left(2^{-1},(8/\eta)\log(4\theta^{-1}\beta)\right)$ which is of order d^2). In the case of constant step sizes $\gamma_k = \gamma$ for all $k \geq 0$, we adapt Lemma 5.4 to the bound given by 5.18.

	d	ε	L
γ	$\mathcal{O}(d^{-3})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$	$\mathcal{O}(L^{-2})$
p	$\mathcal{O}(d^5)$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$	$\mathcal{O}(L^2)$

Table 5.1: For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Corollary 5.19

Corollary 5.19. *Assume **L** 1 and **H** 6. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $\varepsilon > 0$, we get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ if*

$$p > T\gamma^{-1} \quad \text{and} \quad \gamma \leq \frac{-d + \sqrt{d^2 + (2/3)A(\gamma,x)\varepsilon^2(L^2T)^{-1}}}{2A(\gamma,x)/3} \wedge L^{-1} , \tag{5.32}$$

where

$$T = \max\left\{4\theta^{-1}\log\left(4\varepsilon^{-1}\tilde{\Lambda}(x)\right), \log(8\varepsilon^{-1})\Big/(-\log(\varpi))\right\}$$
$$\tilde{\Lambda}(x) = (1/2)(G(\lambda,\gamma_1,c,W_{\mathrm{c}}(x)) + \theta^{-1}\beta) + 2\theta^{-1}\beta \mathrm{e}^{4\theta^{-1}\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))} ,$$

where $A(\gamma,x)$, ϖ are given by (5.27), (5.30a) respectively, the functions G and W_{c} are defined in (5.8) and (5.25), the constants $\lambda, c, \theta, \beta$ in Proposition 5.13 and Proposition 5.16 respectively.

Proof. The proof follows the same line as the one of Lemma 5.4 using Corollary 5.18 and that $\sup_{n\geq 0}\Lambda(\delta Q_\gamma^n) < \tilde{\Lambda}(x)$ for all $x \in \mathbb{R}^d$. □

In particular, with the notation of Corollary 5.19, since $\max(\log(\beta), \log(c))$ and $-(\log(\varpi))^{-1}$ are of order d and d^2 as d goes to infinity respectively, T is of order d^2. Therefore, γ defined by (5.32) is of order d^{-3} which implies a number of iteration p of order d^5 to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$; see also Table 5.1.

Corollary 5.19 can be compared with the results which establishes the dependency on the dimension for two kinds of Metropolis-Hastings algorithms to sample from a log-concave density: the random walk Metropolis algorithm (RWM) and the hit-and-run algorithm. It has been shown in [LV07, Theorem 2.1] that for $\varepsilon > 0$, the hit-and-run and the RWM reach a ball centered at π, of radius ε for the total variation distance, in a number of iteration p of order d^4 as d goes to infinity. It should be stressed that [LV07, Theorem 2.1] does not assume any kind of smoothness about the density π contrary to Theorem 5.17. However, this result assumes that the target distribution is near-isotropic, i.e. there exists $C \geq 0$ which does not depend on the dimension such that for all $x \in \mathbb{R}^d$,

$$C^{-1}\|x\|^2 \leq \int_{\mathbb{R}^d} \langle x, y\rangle^2 \pi(\mathrm{d}y) \leq C\|x\|^2 .$$

Note that this condition implies that the variance of π is upper bounded by Cd. With the same kind of assumption, we can improve the dependence on the dimension in the bounds given by Corollary 5.19.

	d	ε	L
γ	$\mathcal{O}(d^{-2})$	$\mathcal{O}(\varepsilon^{4})$	$\mathcal{O}(L^{-2})$
p	$\mathcal{O}(d^{3})$	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{O}(L^{2})$

Table 5.2: For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Theorem 5.20

H 7. *There exists a constant* $\mathsf{C} \in \mathbb{R}_+$ *independent of the dimension such that*

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \pi(\mathrm{d}y) \leq \mathsf{C}^2 d \,.$$

Theorem 5.20. *Assume **L** 1, **H** 6 and **H** 7. Let* $(\gamma_k)_{k\geq 0}$ *be a sequence of nonnegative step sizes. Then for all* $\varepsilon > 0$*, we get* $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ *if* p *and* γ *satisfy* (5.32) *with*

$$T = 4\varepsilon^{-2}\pi^{-1} \max\left\{\|x - x^\star\|, \mathsf{C}d^{1/2}\right\}^2 \,.$$

Proof. The proof is postponed to Section 5.4.12. □

Since T is of order d, therefore γ defined by (5.32) is of order d^{-2} which implies a number of iteration p of order d^3 to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$. However the dependence on ε is less good than in Corollary 5.19. Indeed since T is of order ε^{-2}, γ defined by (5.32) is of order ε^{-4} which implies a number of iteration p of order ε^{-6} to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$; see also Table 5.2. Therefore, we have a better dependency on the dimension than [LV07, Theorem 2.1]. Regarding the dependence on ε [LV07, Theorem 2.1] does not give the order of the bounds and we cannot compare to this result.

To conclude our study on convex potential, we also mention [BDJ98] which studies the sampling of the uniform distribution over a convex subset $\mathsf{K} \subset \mathbb{R}^d$ using coupling techniques. Let $C > 0$. A convex set $\mathsf{K} \subset \mathbb{R}^d$ is C-well rounded if $\mathrm{B}(0,1) \subset \mathsf{K} \subset \mathrm{B}(0, Cd)$. [BDJ98] shows that a number of iteration of order d^9 as d goes to infinity is sufficient to sample uniformly over any C-well rounded convex set. Comparison with our result is difficult since we assume that π is positive on $\mathbb{R}^d$, continuously differentiable, while [BDJ98] studies the case of uniform distributions over a convex body. An adaptation of our result to non continuously differentiable potentials will appear in a forthcoming paper [DMP].

5.3.3 Strongly log-concave densities

More precise bounds can be obtained in the case where U is assumed to be strongly convex outside some ball; this assumption has been considered by [Ebe15] for convergence in the Wasserstein distance; see also [BGG12].

H 8 (M_{s})**.** *U is convex and there exist $M_{\mathrm{s}} \geq 0$ and $m > 0$, such that for all $x, y \in \mathbb{R}^d$ satisfying $\|x - y\| \geq M_{\mathrm{s}}$,*

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq m \|x - y\|^2 \ .$$

We will see in the sequel that under this assumption the convergence rate in (5.3) does not depend on the dimension d but only on the constants m and M_{s}.

Proposition 5.21. *Assume **L**1 and **H**8(M_{s}). Let $\bar{\gamma} \in (0, 2mL^{-2})$. For all $\gamma \in (0, \bar{\gamma}]$, $V(x) = \|x - x^\star\|^2$ satisfies (5.6) with $\lambda = \mathrm{e}^{-2m + \bar{\gamma} L^2}$ and $\mathrm{c} = 2(d + m M_{\mathrm{s}}^2)$.*

Proof. The proof is postponed to Section 5.4.13. □

Theorem 5.22. *Assume **L**1 and **H**8(M_{s}). Let $(\gamma_k)_{k \geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq \bar{\gamma}$, $\bar{\gamma} \in (0, 2mL^{-2})$. Then, for all $n \geq 0$, $p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (5.10) holds with*

$$\begin{aligned}
\log(\kappa) &= -(m/2)\log(2) \\
&\qquad \times \left[\log\left\{\left(1 + \mathrm{e}^{m\omega(2^{-1}, \max(1, M_{\mathrm{s}}))/4}\right)(1 + \max(1, M_{\mathrm{s}}))\right\} + \log(2)\right]^{-1} \\
C(\delta_x Q_\gamma^n) &\leq 3 + \left(d/m + M_{\mathrm{s}}^2\right)^{1/2} + F^{1/2}(\lambda, \Gamma_{1,n}, \mathrm{c}, \gamma_1, \|x - x^\star\|^2) \\
A(\gamma, x) &\leq L^2\, G(\lambda, \mathrm{c}, \gamma_1, \|x - x^\star\|^2) \ ,
\end{aligned}$$

where F, G, ω are defined by (5.7), (5.8), (5.29) respectively, and λ, c are given in Proposition 5.21.

Proof. The proof is postponed to Section 5.5.1. □

Note that the conditions of Proposition 5.3-(i) are fulfilled. For constant step sizes $\gamma_k = \gamma$ for all $k \geq 1$, Lemma 5.4 and Lemma 5.5 can be applied. We give in Table 5.3 the dependency of the step size $\gamma > 0$ and the minimum number of iterations $p \geq 0$, provided in Lemma 5.4, on the dimension d and the other constants related to U, to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$, for a target precision $\varepsilon > 0$. We can see that the dependency on the dimension is milder than for the convex case. The number of iteration requires to reach a target precision ε is just of order $\mathcal{O}(d \log(d))$.

Consider the case where π is the d-dimensional standard Gaussian distribution. Then for all $p \in \mathbb{N}$, $\gamma \in (0,1)$ and $x \in \mathbb{R}^d$, $\delta_x R_\gamma^p$ is the d-dimensional Gaussian distribution with mean $(1-\gamma)^p x$ and covariance matrix $\sigma_\gamma \mathrm{I}_d$, with $\sigma_\gamma = (1 - (1-\gamma)^{2(p+1)})(1 - \gamma/2)^{-1}$. Therefore using the Pinsker inequality, we get:

$$\begin{aligned}
\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}}^2 &\leq 2\,\mathrm{KL}\left(\delta_x R_\gamma^p \middle| \pi\right) \\
&\leq d\left[\log(\sigma_\gamma) - 1 + \sigma_\gamma^{-1}\left\{1 + (1-\gamma)^{2p}\|x\|^2 d^{-1}\right\}\right] \ .
\end{aligned}$$

	d	ε	L	m	M_{s}
γ	$\mathcal{O}(d^{-1})$	$\mathcal{O}(\varepsilon^2/\log(\varepsilon^{-1}))$	$\mathcal{O}(L^{-2})$	$\mathcal{O}(m)$	$\mathcal{O}(M_{\mathrm{s}}^{-4})$
p	$\mathcal{O}(d\log(d))$	$\mathcal{O}(\varepsilon^{-2}\log^2(\varepsilon^{-1}))$	$\mathcal{O}(L^2)$	$\mathcal{O}(m^{-2})$	$\mathcal{O}(M_{\mathrm{s}}^{8})$

Table 5.3: For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Theorem 5.22

Using the inequalities for all $t \in (0,1)$, $(1-t)^{-1} \leq 1 + t(1-t)^{-2}$ and for all $s \in (0,1/2)$, $-\log(1-s) \leq s + 2s^2$, we have:

$$\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}}^2 \leq d\left\{\gamma^2/2 + (1-\gamma)^{2(p+1)}(1-\gamma/2)(1-(1-\gamma)^{2(p+1)})^{-2}\right\} + \sigma_\gamma^{-1}(1-\gamma)^{2p}\|x\|^2 .$$

This inequality implies that in order to have $\|\delta_x R_\gamma - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$, the step size γ has to be of order $d^{-1/2}$ and p of order $d^{1/2}\log(d)$. Therefore, the dependency on the dimension reported in Table 5.3 does not match this particular example. However it does not imply that this dependency can be improved.

5.3.4 Bounded perturbation of strongly log-concave densities

We now consider the case where U is a bounded perturbation of a strongly convex potential.

H9. *The potential U may be expressed as $U = U_1 + U_2$, where*

*(a) $U_1 : \mathbb{R}^d \to \mathbb{R}$ satisfies **H**8(0) (i.e. is strongly convex) and there exists $L_1 \geq 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U_1(x) - \nabla U_1(y)\| \leq L_1 \|x - y\| .$$

(b) $U_2 : \mathbb{R}^d \to \mathbb{R}$ is continuously differentiable and $\|U_2\|_\infty + \|\nabla U_2\|_\infty < +\infty$.

The probability measure π is said to satisfy a log-Sobolev inequality with constant $C_{\mathrm{LS}} > 0$ if for all locally Lipschitz function $h : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathrm{Ent}_\pi\left(h^2\right) \leq 2C_{\mathrm{LS}} \int \|\nabla h\|^2 \mathrm{d}\pi .$$

Then [CG09, Theorem 2.7] shows that for all $t \geq 0$ and any probability measure $\mu_0 \ll \pi$ satisfying $\mathrm{d}\mu_0/\mathrm{d}\pi \log(\mathrm{d}\mu_0/\mathrm{d}\pi) \in \mathrm{L}^1(\pi)$, we have

$$\|\mu_0 P_t - \pi\|_{\mathrm{TV}} \leq \mathrm{e}^{-t/C_{\mathrm{LS}}}\left\{2\,\mathrm{Ent}_\pi\left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\pi}\right)\right\}^{1/2} . \tag{5.33}$$

Under **H** 9, [BGL14, Corollary 5.7.2] and the Holley-Stroock perturbation principle [HS87, p. 1184], π satisfies a log-Sobolev inequality with a constant which only depends on the strong convexity constant m of U_1 and $\mathrm{osc}_{\mathbb{R}^d}(U_2)$. Define

$$\varpi = \frac{2mL_1}{m+L_1} \; . \tag{5.34}$$

Denote by $x_1^\star$ the minimizer of U_1.

Proposition 5.23. *Assume **H** 9. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L_1)$. Then for all $p \geq 1$ and $x \in \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2 \, Q_\gamma^p(x, \mathrm{d}y) \leq \prod_{k=1}^{p} (1 - \varpi\gamma_k/2) \, \|x - x_1^\star\|^2 \\ + 2\varpi^{-1}(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) \; .$$

Proof. The proof is postponed to Section 6.8.2. □

Theorem 5.24. *Assume **L** 1 and **H** 9. Let $(\gamma_k)_{k\in\mathbb{N}^*}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L_1)$. Then, for all $n, p \geq 1$, $n < p$, and $x \in \mathbb{R}^d$, (5.10) holds with*

$$\begin{aligned} &-\log(\kappa) = m\exp\{-\mathrm{osc}_{\mathbb{R}^d}(U_2)\} \\ &C^2(\delta_x Q_\gamma^n) \leq L_1 \mathrm{e}^{-\varpi\Gamma_n/2} \, \|x - x_1^\star\|^2 + L_1\gamma_n(\gamma_n + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2 + 2\mathrm{osc}_{\mathbb{R}^d}(U_2) \\ &+ 2L_1\varpi^{-1}(1 - \varpi\gamma_n)(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) - d(1 + \log(2\gamma_n m) - 2L_1\gamma_n) \\ &A(\gamma, x) \leq 2L_1^2 \left\{ \|x_1^\star - x^\star\|^2 + 2\varpi^{-1}(2d + (\gamma_1 + 2\varpi^{-1}) \, \|\nabla U_2\|_\infty^2) \right\} + 2 \, \|\nabla U_2\|_\infty^2 \; , \end{aligned} \tag{5.35}$$

where ϖ is defined in (5.34).

Proof. The proof is postponed to Section 5.4.15. □

Note that by (5.35), $\sup_{n\geq 1}\{C(\delta_x Q_\gamma^n)/(-\log(\gamma_n))\} < +\infty$, therefore Proposition 5.3-(i) can be applied and $\lim_{p\to+\infty} \|\delta_\gamma Q_\gamma^p - \pi\|_{\mathrm{TV}} = 0$ if $\lim_{k\to+\infty} \gamma_k = 0$ and $\lim_{k\to+\infty} \Gamma_k = +\infty$.

5.4 Proofs

5.4.1 Proof of Lemma 5.1

By a straightforward induction, we get for all $n \geq 0$ and $x \in \mathbb{R}^d$,

$$Q_\gamma^n V(x) \leq \lambda^{\Gamma_n} V(x) + c\sum_{i=1}^{n} \gamma_i \lambda^{\Gamma_{i+1,n}} \; . \tag{5.36}$$

Note that for all $n \geq 1$, we have since $(\gamma_k)_{k\geq 1}$ is nonincreasing and for all $t \geq 0$, $\lambda^t = 1 + \int_0^t \lambda^s \log(\lambda) \mathrm{d}s$,

$$\begin{aligned}
&\sum_{i=1}^{n} \gamma_i \lambda^{\Gamma_{i+1,n}} \leq \sum_{i=1}^{n} \gamma_i \prod_{j=i+1}^{n} (1 + \lambda^{\gamma_1} \log(\lambda)\gamma_j) \\
&\leq (-\lambda^{\gamma_1} \log(\lambda))^{-1} \sum_{i=1}^{n} \gamma_i \left\{ \prod_{j=i+1}^{n} (1 + \lambda^{\gamma_1} \log(\lambda)\gamma_j) - \prod_{j=i}^{n} (1 + \lambda^{\gamma_1} \log(\lambda)\gamma_j) \right\} \\
&\leq (-\lambda^{\gamma_1} \log(\lambda))^{-1} .
\end{aligned}$$

The proof is then completed using this inequality in (5.36).

5.4.2 Proof of Lemma 5.6

By **L**1, **H**5, the Cauchy-Schwarz inequality and $\nabla U(x^\star) = 0$, for all $x \in \mathbb{R}^d$, $\|x\| \geq M_\rho$, we have

$$\begin{aligned}
U(x) - U(x^\star) &= \int_0^1 \langle \nabla U(x^\star + t(x - x^\star)), x - x^\star \rangle \, \mathrm{d}t \\
&\geq \int_0^{\frac{M_\rho}{\|x - x^\star\|}} \langle \nabla U(x^\star + t(x - x^\star)), x - x^\star \rangle \, \mathrm{d}t \\
&\qquad + \int_{\frac{M_\rho}{\|x - x^\star\|}}^1 \langle \nabla U(x^\star + t(x - x^\star)), t(x - x^\star) \rangle \, \mathrm{d}t \\
&\geq -M_\rho^2 L/2 + \rho \|x - x^\star\|^\alpha (\alpha + 1)^{-1} \left\{ 1 - (M_\rho / \|x - x^\star\|)^{\alpha+1} \right\} .
\end{aligned}$$

On the other hand using again **L**1, the Cauchy-Schwarz inequality and $\nabla U(x^\star) = 0$, for all $x \in \mathrm{B}(x^\star, M_\rho)$,

$$U(x) - U(x^\star) = \int_0^1 \langle \nabla U(x^\star + t(x - x^\star)), x - x^\star \rangle \, \mathrm{d}t \geq -M_\rho^2 L/2 ,$$

which concludes the proof.

5.4.3 Proof of Proposition 5.7

For all $x \in \mathbb{R}^d$, we have

$$\mathscr{A}^{\mathrm{L}} V_\varsigma(x) = \varsigma(1 - \varsigma) \left\{ - \|\nabla U(x)\|^2 + (1 - \varsigma)^{-1} \Delta U(x) \right\} V_\varsigma(x) .$$

If $\alpha > 1$, by the Cauchy-Schwarz inequality, under **L**1-**H**5 for all $x \in \mathbb{R}^d$, $\Delta U(x) \leq dL$ and $\|\nabla U(x)\| \geq \rho \|x - x^\star\|^{\alpha-1}$ for $\|x - x^\star\| \geq M_\rho$. Then, for all $x \notin \mathcal{E}_\varsigma$,

$$\mathscr{A}^{\mathrm{L}} V_\varsigma(x) \leq \varsigma(1 - \varsigma) \left\{ -\rho \|x - x^\star\|^{2(\alpha-1)} + (1 - \varsigma)^{-1} dL \right\} V_\varsigma(x) \leq -\varsigma dL V_\varsigma(x) ,$$

and $\sup_{\{x \in \mathcal{E}_\varsigma\}} \mathscr{A}^{\mathrm{L}} V_\varsigma(x) \leq \varsigma dL \sup_{\{y \in \mathcal{E}_\varsigma\}} \{V_\varsigma(y)\}$.

5.4.4 Proof of Proposition 5.8

By **H**5, for all $x \notin \mathrm{B}(x^\star, M_\rho)$,

$$\|\nabla U(x)\| \geq \rho \|x - x^\star\|^{\alpha-1} \; . \tag{5.37}$$

Since under **L**1, for all $x, y \in \mathbb{R}^d$, $U(y) \leq U(x) + \langle \nabla U(x), y - x\rangle + (L/2)\|y - x\|^2$, we have for all $\gamma \in (0, \bar{\gamma})$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}
&R_\gamma V(x)/V(x) \\
&= (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(\{U(y) - U(x)\}/2 - (4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2 \right) \mathrm{d}y \\
&\leq (4\pi\gamma)^{-d/2} \int_{\mathbb{R}^d} \exp\left(-4^{-1}\gamma \|\nabla U(x)\|^2 - (4\gamma)^{-1}(1 - \gamma L) \|y - x\|^2 \right) \mathrm{d}y \\
&\leq (1 - \gamma L)^{-d/2} \exp(-4^{-1}\gamma \|\nabla U(x)\|^2) \; ,
\end{aligned}$$

where we used in the last line that $\gamma < L^{-1}$. Since $\log(1 - L\gamma) = -L\int_0^\gamma (1 - Lt)^{-1} \mathrm{d}t$, for all $\gamma \in (0, \bar{\gamma}]$, $\log(1 - L\gamma) \geq -L\gamma(1 - L\bar{\gamma})^{-1}$. Using this inequality, we get

$$R_\gamma V(x)/V(x) \leq \lambda^{-\gamma} \exp\left(-4^{-1}\gamma \|\nabla U(x)\|^2 \right) \; . \tag{5.38}$$

By (5.37), for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq K$, we have

$$R_\gamma V(x) \leq \lambda^\gamma V(x) \; . \tag{5.39}$$

Also by (5.38) and since for all $t \geq 0$, $\mathrm{e}^t - 1 \leq t\mathrm{e}^t$, we get for all $x \in \mathbb{R}^d$

$$R_\gamma V(x) - \lambda^\gamma V(x) \leq \lambda^\gamma(\lambda^{-2\gamma} - 1)V(x) \leq -2\gamma \log(\lambda)\lambda^{-\bar{\gamma}} V(x) \; .$$

The proof is completed combining the last inequality and (5.39).

5.4.5 Proof of Theorem 5.9

We first bound $A(\gamma, x)$ for all $x \in \mathbb{R}^d$. Let $x \in \mathbb{R}^d$. By **L**1, we have

$$\mathbb{E}_x[\|\nabla U(X_k)\|^2] \leq L^2 \mathbb{E}_x[\|X_k - x^\star\|^2] \; . \tag{5.40}$$

Consider now the function $\phi_\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ defined for all $t \geq 0$ by $\phi_\alpha(t) = \exp(A_\alpha(t + B_\alpha)^{\alpha/2})$ where $A_\alpha = \rho/(2(\alpha+1))$ and $B_\alpha = \{(2-\alpha)/(\alpha A_\alpha)\}^{2/\alpha}$. Since ϕ_α is convex and invertible on $\mathbb{R}_+$, we get using the Jensen inequality and Lemma 5.6 for all $k \geq 0$:

$$\mathbb{E}_x[\|X_k - x^\star\|^2] \leq \phi_\alpha^{-1}\left(\mathbb{E}_x[\phi_\alpha\left(\|X_k - x^\star\|^2\right)] \right) \leq \phi_\alpha^{-1}\left(\mathrm{e}^{a_\alpha/2 + B_\alpha^{\alpha/2}} \mathbb{E}_x[V(X_k)] \right) \; ,$$

where $V(x) = \exp(U(x)/2)$. Using that for all $t \geq 0$, $\phi_\alpha^{-1}(t) \leq (A_\alpha^{-1}\log(t))^{2/\alpha}$ and Lemma 5.1, we get

$$\sup_{k \geq 0} \mathbb{E}_x[\|X_k - x^\star\|^2] \leq \left(A_\alpha^{-1} \left[a_\alpha/2 + B_\alpha^{\alpha/2} + \log\{G(\lambda, c(\gamma_1), V(x))\} \right] \right)^{2/\alpha} \; .$$

Eq. (5.19) follows from Proposition 5.7, Proposition 5.8 and Lemma 5.1.

5.4.6 Proof of Theorem 5.10

Lemma 5.25. *Let μ and ν be two probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $V : \mathbb{R}^d \to [1, \infty)$ be a measurable function. Then*

$$\|\mu - \nu\|_V \leq \sqrt{2} \left\{ \nu(V^2) + \mu(V^2) \right\}^{1/2} \mathrm{KL}^{1/2}(\mu|\nu) .$$

Proof. Without losing any generality, we assume that $\mu \ll \nu$. For all $t \in [0,1]$, $t\log(t) - t + 1 = \int_t^1 (u-t)u^{-1} \mathrm{d}u \geq 2^{-1}(1-t)^2$, and on $[1, +\infty)$, $t \mapsto 2(1+t)(t\log(t) - t + 1) - (1-t)^2$ is nonincreasing. Therefore, for all $t \geq 0$,

$$|1 - t| \leq (2(1+t)(t\log(t) - t + 1))^{1/2} . \tag{5.41}$$

Then, we have:

$$\begin{aligned} \|\mu - \nu\|_V &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x) \mathrm{d}\mu(x) - \int_{\mathbb{R}^d} f(x) \mathrm{d}\nu(x) \right| \\ &= \sup_{f \in \mathbb{F}(\mathbb{R}^d), \|f\|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x) \left\{ \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1 \right\} \mathrm{d}\nu(x) \right| \leq \int_{\mathbb{R}^d} V(x) \left| \frac{\mathrm{d}\mu}{\mathrm{d}\nu} - 1 \right| \mathrm{d}\nu(x) . \end{aligned}$$

Using (5.41) and the Cauchy-Schwarz inequality in the previous inequality concludes the proof. □

Proof of Theorem 5.10. First note that by the triangle inequality and Proposition 5.7, for all $p \geq 1$

$$\|\pi - \delta_x Q_\gamma^p\|_{V^{1/2}} \leq C_{1/4} \kappa^{p\gamma} V^{1/2}(x) + \|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\|_{V^{1/2}} . \tag{5.42}$$

We now bound the second term of the right hand side. Let $\mathrm{k}_\gamma = \lceil \gamma^{-1} \rceil$ and q_γ and r_γ be respectively the quotient and the remainder of the Euclidean division of p by k_γ. The triangle inequality implies $\|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\|_{V^{1/2}} \leq A + B$ with

$$\begin{aligned} A &= \left\| \delta_x Q_\gamma^{(\mathrm{q}_\gamma - 1)\mathrm{k}_\gamma} P_{\Gamma_{(\mathrm{q}_\gamma - 1)\mathrm{k}_\gamma, p}} - \delta_x Q_\gamma^{(\mathrm{q}_\gamma - 1)\mathrm{k}_\gamma} Q_\gamma^{(\mathrm{q}_\gamma - 1)\mathrm{k}_\gamma + 1, p} \right\|_{V^{1/2}} \\ B &= \sum_{i=1}^{\mathrm{q}_\gamma} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma + 1, p}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} P_{\Gamma_{i\mathrm{k}_\gamma + 1, p}} \right\|_{V^{1/2}} . \end{aligned}$$

It follows from Proposition 5.7 and $\mathrm{k}_\gamma \geq \gamma^{-1}$ that

$$B \leq \sum_{i=1}^{\mathrm{q}_\gamma} C_{1/4} \kappa^{\mathrm{q}_\gamma - i} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma + 1, i\mathrm{k}_\gamma}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} \right\|_{V^{1/2}} . \tag{5.43}$$

We now bound each term of the sum in the right hand side. For all initial distribution ν_0 on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $i, j \geq 1$, $i < j$, it follows from Lemma 5.25, [Kul97, Theorem 4.1, Chapter 2] and (5.15):

$$\begin{aligned}
&\|\nu_0 Q_\gamma^{i,j} - \nu_0 P_{\Gamma_{i,j}}\|^2_{V^{1/2}} \leq 2\left(\nu_0 Q_\gamma^{i,j}(V) + \nu_0 P_{\Gamma_{i,j}}(V)\right) \mathrm{KL}(\nu_0 Q_\gamma^{i,j} | \nu_0 P_{\Gamma_{i,j}}) \\
&\leq 2L^2 \left(\nu_0 Q_\gamma^{i,j}(V) + \nu_0 P_{\Gamma_{i,j}}(V)\right) \\
&\qquad \times (j-i)\left(\gamma^2 d + (\gamma^3/3) \sup_{k \in \{i,\cdots,j\}} \nu_0 Q_\gamma^{i,k-1}(\|\nabla U\|^2)\right) .
\end{aligned}$$

Proposition 5.7 implies by the proof of [MT93a, Theorem 6.1] that for all $y \in \mathbb{R}^d$ and $t \geq 0$: $P_t V(y) \leq V(y) + \beta_{1/2}/\theta_{1/2}$. Then, using Proposition 5.8, Lemma 5.1 and $\mathrm{k}_\gamma \geq \gamma^{-1}$ in (5.43), we get

$$\begin{aligned}
&\sup_{i \in \{1,\cdots,\mathrm{q}_\gamma\}} \left\| \delta_x Q_\gamma^{(i-1)\mathrm{k}_\gamma} P_{\Gamma_{(i-1)\mathrm{k}_\gamma+1, i\mathrm{k}_\gamma}} - \delta_x Q_\gamma^{i\mathrm{k}_\gamma} \right\|^2_{V^{1/2}} \\
&\qquad \leq 2^{-1}(1+\gamma)L^2 \left\{ 2G(\lambda, c, V(x)) + \beta_{1/2}/\theta_{1/2} \right\} \\
&\qquad\qquad \times \left\{ \gamma d + 3^{-1}\gamma^2 \|\nabla U\|^2_{V^{1/2}} G(\lambda, c, V(x)) \right\} .
\end{aligned}$$

Finally, A can be bounded along the same lines. □

5.4.7 Proof of Theorem 5.12

Denote for $\gamma > 0$, $r_\gamma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the transition density of R_γ defined for $x, y \in \mathbb{R}^d$ by

$$r_\gamma(x,y) = (4\pi\gamma)^{-1} \exp(-(4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2) . \tag{5.44}$$

For all $n \geq 1$, denote by $q_\gamma^n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ the transition density associated with Q_γ^n defined by induction by: for all $x, y \in \mathbb{R}^d$

$$q_\gamma^1(x,y) = r_{\gamma_1}(x,y) , \qquad q_\gamma^{n+1}(x,y) = \int_{\mathbb{R}^d} q_\gamma^n(x,z) r_{\gamma_{n+1}}(z,y) \mathrm{d}z \text{ for } n \geq 1 . \tag{5.45}$$

Lemma 5.26. *Assume* **L1**. *Let* $(\gamma_k)_{k \geq 1}$ *be a nonincreasing sequence with* $\gamma_1 < L$. *Then for all* $n \geq 1$ *and* $x, y \in \mathbb{R}^d$,

$$q_\gamma^n(x,y) \leq \frac{\exp\left(2^{-1}(U(x) - U(y)) - (2\sigma_{\gamma,n})^{-1} \|y - x\|^2\right)}{(2\pi\sigma_{\gamma,n} \prod_{i=1}^n (1 - L\gamma_i))^{d/2}} ,$$

where $\sigma_{\gamma,n} = \sum_{i=1}^n 2\gamma_i (1 - L\gamma_i)^{-1}$.

Proof. Under **L1**, we have for all $x, y \in \mathbb{R}^d$, $U(y) \leq U(x) + \langle \nabla U(x), y - x \rangle + (L/2) \|y - x\|^2$, which implies that for all $\gamma \in (0, L^{-1})$

$$r_\gamma(x,y) \leq (4\pi\gamma)^{-d/2} \exp\left(2^{-1}(U(x) - U(y)) - (1 - L\gamma)(4\gamma)^{-1} \|y - x\|^2\right) . \tag{5.46}$$

Then, the proof of the claimed inequality is by induction. By (5.46), the inequality holds for $n = 1$. Now assume that it holds for $n \geq 1$. By induction hypothesis and (5.46) applied for $\gamma = \gamma_{n+1}$, we have

$$\begin{aligned} q_\gamma^{n+1}(x,y) &\leq (4\pi\gamma_{n+1})^{-d/2} \left\{ 2\pi\sigma_{\gamma,n} \prod_{i=1}^{n}(1-L\gamma_i) \right\}^{-d/2} \exp\left(2^{-1}(U(x)-U(y))\right) \\ &\quad \times \int_{\mathbb{R}^d} \exp\left(-(2\sigma_{\gamma,n})^{-1}\|z-x\|^2 - (1-L\gamma_{n+1})(4\gamma_{n+1})^{-1}\|z-y\|^2\right) \mathrm{d}z \\ &\leq (4\pi\gamma_{n+1})^{-d/2} \left\{ 2\pi\sigma_{\gamma,n} \prod_{i=1}^{n}(1-L\gamma_i) \right\}^{-d/2} (\sigma_{\gamma,n}^{-1} + (1-L\gamma_{n+1})/(2\gamma_{n+1}))^{-d/2} \\ &\quad \times (2\pi)^{d/2} \exp\left(2^{-1}(U(x)-U(y)) - (2\sigma_{\gamma,n+1})^{-1}\|y-x\|^2\right) . \end{aligned}$$

Rearranging terms in the last inequality concludes the proof. $\square$

Lemma 5.27. *Assume **L** 1 and **H** 5. Then* $\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)}\mathrm{d}y \leq \vartheta_U$ *where*

$$\vartheta_U \overset{\text{def}}{=} \mathrm{e}^{a_\alpha} \frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d \boldsymbol{\Gamma}((d+1)/2)} , \tag{5.47}$$

and a_α is given in (5.18).

Proof. By Lemma 5.6, for all $x \in \mathbb{R}^d$, $U(x) \geq \rho\|x-x^\star\|/(\alpha+1) - a_\alpha$. Using the spherical coordinates, we get

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)}\mathrm{d}y \leq \mathrm{e}^{a_\alpha} \left\{ (2\pi)^{(d+1)/2}/\boldsymbol{\Gamma}((d+1)/2) \right\} \int_0^{+\infty} \mathrm{e}^{-\rho t/(\alpha+1)} t^{d-1} \mathrm{d}t .$$

Then the proof is concluded by a straightforward calculation. $\square$

Corollary 5.28. *Assume **L** 1 and **H** 5. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 < L$. Then for all $n \geq 1$ and $x \in \mathbb{R}^d$,*

$$\operatorname{Var}_\pi \left\{ \frac{\mathrm{d}\delta_x Q_\gamma^n}{\mathrm{d}\pi} \right\} \leq (\vartheta_U \exp(U(x))) \left(4\pi \left\{ \prod_{k=1}^{n}(1-L\gamma_k) \right\}^2 \sum_{i=1}^{n} \frac{\gamma_i}{1-L\gamma_i} \right)^{-d/2} ,$$

where ϑ_U is given by (5.47).

Proof of Theorem 5.12. We bound the two terms of the right hand side of (5.10). The first term is dealt with the same reasoning as for the proof of Theorem 5.9. Regarding the second term, by [Bak+08, Theorem 1.4], π satisfies a Poincaré inequality with constant $\log^{-1}(\kappa)$. Then, the claimed bound follows from (5.22) and Corollary 5.28. $\square$

5.4.8 Proof of Proposition 5.13

Set $\chi = \eta/4$ and for all $x \in \mathbb{R}^d$, $\phi(x) = (\|x - x^\star\|^2 + 1)^{1/2}$. Since ϕ is 1-Lipschitz, we have by the log-Sobolev inequality [BLM13, Theorem 5.5] for all $x \in \mathbb{R}^d$,

$$R_\gamma W_c(x) \leq e^{\chi R_\gamma \phi(x) + \chi^2 \gamma} \leq e^{\chi \sqrt{\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1} + \chi^2 \gamma} . \tag{5.48}$$

Under **L**1 since U is convex and $x^\star$ is a minimizer of U, [Nes04, Theorem 2.1.5 Equation (2.1.7)] shows that for all $x \in \mathbb{R}^d$,

$$\langle \nabla U(x), x - x^\star \rangle \geq (2L)^{-1} \|\nabla U(x)\|^2 + \eta \|x - x^\star\| \mathbb{1}_{\{\|x - x^\star\| \geq M_\eta\}} ,$$

which implies that for all $x \in \mathbb{R}^d$ and $\gamma \in (0, L^{-1}]$, we have

$$\|x - \gamma \nabla U(x) - x^\star\|^2 \leq \|x - x^\star\|^2 - 2\gamma\eta \|x - x^\star\| \mathbb{1}_{\{\|x - x^\star\| \geq M_\eta\}} . \tag{5.49}$$

Using this inequality and for all $u \in [0,1]$, $(1-u)^{1/2} - 1 \leq -u/2$, we have for all $x \in \mathbb{R}^d$, satisfying $\|x - x^\star\| \geq R_c = \max(1, 2d\eta^{-1}, M_\eta)$,

$$\begin{aligned}
&\left(\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1\right)^{1/2} - \phi(x) \\
&\quad \leq \phi(x) \left\{ \left(1 - 2\gamma \phi^{-2}(x)(\eta \|x - x^\star\| - d)\right)^{1/2} - 1 \right\} \\
&\quad \leq -\gamma \phi^{-1}(x)(\eta \|x - x^\star\| - d) \leq -(\eta\gamma/2) \|x - x^\star\| \phi^{-1}(x) \leq -2^{-3/2} \eta\gamma .
\end{aligned}$$

Combining this inequality and (5.48), we get for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq R_c$,

$$R_\gamma W_c(x) / W_c(x) \leq e^{\gamma\chi(\chi - 2^{-3/2}\eta)} = \lambda^\gamma .$$

By (5.49) and the inequality for all $a, b \geq 0$, $\sqrt{a + 1 + b} - \sqrt{1 + b} \leq a/2$, we get for all $x \in \mathbb{R}^d$,

$$\sqrt{\|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d + 1} - \phi(x) \leq \gamma d .$$

Then using this inequality in (5.48), we have for all $x \in \mathbb{R}^d$,

$$R_\gamma W_c(x) \leq \lambda^\gamma W_c(x) + \left(e^{\chi\gamma(d+\chi)} - \lambda^\gamma\right) e^{\eta(R_c^2+1)^{1/2}/4} \mathbb{1}_{B(x^\star, R_c)}(x) .$$

Using the inequality for all $t \geq 0$, $e^t - 1 \leq t e^t$ concludes the proof.

5.4.9 Proof of Corollary 5.14

We preface the proof by a Lemma.

Lemma 5.29. *Assume* ***L*** *1 and that U is convex. Let $(\gamma_k)_{k \in \mathbb{N}^*}$ be a nonincreasing sequence with $\gamma_1 \leq L^{-1}$. For all $n \geq 0$ and $x \in \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 Q_\gamma^n(x, dy) \leq \left\{ 4\eta^{-1} \left[1 + \log\left\{G(\lambda, c, \gamma_1, W_c(x))\right\}\right] \right\}^2 ,$$

where W_c, λ, c are given in (5.25) and Proposition 5.13 respectively.

Proof. Let $n \geq 0$ and $x \in \mathbb{R}^d$. Consider the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by for all $t \in \mathbb{R}$, $\phi(t) = \exp\left\{(\eta/4)(t + (4/\eta)^2)^{1/2}\right\}$. Since this function is convex on $\mathbb{R}_+$, we have by the Jensen inequality and the inequality for all $t \geq 0$, $\phi(t) \leq \mathrm{e}^{1+(\eta/4)(t+1)^{1/2}}$,

$$\phi\left(\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, Q_\gamma^n(x, \mathrm{d}y)\right) \leq \mathrm{e}^1 Q_\gamma^n W_{\mathrm{c}}(x) \; .$$

The proof is then completed using Proposition 5.13, Lemma 5.1 and that ϕ is one-to-one with for all $t \geq 1$, $\phi^{-1}(t) \leq \left(4\eta^{-1}\log(t)\right)^2$. □

Proof of Corollary 5.14. Using $\nabla U(x^\star) = 0$, **L**1 and Lemma 5.29, we have for all $k \geq 0$,

$$\int_{\mathbb{R}^d} \|\nabla U(y)\|^2 \, Q_\gamma^k(x, \mathrm{d}y) \leq L^2 \left(4\eta^{-1}\left\{1 + \log\left\{G(\lambda, c, \gamma_1, W_{\mathrm{c}}(x))\right\}\right\}\right)^2 \; .$$

□

5.4.10 Proof of Theorem 5.15

We preface the proof by a Lemma.

Lemma 5.30. *Assume **L**1 and that U is convex. Then*

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \leq \left(\frac{(2\pi)^{(d+1)/2}(d-1)!}{\eta^d \boldsymbol{\Gamma}((d+1)/2)} + \frac{\pi^{d/2} M_\eta^d}{\boldsymbol{\Gamma}(d/2+1)}\right) \; . \tag{5.50}$$

Proof. By (5.24) and $U(x^\star) = 0$, we have

$$\int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y \leq \int_{\mathbb{R}^d} \mathrm{e}^{-\eta\|y - x^\star\|} \mathrm{d}y + \int_{\mathbb{R}^d} \mathbb{1}_{\{\|y - x^\star\| \leq M_\eta\}} \mathrm{d}y \; .$$

Then the proof is concluded using the spherical coordinates. □

Proof of Theorem 5.15. By [Bob99, Theorem 1.2], π satisfies a Poincaré inequality with constant $\log^{-1}(\kappa)$. Therefore, the second term in (5.10) is dealt as in the proof of Theorem 5.12 using (5.22), Lemma 5.30 and Lemma 5.27. □

5.4.11 Proof of Proposition 5.16

For all $x \in \mathbb{R}^d$, we have

$$\begin{aligned} \mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) = \frac{\eta W_{\mathrm{c}}(x)}{4(\|x - x^\star\|^2 + 1)^{1/2}} \Big\{ &(\eta/4)(\|x - x^\star\|^2 + 1)^{-1/2} \|x - x^\star\|^2 \\ &- \langle \nabla U(x), x - x^\star \rangle - (\|x - x^\star\|^2 + 1)^{-1} \|x - x^\star\|^2 + d \Big\} \; . \end{aligned}$$

By (5.24), $\langle \nabla U(x), x - x^\star \rangle \geq \eta \|x - x^\star\|$ for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\eta$. Then, for all x, $\|x - x^\star\| \geq K = \max(M_\eta, 4d/\eta, 1)$, $\mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) \leq -(\eta^2/8) W_{\mathrm{c}}(x)$. In addition, since U is convex and $\nabla U(x^\star) = 0$, for all $x \in \mathbb{R}^d$, $\langle \nabla U(x), x - x^\star \rangle \geq 0$ and we get $\sup_{\{x \in \mathcal{E}\}} \mathscr{A}^{\mathrm{L}} W_{\mathrm{c}}(x) \leq \beta$.

5.4.12 Proof of Theorem 5.20

Proposition 5.31. *Assume **L**1, **H**6 and **H**7. Let $(\gamma_k)_{k\geq 0}$ be a sequence of nonnegative step sizes. Then for all $x \in \mathbb{R}^d$, $n \geq 0$, $p \geq 1$, $n < p$,*

$$\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq (2\pi(p-n))^{-1/2} \{\|x - x^\star\| + \mathsf{C}d\} + 2^{-1/2} L \left(\sum_{k=n}^{p-1} \left\{ (\gamma_{k+1}^3/3) A(\gamma, x) + d\gamma_{k+1}^2 \right\} \right)^{1/2} ,$$

where $A(\gamma, x)$ is given by (5.27).

Proof. Note if $(Y_t)_{t\geq 0}$ is a solution of (5.1), then $(Y_{t/2})_{t\geq 0}$ is a weak solution of the rescaled Langevin diffusion:

$$\mathrm{d}\tilde{Y}_t = -(1/2)\nabla U(\tilde{Y}_t)\mathrm{d}t + \mathrm{d}B_t^d \; . \tag{5.51}$$

Then by Proposition 5.34 and the inequality for all $s \geq 0$, $\mathbf{\Phi}(s) - 1/2 \leq (2\pi)^{-1/2} s$, we have for all $x, y \in \mathbb{R}^d$ and $t > 0$

$$\|\delta_x P_t - \delta_y P_t\|_{\mathrm{TV}} \leq \frac{\|x - y\|}{(4\pi t)^{1/2}} \; .$$

Therefore by the triangle inequality, the Cauchy-Schwarz inequality and **H**7, we have

$$\|\delta_x P_t - \pi\|_{\mathrm{TV}} \leq \frac{\|x - x^\star\| + \mathsf{C}d^{1/2}}{(4\pi t)^{1/2}} \; .$$

Then the proof follows the same lines as the proof of Proposition 5.2. □

Proof of Theorem 5.20. The proof is a straightforward consequence of Proposition 5.31. □

5.4.13 Proof of Proposition 5.21

Under **L**1, using that $\nabla U(x^\star) = 0$, we get for all $x \in \mathbb{R}^d$,

$$\begin{aligned}\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma(x, \mathrm{d}y) &= \|x - x^\star + \gamma(\nabla U(x^\star) - \nabla U(x))\|^2 + 2\gamma d \\ &\leq (1 + (L\gamma)^2) \|x - x^\star\|^2 - 2\gamma \langle \nabla U(x) - \nabla U(x^\star), x - x^\star \rangle + 2\gamma d \; . \end{aligned} \tag{5.52}$$

Then for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_{\mathrm{s}}$, we get using for all $t \geq 0$, $1 - t \leq \mathrm{e}^{-t}$

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma(x, \mathrm{d}y) \leq \lambda^\gamma \|x - x^\star\|^2 + 2\gamma d \; .$$

Using again (5.52) and the convexity of U, it yields for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \leq M_{\mathrm{s}}$,

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma(x, \mathrm{d}y) \leq \gamma \mathrm{c} \; ,$$

which concludes the proof.

5.4.14 Proof of Theorem 6.2

We preface the proof by a lemma.

Lemma 5.32. *Assume* ***H****9. Then, for all* $x \in \mathbb{R}^d$*,*

$$\|x - \gamma\nabla U(x) - x_1^\star\|^2 \leq (1 - \varpi\gamma/2)\,\|x - x_1^\star\|^2 + \gamma(\gamma + 2\varpi^{-1})\,\|\nabla U_2\|_\infty^2 \ .$$

Proof. Using that for all $y, z \in \mathbb{R}^d$, $\|y+z\|^2 \leq (1+\varpi\gamma/2)\,\|y\|^2 + (1 + 2(\varpi\gamma)^{-1})\,\|z\|^2$, we get under **H**9-(b):

$$\|x - \gamma\nabla U(x) - x_1^\star\|^2 \leq (1 + \varpi\gamma/2)\,\|x - \gamma\nabla U_1(x) - x_1^\star\|^2 \\ + \gamma(\gamma + 2\varpi^{-1})\,\|\nabla U_2\|_\infty^2 \ . \quad (5.53)$$

By [Nes04, Theorem 2.1.12, Theorem 2.1.9], **H**9-(b) implies that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U_1(y) - \nabla U_1(x), y - x\rangle \geq (\varpi/2)\,\|y - x\|^2 + \frac{1}{m + L_1}\,\|\nabla U_1(y) - \nabla U_1(x)\|^2 \ ,$$

Using this inequality and $\nabla U_1(x_1^\star) = 0$ in (5.53) concludes the proof. □

Proof of Theorem 6.2. For any $\gamma \in (0, 2/(m+L_1))$, we have for all $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2\, R_\gamma(x, \mathrm{d}y) = \|x - \gamma\nabla U(x) - x_1^\star\|^2 + 2\gamma d \\ \leq (1 - \varpi\gamma/2)\,\|x - x_1^\star\|^2 + \gamma\left\{(\gamma + 2\varpi^{-1})\,\|\nabla U_2\|_\infty^2 + 2d\right\} \ ,$$

where we have used Lemma 5.32 for the last inequality. Since $\gamma_1 \leq 2/(m+L_1)$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing, by a straightforward induction, for $p \geq 1$ and $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \|y - x_1^\star\|^2\, Q_\gamma^p(x, \mathrm{d}y) \leq \prod_{k=1}^{p} (1 - \varpi\gamma_k/2)\,\|x - x_1^\star\|^2 \\ + ((\gamma_1 + 2\varpi^{-1})\,\|\nabla U_2\|_\infty^2 + 2d) \sum_{i=n}^{p} \prod_{k=i+1}^{p} (1 - \varpi\gamma_k/2)\gamma_i \ . \quad (5.54)$$

Consider the second term in the right hand side of (5.54). Since $\gamma_1 \leq 2/(m+L_1)$, $m \leq L_1$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing, $\max_{k\geq 1}\gamma_k \leq \varpi^{-1}$ and therefore:

$$\sum_{i=n}^{p} \prod_{k=i+1}^{p} (1 - \varpi\gamma_k/2)\gamma_i \\ \leq \varpi^{-1} \sum_{i=n}^{p} \left\{ \prod_{k=i+1}^{p} (1 - \varpi\gamma_k/2) - \prod_{k=i}^{p} (1 - \varpi\gamma_k/2) \right\} \leq 2\varpi^{-1} \ .$$

□

5.4.15 Proof of Theorem 5.24

We preface the proof of the Theorem by a preliminary lemma.

Lemma 5.33. *Assume* $\boldsymbol{H}9$. *Let* $\gamma \in (0, 2/(m+L_1))$, *then for all* $x \in \mathbb{R}^d$,

$$\operatorname{Ent}_\pi\left(\frac{\mathrm{d}\delta_x R_\gamma}{\mathrm{d}\pi}\right) \leq (L_1/2)\left\{(1-\varpi\gamma/2)\left\|x-x_1^\star\right\|^2+\gamma(\gamma+2\varpi^{-1})\left\|\nabla U_2\right\|_\infty^2\right\} \\ +\operatorname{osc}_{\mathbb{R}^d}(U_2)-(d/2)(1+\log(2\gamma m)-2L_1\gamma)\ .$$

Proof. Let $\gamma \in (0, 2/(m+L_1))$ and r_γ be the transition density of R_γ given by (5.44). Under **H**9-(a) by [Nes04, Theorems 2.1.8-2.1.9], we have for all $x \in \mathbb{R}^d$,

$$U_1(x) \leq U_1(x_1^\star) + (L_1/2)\left\|x-x_1^\star\right\|^2\ . \tag{5.55}$$

Therefore we have for all $x \in \mathbb{R}^d$

$$\operatorname{Ent}_\pi\left(\frac{\mathrm{d}\delta_x R_\gamma}{\mathrm{d}\pi}\right) = \int_{\mathbb{R}^d} \log(r_\gamma(x,y)/\pi(y)) r_\gamma(x,y)\mathrm{d}x \\ \leq R_\gamma\psi(x) - (d/2)(1+\log(4\pi\gamma))\ , \tag{5.56}$$

where $\psi : \mathbb{R}^d \to \mathbb{R}$ is the function defined for all $y \in \mathbb{R}^d$ by

$$\psi(y) = U_2(y) + U_1(x_1^\star) + (L_1/2)\left\|y-x_1^\star\right\|^2 + \log\left(\int_{\mathbb{R}^d} \mathrm{e}^{-U(z)}\mathrm{d}z\right)\ .$$

By **H**9-(b) and Lemma 5.32, we get for all $x \in \mathbb{R}^d$:

$$\begin{aligned} R_\gamma\psi(x) &\leq (L_1/2)\left\|x-\gamma\nabla U(x)-x_1^\star\right\|^2 + \log\left(\int_{\mathbb{R}^d} \mathrm{e}^{-U_1(z)+U_1(x_1^\star)}\mathrm{d}z\right) \\ &\qquad + \operatorname{osc}_{\mathbb{R}^d}(U_2) + dL_1\gamma \\ &\leq (L_1/2)\left\{(1-\varpi\gamma/2)\left\|x-x_1^\star\right\|^2+\gamma(\gamma+2\varpi^{-1})\left\|\nabla U_2\right\|_\infty^2\right\} \\ &\qquad + \operatorname{osc}_{\mathbb{R}^d}(U_2) + dL_1\gamma\ . \end{aligned}$$

Plugging this bound in (5.56) gives the desired result. □

Proof of Theorem 5.24. We first deal with the second term in the right hand side of (5.10). Under **H**9, [BGL14, Corollary 5.7.2] and the Holley-Stroock perturbation principle [HS87, p. 1184] show that π satisfies a log-Sobolev inequality with constant $C_{\mathrm{LS}} = -\log^{-1}(\kappa)$. So by (5.33) we have

$$\left\|\delta_x Q_\gamma^n P_t - \pi\right\|_{\mathrm{TV}} \leq \kappa^t\left\{2\operatorname{Ent}_\pi\left(\frac{\mathrm{d}\delta_x Q_\gamma^n}{\mathrm{d}\pi}\right)\right\}^{1/2}\ .$$

We now bound $\mathrm{Ent}_\pi\left(\mathrm{d}\delta_x Q_\gamma^n/\mathrm{d}\pi\right)$ which will imply the upper bound of $C(\delta_x Q_\gamma^n)$. We proceed by induction. For $n = 1$, it is Lemma 5.33. For $n \geq 2$, by (5.45) and the Jensen inequality applied to the convex function $t \mapsto t\log(t)$, we have for all $x \in \mathbb{R}^d$ and $n \geq 1$,

$$\begin{aligned}
&\mathrm{Ent}_\pi\left(\mathrm{d}\delta_x Q_\gamma^n/\mathrm{d}\pi\right) \\
&= \int_{\mathbb{R}^d} \log\left\{\pi^{-1}(y)\int_{\mathbb{R}^d} q_\gamma^{n-1}(x,z) r_{\gamma_n}(z,y)\mathrm{d}z\right\}\int_{\mathbb{R}^d} q_\gamma^{n-1}(x,z) r_{\gamma_n}(z,y)\mathrm{d}z\mathrm{d}y \\
&\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \log\left\{r_{\gamma_n}(z,y)\pi^{-1}(y)\right\} q_\gamma^{n-1}(x,z) r_{\gamma_n}(z,y)\mathrm{d}z\mathrm{d}y \;. \qquad (5.57)
\end{aligned}$$

Using Fubini's theorem, Lemma 5.33, Theorem 6.2, and the inequality $t \geq 0$, $1-t \leq \mathrm{e}^{-t}$ in (5.57) concludes the proof of (5.35).

Finally, $A(\gamma, x)$ is bounded using the inequality for all $y, z \in \mathbb{R}^d$, $\|y+z\|^2 \leq 2(\|y\|^2 + \|z\|^2)$, **H**9 and Theorem 6.2. □

5.5 Quantitative convergence bounds in total variation for diffusions

In this part, we derived quantitative convergence results in total variation norm for d-dimensional SDEs of the form

$$\mathrm{d}\mathbf{X}_t = b(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}B_t^d \;, \qquad (5.58)$$

started at $\mathbf{X}_0$, where $(B_t^d)_{t\geq 0}$ is a d-dimensional standard Brownian motion and $b : \mathbb{R}^d \to \mathbb{R}^d$ satisfies the following assumptions.

G 1. *b is Lipschitz and for all $x, y \in \mathbb{R}^d$, $\langle b(x) - b(y), x - y\rangle \leq 0$.*

Under **G**1, [IW89, Theorems 2.4-3.1-6.1, Chapter IV] imply that there exists a unique solution $(\mathbf{X}_t)_{t\geq 0}$ to (5.58) for all initial point $x \in \mathbb{R}^d$, which is strongly Markovian. Denote by $(\mathbf{P}_t)_{t\geq 0}$ the transition semigroup associated with (5.58). To derive explicit bound for $\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}}$, we use the coupling by reflection, introduced in [LR86] to show convergence in total variation norm for solution of SDE, and recently used by [Ebe15] to obtain exponential convergence in the Wasserstein distance of order 1. This coupling is defined as (see [CL89, Example 3.7]) the unique strong Markovian process $(\mathbf{X}_t, \mathbf{Y}_t)_{t\geq 0}$ on $\mathbb{R}^{2d}$, solving the SDE:

$$\begin{cases} \mathrm{d}\mathbf{X}_t &= b(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}B_t^d \\ \mathrm{d}\mathbf{Y}_t &= b(\mathbf{Y}_t)\mathrm{d}t + (\mathrm{Id} - 2e_t e_t^T)\mathrm{d}B_t^d \;, \end{cases} \quad \text{where } e_t = e(\mathbf{X}_t - \mathbf{Y}_t) \qquad (5.59)$$

with $e(z) = z/\|z\|$ for $z \neq 0$ and $e(0) = 0$ otherwise. Define the coupling time

$$\tau_c = \inf\{s \geq 0 \mid \mathbf{X}_s \neq \mathbf{Y}_s\} \;. \qquad (5.60)$$

By construction $\mathbf{X}_t = \mathbf{Y}_t$ for $t \geq \tau_c$. We denote in the sequel by $\tilde{\mathbb{P}}_{(x,y)}$ and $\tilde{\mathbb{E}}_{(x,y)}$ the probability and the expectation associated with the SDE (5.59) started at $(x, y) \in \mathbb{R}^{2d}$ on the canonical space of continuous function from $\mathbb{R}_+$ to $\mathbb{R}^{2d}$. We denote by $(\widetilde{\mathcal{F}}_t)_{t\geq 0}$ the canonical filtration. Since $\bar{B}^d_t = \int_0^t (\mathrm{Id} - 2e_s e_s^T) \mathrm{d}B^d_s$ is a d-dimensional Brownian motion, the marginal processes $(\mathbf{X}_t)_{t\geq 0}$ and $(\mathbf{Y}_t)_{t\geq 0}$ are under $\tilde{\mathbb{P}}_{(x,y)}$ weak solutions to (5.58) started at x and y respectively. The results in [LR86] are derived under less stringent conditions than **G**1, but do not provide quantitative estimates.

Proposition 5.34 ([LR86, Example 5])**.** *Assume* ***G****1 and let* $(\mathbf{X}_t, \mathbf{Y}_t)_{t\geq 0}$ *be the solution of* (5.59)*. Then for all* $t \geq 0$ *and* $x, y \in \mathbb{R}^d$*, we have*

$$\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t) = \tilde{\mathbb{P}}_{(x,y)}(\mathbf{X}_t \neq \mathbf{Y}_t) \leq 2\left(\mathbf{\Phi}\left\{ \left(2t^{1/2}\right)^{-1} \|x - y\| \right\} - 1/2 \right) .$$

Proof. For $t < \tau_c$, $\mathbf{X}_t - \mathbf{Y}_t$ is the solution of the SDE

$$\mathrm{d}\{\mathbf{X}_t - \mathbf{Y}_t\} = \{b(\mathbf{X}_t) - b(\mathbf{Y}_t)\} \,\mathrm{d}t + 2e_t \mathrm{d}B^1_t ,$$

where $B^1_t = \int_0^t \mathbb{1}_{\{s<\tau_c\}} e_s^T \mathrm{d}B^d_s$. Using the Itô's formula and **G**1, we have for all $t < \tau_c$,

$$\|\mathbf{X}_t - \mathbf{Y}_t\| = \|x - y\| + \int_0^t \langle b(\mathbf{X}_s) - b(\mathbf{Y}_s), e_s\rangle \,\mathrm{d}s + 2B^1_t \leq \|x - y\| + 2B^1_t .$$

Therefore, for all $x, y \in \mathbb{R}^d$ and $t \geq 0$, we get

$$\begin{aligned}\tilde{\mathbb{P}}_{(x,y)}(\tau_c > t) &\leq \tilde{\mathbb{P}}_{(x,y)}\left(\min_{0\leq s\leq t} B^1_s \geq \|x - y\|/2 \right) \\ &= \tilde{\mathbb{P}}_{(x,y)}\left(\max_{0\leq s\leq t} B^1_s \leq \|x - y\|/2 \right) = \tilde{\mathbb{P}}_{(x,y)}\left(|B^1_t| \leq \|x - y\|/2 \right) ,\end{aligned}$$

where we have used the reflection principle in the last identity. □

Define for $R > 0$ the set $\Delta_R = \{x, y \in \mathbb{R}^d \mid \|x - y\| \leq R\}$. Proposition 5.34 and Lindvall's inequality give that, for all $\epsilon \in (0, 1)$ and $t \geq \omega(\epsilon, R)$,

$$\sup_{(x,y)\in\Delta_R} \|\mathbf{P}_t(x, \cdot) - \mathbf{P}_t(y, \cdot)\|_{\mathrm{TV}} \leq 2(1 - \epsilon) , \tag{5.61}$$

where ω is defined in (5.29). To obtain quantitative exponential bounds in total variation for any $x, y \in \mathbb{R}^d$, it is required to control some exponential moments of the successive return times to Δ_R. This is first achieved by using a drift condition for the generator $\mathscr{A}$ associated with the SDE (5.58) defined for all $f \in C^2(\mathbb{R}^d)$ by

$$\mathscr{A} f = \langle b, \nabla f\rangle + (1/2)\Delta f .$$

Consider the following assumption:

G 2. *(i) There exist a twice continuously differentiable function $V : \mathbb{R}^d \mapsto [1, \infty)$ and constants $\theta > 0$, $\beta \geq 0$ such that*

$$\mathscr{A} V \leq -\theta V + \beta \; . \tag{5.62}$$

(ii) There exists $\delta > 0$ and $\mathrm{R} > 0$ such that $\Theta \subset \Delta_{\mathrm{R}}$ where

$$\Theta = \{(x, y) \in \mathbb{R}^{2d} \mid V(x) + V(y) \leq 2\theta^{-1}\beta + \delta\} \; . \tag{5.63}$$

For $t > 0$, and G a closed subset of $\mathbb{R}^{2d}$, define by $\mathrm{T}_1^{G,t}$ the first return time to G delayed by t:

$$\mathrm{T}_1^{G,t} = \inf\{s \geq t \mid (\mathbf{X}_s, \mathbf{Y}_s) \in G\} \; .$$

For $j \geq 2$, define recursively the j-th return time to G delayed by t by

$$\mathrm{T}_j^{G,t} = \inf\{s \geq \mathrm{T}_{j-1}^{G,t} + t \mid (\mathbf{X}_s, \mathbf{Y}_s) \in G\} = \mathrm{T}_{j-1}^{G,t} + \mathrm{T}_1^{G,t} \circ \mathrm{S}_{\mathrm{T}_{j-1}^{G,t}} \; , \tag{5.64}$$

where S is the shift operator on the canonical space. By [EK86, Proposition 1.5 Chapter 2], the sequence $(\mathrm{T}_j^{G,t})_{j\geq 1}$ is a sequence of stopping time with respect to $(\tilde{\mathcal{F}}_t)_{t\geq 0}$.

Proposition 5.35. *Assume **G** 1 and **G** 2. For all $x, y \in \mathbb{R}^d$, $\epsilon \in (0, 1)$ and $j \geq 1$, we have*

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}\mathrm{T}_j^{\Theta,\omega(\epsilon,\mathrm{R})}}\right] \leq \{\mathrm{K}(\epsilon)\}^{j-1}\left\{(1/2)(V(x) + V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} \; ,$$

$$\tilde{\theta} = \theta^2\delta(2\beta + \theta\delta)^{-1} \; , \; \mathrm{K}(\epsilon) = \tilde{\theta}^{-1}\beta\left(1 + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\right) + \delta/2 \; , \tag{5.65}$$

where ω is defined in (5.29).

Proof. For notational simplicity, set $\mathrm{T}_j = \mathrm{T}_j^{\Theta,\omega(\epsilon,\mathrm{R})}$. Note that for all $x, y \in \mathbb{R}^d$,

$$\mathscr{A}V(x) + \mathscr{A}V(y) \leq -\tilde{\theta}(V(x) + V(y)) + 2\beta\mathbb{1}_{\Theta}(x, y) \; .$$

Then by the Dynkin formula (see e.g. [MT93b, Eq. (8)]) the process

$$t \mapsto (1/2)\mathrm{e}^{\tilde{\theta}(\mathrm{T}_1\wedge t)}\left\{V\left(\mathbf{X}_{\mathrm{T}_1\wedge t}\right) + V\left(\mathbf{Y}_{\mathrm{T}_1\wedge t}\right)\right\} \; , \quad t \geq \omega(\epsilon, \mathrm{R}) \; ,$$

is a positive supermartingale. Using the optional stopping theorem and the Markov property, we have, using that for all $t \geq 0$ $\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}t}V(X_t)\right] \leq V(x) + \beta\tilde{\theta}^{-1}\mathrm{e}^{\tilde{\theta}t}$,

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{\tilde{\theta}\mathrm{T}_1}\right] \leq (1/2)(V(x) + V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta \; .$$

The result then follows from this inequality and the strong Markov property. □

Theorem 5.36. *Assume* **G***1 and* **G***2. Then for all $\epsilon \in (0,1)$, $t \geq 0$ and $x, y \in \mathbb{R}^d$,*

$$\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}} \leq \mathrm{e}^{-\tilde{\theta}t/2}\left\{(1/2)(V(x)+V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} + 2\kappa^t \ ,$$

where ω is defined in (5.29), $\tilde{\theta}, \mathrm{K}(\epsilon)$ in (5.65) and

$$\log(\kappa) = (\tilde{\theta}/2)\log(1-\epsilon)\{\log(\mathrm{K}(\epsilon)) - \log(1-\epsilon)\}^{-1} \ .$$

Proof. Let $x, y \in \mathbb{R}^d$ and $t \geq 0$. For all $\ell \geq 1$ and $\epsilon \in (0,1)$,

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t\right) \leq \tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t, \mathrm{T}_\ell \leq t\right) + \tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_\ell > t\right) \ , \tag{5.66}$$

where $\mathrm{T}_\ell = \mathrm{T}_\ell^{\Theta,\omega(\epsilon,\mathrm{R})}$. We now bound the two terms in the right hand side of this equation. For the first term, since $\Theta \subset \Delta_{\mathrm{R}}$, by (5.61), we have conditioning successively on $\widetilde{\mathcal{F}}_{\mathrm{T}_j}$, for $j = \ell, \ldots, 1$, and using the strong Markov property,

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t, \mathrm{T}_\ell \leq t\right) \leq (1-\epsilon)^\ell \ . \tag{5.67}$$

For the second term, using Proposition 5.35 and the Markov inequality, we get

$$\begin{aligned}\tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_\ell > t\right) &\leq \tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_1 > t/2\right) + \tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_\ell - \mathrm{T}_1 > t/2\right) \\ &\leq \mathrm{e}^{-\tilde{\theta}t/2}\left\{(1/2)(V(x)+V(y)) + \mathrm{e}^{\tilde{\theta}\omega(\epsilon,\mathrm{R})}\tilde{\theta}^{-1}\beta\right\} + \mathrm{e}^{-\tilde{\theta}t/2}\{\mathrm{K}(\epsilon)\}^{\ell-1} \ .\end{aligned}$$

The proof is completed combining this inequality and (5.67) in (5.66) and taking $\ell = \left\lceil 2^{-1}\tilde{\theta}t/(\log(\mathrm{K}(\epsilon)) - \log(1-\epsilon))\right\rceil$. □

More precise bounds can be obtained under more stringent assumption on the drift b; see [BGG12] and [Ebe15].

G 3. *There exist $\tilde{M}_{\mathrm{s}} \geq 1$ and $\tilde{m}_{\mathrm{s}} > 0$, such that for all $x, y \in \mathbb{R}^d$, $\|x-y\| \geq \tilde{M}_{\mathrm{s}}$,*

$$\langle b(x) - b(y), x - y\rangle \leq -\tilde{m}_{\mathrm{s}}\|x-y\|^2 \ .$$

Proposition 5.37. *Assume* **G***1 and* **G***3.*

(a) For all $x, y \in \mathbb{R}^d$ and $\epsilon \in (0,1)$

$$\tilde{\mathbb{E}}_{(x,y)}\left[\exp\left(\frac{\tilde{m}_{\mathrm{s}}}{2}\left(\tau_c \wedge \mathrm{T}_1^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}\right)\right)\right] \leq 1 + \|x-y\| + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2} \ .$$

(b) For all $x, y \in \mathbb{R}^d$, $\epsilon \in (0,1)$ and $j \geq 1$

$$\begin{aligned}\tilde{\mathbb{E}}_{(x,y)}&\left[\exp\left((\tilde{m}_{\mathrm{s}}/2)\left(\tau_c \wedge \mathrm{T}_j^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}\right)\right)\right] \\ &\leq \{\mathrm{D}(\epsilon)\}^{j-1}\left\{1 + \|x-y\| + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2}\right\} \ ,\end{aligned}$$

$$\mathrm{D}(\epsilon) = (1 + \mathrm{e}^{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})/2})(1+\tilde{M}_{\mathrm{s}}) \ , \tag{5.68}$$

where ω is given in (5.29).

Proof. In the proof, we set $\mathrm{T}_j = \mathrm{T}_j^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}$.

(a) Consider the sequence of increasing stopping time

$$\tau_k = \inf\{t > 0 \mid \|\mathbf{X}_t - \mathbf{Y}_t\| \notin \left[k^{-1}, k\right]\} , \quad k \geq 1 ,$$

and set $\zeta_k = \tau_k \wedge \mathrm{T}_1$. We derive a bound on $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)\zeta_k\}]$ independent on k. Since $\lim_{k\to+\infty} \uparrow \tau_k = \tau_c$ almost surely, the monotone convergence theorem implies that the same bound holds for $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)(\tau_c \wedge \mathrm{T}_1)\}]$. Set now $\mathrm{W}_{\mathrm{s}}(x,y) = 1 + \|x-y\|$. Since $\mathrm{W}_{\mathrm{s}} \geq 1$ and $\tau_c < \infty$ a.s by Proposition 5.34, it suffices to give a bound on $\tilde{\mathbb{E}}_{(x,y)}[\exp\{(\tilde{m}_{\mathrm{s}}/2)\zeta_k\}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})]$. By Itô's formula, we have for all $v, t \leq \tau_c$, $v \leq t$

$$\begin{aligned} \mathrm{e}^{\tilde{m}_{\mathrm{s}} t/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_t, \mathbf{Y}_t) = \mathrm{e}^{\tilde{m}_{\mathrm{s}} v/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_v, \mathbf{Y}_v) + (\tilde{m}_{\mathrm{s}}/2) \int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}} u/2}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_u, \mathbf{Y}_u)\mathrm{d}u \\ + \int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}} u/2} \langle b(\mathbf{X}_u) - b(\mathbf{Y}_u), e_u\rangle \,\mathrm{d}u + 2\int_v^t \mathrm{e}^{\tilde{m}_{\mathrm{s}} u/2}\mathrm{d}B_u^1 . \end{aligned} \tag{5.69}$$

Using $\mathbf{G}3(b)$, we have for all $k \geq 1$ and $t_{\mathrm{s}} = \omega(\epsilon, \tilde{M}_{\mathrm{s}}) \leq v \leq t$

$$\begin{aligned} \mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\zeta_k\wedge t)}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge t}, \mathbf{Y}_{\zeta_k\wedge t}) \leq \mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\zeta_k\wedge v)}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge v}, \mathbf{Y}_{\zeta_k\wedge v}) \\ + 2\int_{\zeta_k\wedge v}^{\zeta_k\wedge t} \mathrm{e}^{\tilde{m}_{\mathrm{s}} u/2}\mathrm{d}B_u^1 . \end{aligned}$$

So the process

$$\{\exp\left((\tilde{m}_{\mathrm{s}}/2)(\zeta_k\wedge t)\right)\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k\wedge t}, \mathbf{Y}_{\zeta_k\wedge t})\}_{t\geq t_{\mathrm{s}}} ,$$

is a positive supermartingale and by the optional stopping theorem, we get

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)\zeta_k}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})\right] \leq \tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\tau_k\wedge t_{\mathrm{s}})}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\tau_k\wedge t_{\mathrm{s}}}, \mathbf{Y}_{\tau_k\wedge t_{\mathrm{s}}})\right] , \tag{5.70}$$

where we used that $\zeta_k \wedge t_{\mathrm{s}} = \tau_k \wedge t_{\mathrm{s}}$. By (5.69), $\mathbf{G}1$ and $\mathbf{G}3$, we have

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\tau_k\wedge t_{\mathrm{s}})}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\tau_k\wedge t_{\mathrm{s}}}, \mathbf{Y}_{\tau_k\wedge t_{\mathrm{s}}})\right] \leq \mathrm{W}_{\mathrm{s}}(x,y) + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}} t_{\mathrm{s}}/2} ,$$

and (5.70) becomes

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)\zeta_k}\mathrm{W}_{\mathrm{s}}(\mathbf{X}_{\zeta_k}, \mathbf{Y}_{\zeta_k})\right] \leq \mathrm{W}_{\mathrm{s}}(x,y) + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\tilde{m}_{\mathrm{s}} t_{\mathrm{s}}/2} .$$

(b) The proof is by induction. The case $j = 1$ has been established above. Now let $j \geq 2$. Since on the event $\{\tau_c > \mathrm{T}_{j-1}\}$, we have

$$\tau_c \wedge \mathrm{T}_j = \mathrm{T}_{j-1} + (\tau_c \wedge \mathrm{T}_1) \circ \mathrm{S}_{\mathrm{T}_{j-1}} ,$$

where S is the shift operator, we have conditioning on $\tilde{\mathcal{F}}_{\mathrm{T}_{j-1}}$, using the strong Markov property, Proposition 5.34 and the first part,

$$\tilde{\mathbb{E}}_{(x,y)}\left[\mathbb{1}_{\{\tau_c > \mathrm{T}_{j-1}\}}\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)(\tau_c\wedge \mathrm{T}_j)}\right] \leq \mathrm{D}(\epsilon)\,\tilde{\mathbb{E}}_{(x,y)}\left[\mathbb{1}_{\{\tau_c > \mathrm{T}_{j-1}\}}\mathrm{e}^{(\tilde{m}_{\mathrm{s}}/2)\mathrm{T}_{j-1}}\right] .$$

Then the proof follows since $\mathrm{D}(\epsilon) \geq 1$.

□

Theorem 5.38. *Assume **G**1 and **G**3. Then for all $\epsilon \in (0,1)$, $t \geq 0$ and $x, y \in \mathbb{R}^d$,*

$$\|\mathbf{P}_t(x,\cdot) - \mathbf{P}_t(y,\cdot)\|_{\mathrm{TV}} \leq \left\{(1-\epsilon)^{-1} + 1 + \|x-y\|\right\} \kappa^t$$

$$\log(\kappa) = (\tilde{m}_{\mathrm{s}}/2)\log(1-\epsilon)(\log(\mathrm{D}(\epsilon)) - \log(1-\epsilon))^{-1} ,$$

where $\mathrm{D}(\epsilon)$ is defined in (5.68).

Proof. The proof is along the same lines as Theorem 5.36. Set $\mathrm{T}_j = \mathrm{T}_j^{\Delta_{\tilde{M}_{\mathrm{s}}},\omega(\epsilon,\tilde{M}_{\mathrm{s}})}$ for $j \geq 1$. Let $x, y \in \mathbb{R}^d$ and $t \geq 0$. For all $\ell \geq 1$ and $\epsilon \in (0,1)$,

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t\right) \leq \tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t, \mathrm{T}_\ell \leq t\right) + \tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_\ell \wedge \tau_c > t\right) . \tag{5.71}$$

For the first term, by (5.61) we have conditioning successively on $\tilde{\mathcal{F}}_{\mathrm{T}_j}$, for $j = \ell, \cdots, 1$, and using the strong Markov property,

$$\tilde{\mathbb{P}}_{(x,y)}\left(\tau_c > t, \mathrm{T}_\ell \leq t\right) \leq (1-\epsilon)^\ell . \tag{5.72}$$

For the second term, using Proposition 5.37-(b) and the Markov inequality, we get

$$\tilde{\mathbb{P}}_{(x,y)}\left(\mathrm{T}_\ell \wedge \tau_c > t\right) \leq \mathrm{e}^{-\frac{\tilde{m}_{\mathrm{s}} t}{2}} \{\mathrm{D}(\epsilon)\}^{\ell-1} \left\{1 + \|x-y\| + (1+\tilde{M}_{\mathrm{s}})\mathrm{e}^{\frac{\tilde{m}_{\mathrm{s}}\omega(\epsilon,\tilde{M}_{\mathrm{s}})}{2}}\right\} . \tag{5.73}$$

Taking $\ell = \lfloor (\tilde{m}_{\mathrm{s}} t/2)/(\log(\mathrm{D}(\epsilon)) - \log(1-\epsilon)) \rfloor$ and combining (5.72)-(5.73) in (5.71) concludes the proof. □

5.5.1 Proof of Theorem 5.17 and Theorem 5.22

Recall that $(P_t)_{t\geq 0}$ is the Markov semigroup of the Langevin equation associated with U and let $\mathscr{A}^{\mathrm{L}}$ be the corresponding generator. Since $(P_t)_{t\geq 0}$ is reversible with respect to π, we deduce from Theorem 5.36 and Theorem 5.38 quantitative bounds for the exponential convergence of $(P_t)_{t\geq 0}$ to π in total variation noting that if $(Y_t)_{t\geq 0}$ is a solution of (5.1), then $(Y_{t/2})_{t\geq 0}$ is a weak solution of the rescaled Langevin diffusion:

$$\mathrm{d}\tilde{Y}_t = -(1/2)\nabla U(\tilde{Y}_t)\mathrm{d}t + \mathrm{d}B_t^d . \tag{5.74}$$

Proof of Theorem 5.17. Since the generator associated with the SDE (5.74) is $(1/2)\mathscr{A}^{\mathrm{L}}$, Proposition 5.16 shows that (5.62) holds for W_{c} with constants $\theta/2$ and $\beta/2$. Using that for all $a_1, a_2 \in \mathbb{R}$, $\mathrm{e}^{(a_1+a_2)/2} \leq (1/2)(\mathrm{e}^{a_1} + \mathrm{e}^{a_2})$, **G**2-(ii) holds for $\delta = 2\theta^{-1}\beta$ and $\mathrm{R} = (8/\eta)\log(4\theta^{-1}\beta)$. By Theorem 5.36 with $\epsilon = 1/2$, we get for all $x, y \in \mathbb{R}^d$ and $t \geq 0$

$$\begin{aligned}\|P_t(x,\cdot) - P_t(y,\cdot)\|_{\mathrm{TV}} &\leq 2\varpi^t \\ &+ \mathrm{e}^{-\theta t/4}\left\{(1/2)(W_{\mathrm{c}}(x) + W_{\mathrm{c}}(y)) + 2\theta^{-1}\beta \mathrm{e}^{4\theta^{-1}\omega(2^{-1},(8/\eta)\log(4\theta^{-1}\beta))}\right\} ,\end{aligned} \tag{5.75}$$

where ϖ is defined in (5.30a). By [MT93b, Theorem 4.3-(ii)], (5.62) implies that $\int_{\mathbb{R}^d} W_{\mathrm{c}}(y)\pi(\mathrm{d}y) \leq \beta\theta^{-1}$. The proof is then concluded using this bound, (5.75) and that π is invariant for $(P_t)_{t\geq 0}$ □

Proof of Theorem 5.22. By applying Theorem 5.38 with $\epsilon = 1/2$, the triangle inequality and using that π is invariant for $(P_t)_{t \geq 0}$, we have

$$\|P_t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \left\{3 + \|x - x^\star\| + \int_{\mathbb{R}^d} \|y - x^\star\| \, \mathrm{d}\pi(y)\right\} \kappa^t \ .$$

It remains to show that $\int_{\mathbb{R}^d} \|y - x^\star\| \, \mathrm{d}\pi(y) \leq (d/m + M_{\mathrm{s}}^2)^{1/2}$. For this, we establish a drift inequality for the generator $\mathscr{A}^{\mathrm{L}}$ of the Langevin SDE associated with U. Consider the function $W_{\mathrm{s}}(x) = \|x - x^\star\|^2$. For all $x \in \mathbb{R}^d$, we have using $\nabla U(x^\star) = 0$,

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq 2(d - \langle \nabla U(x) - \nabla U(x^\star), x - x^\star\rangle) \ .$$

Therefore by **G**3, for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_{\mathrm{s}}$, we get

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq -2m W_{\mathrm{s}}(x) + 2d \ ,$$

and for all $x \in \mathbb{R}^d$,

$$\mathscr{A}^{\mathrm{L}} W_{\mathrm{s}}(x) \leq -2m W_{\mathrm{s}}(x) + 2(d + m M_{\mathrm{s}}^2) \ .$$

By [MT93b, Theorem 4.3-(ii)], we get $\int_{\mathbb{R}^d} W_{\mathrm{s}}(y) \mathrm{d}\pi(y) \leq d/m + M_{\mathrm{s}}^2$. The bound on $C(\delta_x Q_\gamma^n)$ is a consequence of the Cauchy-Schwarz inequality, Proposition 5.21 and Lemma 5.1. The bound for $A(\gamma, x)$ similarly follows from **L**1, Proposition 5.21 and Lemma 5.1.

□

Acknowledgements

The authors are indebted to Arnaud Guillin for sharing his knowledge of Poincaré and log-Sobolev inequalities. The authors are grateful to Andreas Eberle for very careful readings and many useful comments. The author thank the anonymous referees for their constructive feedback. The work of A.D. and E.M. is supported by the Agence Nationale de la Recherche, under grant ANR-14-CE23-0012 (COSMOS).

Chapter 6

Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm

Alain Durmus[1], Éric Moulines [2]

Abstract

We consider in this chapter the problem of sampling a probability distribution π having a density w.r.t. the Lebesgue measure on $\mathbb{R}^d$, known up to a normalisation factor $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$. Under the assumption that U is continuously differentiable, ∇U is globally Lipschitz and U is strongly convex, we obtain non-asymptotic bounds for the convergence to stationarity in Wasserstein distances and total variation distance of the sampling method based on the Euler discretization of the Langevin stochastic differential equation, for both constant and decreasing step sizes. The dependence on the dimension of the state space of the obtained bounds is studied to demonstrate the applicability of this method in the high dimensional setting. The convergence of an appropriately weighted empirical measure is also investigated and bounds for the mean square error and exponential deviation inequality are reported for functions which are either Lipchitz continuous or measurable and bounded. Some numerical results are presented to illustrate our findings.

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

6.1 Introduction

Let π be a probability distribution on $\mathbb{R}^d$, $d \geq 1$, with density $x \mapsto \mathrm{e}^{-U(x)} / \int_{\mathbb{R}^d} \mathrm{e}^{-U(y)} \mathrm{d}y$ w.r.t. the Lebesgue measure, where U is continuously differentiable, gradient Lipshitz and strongly convex. Consider the Langevin stochastic differential equation associated with π:

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \;, \tag{6.1}$$

where $(B_t)_{t \geq 0}$ is a d-dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, satisfying the usual conditions. Under the stated assumptions on U, π satisfies a log-Sobolev inequalities (see [BCG08; CG09; BGL14]) and the Markov semi-group associated with the Langevin diffusion $(Y_t)_{t \geq 0}$ converges exponentially fast to π with a rate independent of the dimension of the state space. We study in this paper the sampling method based on the Euler-Maruyama discretization scheme associated to the Langevin diffusion, which defines a (possibly) non-homogeneous, discrete-time Markov chain given by

$$X_{k+1} = X_k - \gamma_{k+1} \nabla U(X_k) + \sqrt{2\gamma_{k+1}} Z_{k+1} \tag{6.2}$$

where $(Z_k)_{k \geq 1}$ is an i.i.d. sequence of standard Gaussian random variables and $(\gamma_k)_{k \geq 1}$ is a sequence of stepsizes, which can either be held constant or be chosen to decrease to 0.

This method was originally proposed in the physics literature by [Par81] and introduced in the computational statistics community by [Gre83] and [GM94]. It has been studied in depth by [RT96a], which proposed to use a Metropolis-Hastings step at each iteration to enforce reversibility w.r.t. π leading to the Metropolis Adjusted Langevin Algorithm (MALA). They coin the term *unadjusted* Langevin algorithm (ULA) to stress the fact that the Metropolis-Hastings step is avoided.

We obtain in this paper non-asymptotic and computable bounds between the marginal laws of the Markov chain $(X_n)_{n \geq 0}$ defined by the Euler discretization and the target distribution π in Wasserstein distance and total variation distance for nonincreasing step sizes. When the sequence of step sizes is constant $\gamma_k = \gamma$ for all $k \geq 0$, the Markov chain $(X_n)_{n \geq 0}$ has a unique stationary distribution π_γ (see [RT96a]), which in most of the cases differs from the distribution π. Quantitative estimates between π and π_γ is obtained. When $(\gamma_k)_{k \geq 1}$ decreases to zero and $\sum_{k=1}^{\infty} \gamma_k = \infty$ then we show that the marginal distribution of the non-homogeneous Markov chain $(X_n)_{n \geq 0}$ converges to the target distribution π with explicit expression for the convergence rate.

The paper is organized as follows. In Section 6.2, we study the convergence in the Wasserstein distance of order 2 of the Euler discretization for constant and decreasing stepsizes. In Section 6.3 we provide non-asymptotic bounds of convergence of the weighted empirical measure applied to Lipschitz functions. In Section 6.4, we give non asymptotic bounds in total variation distance between the Euler discretization and π. This study is completed in Section 6.5 by non-asymptotic bounds of convergence of the weighted empirical measure applied to bounded and measurable functions. In Section 6.6 the convergence in the Wasserstein distance of order p, for $p \geq 2$ is investegated. Some numerical illustrations are given Section 6.7 to support our claims. The proofs are given in Section 6.8. Finally in Section 6.9, some results of independent interest, used in some proofs, on functional autoregressive models are gathered. Some technical derivations are carried out in a supplementary paper. [DM15d].

Notations and conventions

Denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$, $\mathbb{F}(\mathbb{R}^d)$ the set of all Borel measurable functions on $\mathbb{R}^d$ and for $f \in \mathbb{F}(\mathbb{R}^d)$, $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. For μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $f \in \mathbb{F}(\mathbb{R}^d)$ a μ-integrable function, denote by $\mu(f)$ the integral of f w.r.t. μ. We say that ζ is a transference plan of μ and ν if it is a probability measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ such that for all measurable set A of $\mathbb{R}^d$, $\zeta(\mathrm{A} \times \mathbb{R}^d) = \mu(\mathrm{A})$ and $\zeta(\mathbb{R}^d \times \mathrm{A}) = \nu(\mathrm{A})$. We denote by $\Pi(\mu, \nu)$ the set of transference plans of μ and ν. Furthermore, we say that a couple of $\mathbb{R}^d$-random variables (X, Y) is a coupling of μ and ν if there exists $\zeta \in \Pi(\mu, \nu)$ such that (X, Y) are distributed according to ζ. For two probability measures μ and ν, we define the Wasserstein distance of order $p \geq 1$ as

$$W_p(\mu, \nu) \stackrel{\text{def}}{=} \left(\inf_{\zeta\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\zeta(x, y) \right)^{1/p} .$$

By [Vil09, Theorem 4.1], for all μ, ν probability measures on $\mathbb{R}^d$, there exists a transference plan $\zeta^\star \in \Pi(\mu, \nu)$ such that for any coupling (X, Y) distributed according to $\zeta^\star$, $W_p(\mu, \nu) = \mathbb{E}[\|X - Y\|^p]^{1/p}$. This kind of transference plan (respectively coupling) will be called an optimal transference plan (respectively optimal coupling) associated with W_p. We denote by $\mathcal{P}_p(\mathbb{R}^d)$ the set of probability measures with finite p-moment: for all $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} \|x\|^p \, \mu(\mathrm{d}x) < +\infty$. By [Vil09, Theorem 6.16], $\mathcal{P}_p(\mathbb{R}^d)$ equipped with the Wasserstein distance W_p of order p is a complete separable metric space.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function, namely there exists $C \geq 0$ such that for all $x, y \in \mathbb{R}^d$, $|f(x) - f(y)| \leq C \|x - y\|$. Then we denote

$$\|f\|_{\mathrm{Lip}} = \inf\{|f(x) - f(y)| \, \|x - y\|^{-1} \mid x, y \in \mathbb{R}^d, x \neq y\} .$$

The Monge-Kantorovich theorem (see [Vil09, Theorem 5.9]) implies that for all μ, ν probabilities measure on $\mathbb{R}^d$,

$$W_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}^d} f(x)\mu(\mathrm{d}x) - \int_{\mathbb{R}^d} f(x)\nu(\mathrm{d}x) \mid f : \mathbb{R}^d \to \mathbb{R} \; ; \; \|f\|_{\mathrm{Lip}} \leq 1 \right\} . \tag{6.3}$$

Denote by $\mathbb{F}_b(\mathbb{R}^d)$ the set of all bounded Borel measurable functions on $\mathbb{R}^d$. For $f \in \mathbb{F}_b(\mathbb{R}^d)$ set $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. For two probability measures μ and ν on $\mathbb{R}^d$, the total variation distance distance between μ and ν is defined by

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{\mathrm{A}\in\mathcal{B}(\mathbb{R}^d)} |\mu(\mathrm{A}) - \nu(\mathrm{A})| = (1/2) \sup_{\substack{f\in\mathbb{F}_b(\mathbb{R}^d) \\ \|f\|_\infty\leq 1}} |\mu(f) - \nu(f)| .$$

By the Monge-Kantorovich theorem the total variation distance between μ and ν can be written on the form:

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf_{\zeta \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_{\mathsf{D}}(x,y) \mathrm{d}\zeta(x,y) ,$$

where $\mathsf{D} = \{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \,|x = y\}$. By [Vil09, Theorem 4.1], for all μ, ν probability measures on $\mathbb{R}^d$, there exists a transference plan $\zeta^\star \in \Pi(\mu,\nu)$ such that for any coupling (X,Y) distributed according to $\zeta^\star$, $\|\mu - \nu\|_{\mathrm{TV}} = \mathbb{P}(X \neq Y)$. This kind of transference plan (respectively coupling) will be called an optimal transference plan (respectively optimal coupling) associated with the total variation distance. For all $x \in \mathbb{R}^d$ and $M > 0$, we denote by $\mathrm{B}(x,M)$, the ball centered at x of radius M. For a subset $\mathrm{A} \subset \mathbb{R}^d$, denote by A^{c} the complementary of A. Let $n, m \in \mathbb{N}^*$ and M be a $n \times n$-matrix, then denote by M^{T} the transpose of M and $\|M\|$ the Frobenius associated with M defined by $\|M\| = \mathrm{Tr}(M^{\mathrm{T}} M)$. Let $n, m \in \mathbb{N}^*$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be a twice continuously differentiable function. Denote by ∇F and $\nabla^2 F$ the Jacobian and the Hessian of F respectively. Denote also by $\vec{\Delta} F$ the vectorial Laplacian of F defined by: for all $x \in \mathbb{R}^d$, $\vec{\Delta} F(x)$ is the vector of $\mathbb{R}^m$ such that for all $i \in \{1, \cdots, m\}$, the i-th component of $\vec{\Delta} F(x)$ is equals to $\sum_{j=1}^d (\partial^2 F_i / \partial x_j^2)(x)$. In the sequel, we take the convention that for $n, p \in \mathbb{N}$, $n < p$ then $\sum_p^n = 0$ and $\prod_p^n = 1$.

6.2 Non-asymptotic bounds in Wasserstein distance of order 2 for ULA

Consider the following assumption on the potential U:

H 10. *The function U is continuously differentiable on $\mathbb{R}^d$ and is gradient Lipschitz, i.e. there exists $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$\|\nabla U(x) - \nabla U(y)\| \leq L \|x - y\| .$$

Under **H** 10, if μ_0 is a probability measure satisfying $\int \|x\|^2 \mu_0(\mathrm{d}x) < \infty$ then by [KS91, Theorem 2.5, Theorem 2.9 Chapter 5] there exists a unique strong solution $(Y_t)_{t \geq 0}$ to (6.1) with initial distribution μ_0. Denote by $(P_t)_{t \geq 0}$ the semi-group associated with (6.1), which is reversible w.r.t. π, and hence admits π as its (unique) invariant measure.

H 11. *U is strongly convex, i.e. there exists $m > 0$ such that for all $x, y \in \mathbb{R}^d$,*

$$U(y) \geq U(x) + \langle \nabla U(x), y - x \rangle + (m/2) \|x - y\|^2 .$$

Under **H** 11, [Nes04, Theorem 2.1.8] shows that U has a unique minimizer $x^\star \in \mathbb{R}^d$. If in addition **H** 10 holds, then [Nes04, Theorem 2.1.12, Theorem 2.1.9] shows that for all $x, y \in \mathbb{R}^d$:

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq (\kappa/2) \|y - x\|^2 + \frac{1}{m + L} \|\nabla U(y) - \nabla U(x)\|^2 , \tag{6.4}$$

$$\langle \nabla U(y) - \nabla U(x), y - x \rangle \geq m \|y - x\|^2 , \tag{6.5}$$

where

$$\kappa = \frac{2mL}{m+L} \; . \tag{6.6}$$

Note that **H** 10 and (6.5) imply that $L \geq m$. We first obtain the geometric rate of convergence to stationarity of the semi-group in Wasserstein distance. It is worthwhile to note that these bounds do not depend on the dimension d.

Theorem 6.1. *Assume **H** 10 and **H** 11.*

(i) For all $p \geq 2$, probability measures μ and $\nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $t \geq 0$,

$$W_p(\mu P_t, \nu P_t) \leq \mathrm{e}^{-mt} W_p(\mu, \nu)$$

(ii) The stationary distribution π satisfies

$$\int_{\mathbb{R}^d} \|x - x^\star\|^2 \, \pi(\mathrm{d}x) \leq d/m \; . \tag{6.7}$$

Proof. Most of the statement is well known; see [BGG12] and the references therein. Nevertheless for completeness, we provide the proof in Section 6.8.1. □

Let $(\gamma_k)_{k\geq 1}$ be a sequence of positive and non-increasing step sizes and for $n, p \in \mathbb{N}$, denote by

$$\Gamma_{n,p} \stackrel{\text{def}}{=} \sum_{k=n}^{p} \gamma_k \; , \qquad \Gamma_n = \Gamma_{1,n} \; . \tag{6.8}$$

For $\gamma > 0$, consider the Markov kernel R_γ given for all $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$R_\gamma(x, \mathrm{A}) = \int_{\mathrm{A}} (4\pi\gamma)^{-d/2} \exp\left(-(4\gamma)^{-1} \|y - x + \gamma \nabla U(x)\|^2\right) \mathrm{d}y \; . \tag{6.9}$$

Under **H** 10 R_γ is strongly Feller, irreducible, strongly aperiodic. The sequence $(X_n)_{n\geq 0}$ given in (6.2) is a Markov chain with respect to the sequence of Markov kernels $(R_{\gamma_n})_{n\geq 1}$. For $p, n \geq 1$, $p \geq n$, define

$$Q_\gamma^{n,p} = R_{\gamma_n} \cdots R_{\gamma_p} \; , \qquad Q_\gamma^n = Q_\gamma^{1,n} \tag{6.10}$$

with the convention that for $n, p \geq 0$, $n < p$, $Q_\gamma^{p,n}$ is the identity operator. The stability of the Euler discretization of a one-dimensional Langevin diffusion with constant step size has been studied in [RT96a, Section 3]; We generalize these results to multidimensional diffusions and decreasing stepsizes.

Theorem 6.2. *Assume **H** 10 and **H** 11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $x^\star$ be the unique minimizer of U. Then for all $x \in \mathbb{R}^d$ and $n, p \in \mathbb{N}^*$*

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, Q_\gamma^{n,p}(x, \mathrm{d}y) \leq \varrho_{n,p}(x) \; ,$$

where $\varrho_{n,p}(x)$ is given by

$$\varrho_{n,p}(x) = \prod_{k=n}^{p}(1-\kappa\gamma_k)\left\|x-x^\star\right\|^2 + 2d\sum_{k=n}^{p}\prod_{i=k+1}^{p}(1-\kappa\gamma_i) \;, \tag{6.11}$$

and κ is defined in (6.6).

Proof. The proof is postponed to Section 6.8.2. □

Theorem 6.3. *Assume* ***H**10* *and* ***H**11*. *For any $\gamma \in (0, 2/(m+L))$, R_γ has a unique stationary distribution π_γ. Moreover, for all $p \geq 1$, $\pi_\gamma \in \mathcal{P}_p(\mathbb{R}^d)$ and for all probability measure $\mu \in \mathcal{P}_{2p}(\mathbb{R}^d)$, we have for all $n \geq 0$:*

$$W_{2p}(\mu R_\gamma^n, \pi_\gamma) \leq (1-\kappa\gamma)^{np} W_{2p}(\mu, \pi_\gamma) \;. \tag{6.12}$$

Proof. The proof is postponed to Section 6.8.3. □

We now proceed to establish explicit bounds for $W_2(\mu_0 Q_\gamma^n, \pi)$, with $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Since π is invariant for P_t for all $t \geq 0$, it suffices to get some bounds on $W_2(\mu_0 Q_\gamma^n, \nu_0 P_{\Gamma_n})$, with $\nu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and take $\nu_0 = \pi$. To do so, we construct a coupling between the diffusion and the linear interpolation of the Euler discretization. In the strongly convex case, an obvious candidate is the synchronous coupling $(Y_t, \overline{Y}_t)_{t\geq 0}$ for all $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1})$ by

$$\begin{cases} Y_t = Y_{\Gamma_n} - \int_{\Gamma_n}^t \nabla U(Y_s)\mathrm{d}s + \sqrt{2}(B_t - B_{\Gamma_n}) \\ \bar{Y}_t = \bar{Y}_{\Gamma_n} - \nabla U(\bar{Y}_{\Gamma_n})(t-\Gamma_n) + \sqrt{2}(B_t - B_{\Gamma_n}) \;, \end{cases} \tag{6.13}$$

where $(\Gamma_n)_{n\geq 1}$ is given in (6.8). Therefore since for all $n \geq 0$, $W_2^2(\mu_0 P_{\Gamma_n}, \nu_0 Q_\gamma^n) \leq \mathbb{E}[\|Y_{\Gamma_n} - \bar{Y}_{\Gamma_n}\|^2]$, where μ_0 and ν_0 are the marginals of ζ_0, we compute an explicit bound of the Wasserstein distance between the sequence of distributions $(\mu_0 Q_\gamma^n)_{n\geq 0}$ and the stationary measure π of the Langevin diffusion (6.1).

Theorem 6.4. *Assume* ***H**10* *and* ***H**11*. *Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $n \geq 1$,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(2)}(\gamma) \;,$$

where

$$u_n^{(1)}(\gamma) \stackrel{\text{def}}{=} \prod_{k=1}^{n}(1-\kappa\gamma_k/2) \tag{6.14}$$

and

$$u_n^{(2)}(\gamma) \stackrel{\text{def}}{=} L^2 \sum_{i=1}^{n} \gamma_i^2 \left\{\kappa^{-1} + \gamma_i\right\} (2d + dL^2\gamma_i/m + dL^2\gamma_i^2/6) \prod_{k=i+1}^{n}(1-\kappa\gamma_k/2) \;, \tag{6.15}$$

where κ is defined in (6.6) .

Proof. The proof is postponed to Section 6.8.4. □

We now consider stepsizes which goes to 0. Under this additional assumption, we may establish the convergence of the sequence $(\mu_0 Q_\gamma^n)_{n\geq 0}$ to π.

Corollary 6.5. *Assume* ***H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Assume that $\lim_{k\to\infty}\gamma_k = 0$ and $\lim_{n\to+\infty}\Gamma_n = +\infty$. Then for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$,*

$$\lim_{n\to\infty} W_2(\mu_0 Q_\gamma^n, \pi) = 0 \ .$$

Proof. The proof is postponed to Section 6.8.5. □

In the case of constant stepsizes $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 6.4, a bound between π and the stationary distribution π_γ of R_γ.

Corollary 6.6. *Assume* ***H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a constant sequence $\gamma_k = \gamma$ for all $k \geq 1$ with $\gamma \leq 1/(m+L)$. Then*

$$W_2^2(\pi, \pi_\gamma) \leq 2\kappa^{-1} L^2 \gamma \left\{\kappa^{-1} + \gamma\right\} (2d + dL^2\gamma/m + dL^2\gamma^2/6) \ .$$

Proof. The proof is postponed to Section 6.8.6. □

We can improve these bound under additional regularity assumptions on the potential U.

H12. *The potential U is three times continuously differentiable and there exists $\tilde{L}$ such that for all $x, y \in \mathbb{R}^d$:*

$$\left\|\nabla^2 U(x) - \nabla^2 U(y)\right\| \leq \tilde{L} \left\|x - y\right\| \ . \tag{6.16}$$

Note that under **H**10 and **H**12, we have that for all $x, y \in \mathbb{R}^d$,

$$\left\|\nabla^2 U(x) y\right\| \leq L \left\|y\right\| \ , \ \left\|\vec{\Delta}(\nabla U)(x)\right\|^2 \leq d\tilde{L}^2 \ . \tag{6.17}$$

Theorem 6.7. *Assume* ***H**10, **H**11 and **H**12. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $n \geq 1$,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + u_n^{(3)}(\gamma) \ ,$$

where $u_n^{(1)}$ is given by (6.14) and

$$u_n^{(3)}(\gamma) \overset{\text{def}}{=} \sum_{i=1}^n d\gamma_i^3 \left\{2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma_i L^4 + 4L^4/(3m)) + \gamma_i L^4(\gamma_i/6 + m^{-1})\right\} \prod_{k=i+1}^n (1-\kappa\gamma_k/2) \ , \tag{6.18}$$

where κ is defined in (6.6) .

Proof. The proof is postponed to Section 6.8.7. □

In the case of constant stepsizes $\gamma_k = \gamma$ for all $k \geq 1$, we can deduce from Theorem 6.7, a sharper bound between π and the stationary distribution π_γ of R_γ.

Corollary 6.8. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a constant sequence $\gamma_k = \gamma$ for all $k \geq 1$ with $\gamma \leq 1/(m+L)$. Then*

$$W_2^2(\pi, \pi_\gamma) \leq 2\kappa^{-1} d\gamma^2 \left\{ 2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma L^4 + 4L^4/(3m)) + \gamma L^4(\gamma/6 + m^{-1}) \right\} .$$

Proof. The proof follows the same line as the proof of Corollary 6.6 and is omitted. □

Let $x^\star$ be the unique minimizer of U. Since for all $y \in \mathbb{R}^d$ $\|x - y\|^2 \leq 2(\|x - x^\star\|^2 + \|x^\star - y\|^2)$, using (6.7), we get:

$$W_2^2(\delta_x, \pi) \leq 2(\|x - x^\star\|^2 + d/m) . \tag{6.19}$$

If $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, we have $W_2^2(\mu_0, \pi) \leq \int \mu_0(\mathrm{d}x) W_2^2(\delta_x, \pi)$. Hence, the bounds provided by Theorem 6.4 and Theorem 6.7 scale linearly with the dimension d. When $\gamma_k = \gamma$ for all $k \geq 1$, (6.14), (6.15) (6.18) imply

$$\begin{cases} u_n^{(1)}(\gamma) = (1 - \kappa\gamma/2)^n , \quad u_n^{(2)}(\gamma) \leq 2\kappa^{-1}\gamma \left\{\kappa^{-1} + \gamma\right\} (2d + dL^2\gamma/m + dL^2\gamma^2/6) , \\ u_n^{(3)}(\gamma) \leq 2\kappa^{-1} d\gamma^2 \left\{ 2L^2 + \kappa^{-1}(\tilde{L}^2/3 + \gamma L^4 + 4L^4/(3m)) + \gamma L^4(\gamma/6 + m^{-1}) \right\} . \end{cases} \tag{6.20}$$

Using this bound, given $\epsilon > 0$, we may determine the smallest number of iterations and an associated step-size γ, starting from x, to approach the stationary distribution in the Wasserstein distance $W_2(\delta_x Q_n^\gamma, \pi)$ with a precision ϵ. Details and further discussions are included in the supplementary paper [DM15d].

Based on Theorem 6.4 and Theorem 6.7, we can obtain explicit bounds for $W_2^2(\delta_x Q_\gamma^n, \pi)$ for all $x \in \mathbb{R}^d$. For simplicity, we consider sequences $(\gamma_k)_{k\geq 1}$ defined for all $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$, for $\gamma_1 < 1/(m+L)$ and $\alpha \in (0,1]$. The order of these bounds is given in Table 6.1 and Table 6.2, see [DM15d, Section 1-2] for details. Two regimes can be observed as in stochastic approximation in the case of Theorem 6.4.

	$\alpha \in (0,1)$	$\alpha = 1$
Order of convergence	$\mathcal{O}(dn^{-\alpha})$	$\mathcal{O}(dn^{-1})$ for $\gamma_1 > 2\kappa^{-1}$ see [DM15d, Section 3]

Table 6.1: Order of convergence of $W_2^2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**10 and **H**11

	$\alpha \in (0,1)$
Order of convergence	$\mathcal{O}(dn^{-2\alpha})$

Table 6.2: Order of convergence of $W_2^2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**10, **H**11 and **H** 12

We now consider the fixed horizon setting. Assuming here that the step sizes $(\gamma_k)_{k\geq 1}$ are defined for $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$ for $\alpha \in [0,1)$, we determine the value of γ_1

	Optimal choice of γ_1	Bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$
$\alpha \in [0,1)$	$\mathcal{O}(n^{\alpha-1}\log(n))$	$\mathcal{O}(dn^{-1}\log(n))$

Table 6.3: Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 6.4

	Optimal choice of γ_1	Bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$
$\alpha \in [0,1)$	$\mathcal{O}(n^{\alpha-1}\log(n))$	$\mathcal{O}(dn^{-2}\log^2(n))$

Table 6.4: Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 6.7

minimizing the upper bound $u_n^{(1)}(\gamma)W_2^2(\mu_0, \pi) + u_n^{(2)}(\gamma)$. The results are summarized in Table 6.3, see [DM15d, Section 1-2] for details.

Moreover, these bounds for a fixed number of iterations implies using the doubling trick (see [HK14]) an anytime algorithm which guarantees for all $n \geq 1$ and $x \in \mathbb{R}^d$ that $W_2(\delta_x Q_\gamma^n, \pi)$ is $\mathcal{O}((\log(n)n^{-1})^{1/2})$ or $\mathcal{O}((\log(n)n^{-1})$.

6.3 Mean square error and concentration for Lipschitz functions

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function and $(X_n)_{n\geq 0}$ the Euler discretization of the Langevin diffusion. In this section we study the approximation of $\int_{\mathbb{R}^d} f(y)\pi(\mathrm{d}y)$ by the weighted average estimator

$$\hat{\pi}_n^N(f) = \sum_{k=N+1}^{N+n} \omega_{k,n}^N f(X_k) \; , \quad \omega_{k,n}^N = \gamma_{k+1}\Gamma_{N+2,N+n+1}^{-1} \; . \tag{6.21}$$

where $N \geq 0$ is the length of the burn-in period, $n \geq 1$ is the number of samples, and for $n, p \in \mathbb{N}$, $\Gamma_{n,p}$ is given by (6.8). In all this section, $\mathbb{P}_x$ and $\mathbb{E}_x$ denote the probability and the expectation respectively, induced on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}})$ by the Markov chain $(X_n)_{n\geq 0}$ started at $x \in \mathbb{R}^d$. We first compute an explicit bound for the Mean Squared Error (MSE) of this estimator defined by:

$$\mathrm{MSE}_f(N,n) = \mathbb{E}_x\left[\left|\hat{\pi}_n^N(f) - \pi(f)\right|^2\right] = \left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 + \mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} \; . \tag{6.22}$$

We first obtain an elementary bound for the bias. For all $k \in \{N+1, \ldots, N+n\}$, let ξ_k be the optimal transference plan between $\delta_x Q_\gamma^k$ and π for W_2. Then by the Jensen inequality and because f is Lipschitz, we have:

$$\left(\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right)^2 = \left(\sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d \times \mathbb{R}^d} \{f(z) - f(y)\} \xi_k(\mathrm{d}z, \mathrm{d}y)\right)^2$$
$$\leq \|f\|_{\mathrm{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d \times \mathbb{R}^d} \|z-y\|^2 \, \xi_k(\mathrm{d}z, \mathrm{d}y) \; .$$

Using Theorem 6.4, we end up with the following bound.

Proposition 6.9. *Assume* ***H****10 and* ***H****11. Let* $(\gamma_k)_{k \geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 1/(m+L)$. *Let* $x^\star$ *be the unique minimizer of* U. *Let* $(X_n)_{n \geq 0}$ *be given by* (6.2) *and started at* $x \in \mathbb{R}^d$. *Then for all* $n, N \geq 0$ *and Lipschitz function* $f : \mathbb{R}^d \to \mathbb{R}$:

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq \|f\|_{\mathrm{Lip}}^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \left\{2(\|x - x^\star\|^2 + d/m) u_k^{(1)}(\gamma) + w_k(\gamma)\right\} \; ,$$

where $u_n^{(1)}(\gamma)$ *is given in* (6.14) *and* $w_n(\gamma)$ *is equal to* $u_n^{(2)}(\gamma)$ *defined by* (6.15) *and to* $u_n^{(3)}(\gamma)$, *defined by* (6.18), *if* ***H****12 holds.*

Consider now the variance term. To control this term, we adapt the proof of [JO10, Theorem 2] for homogeneous Markov chain to our inhomogeneous setting, and we have:

Theorem 6.10. *Assume* ***H****10 and* ***H****11. Let* $(\gamma_k)_{k \geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$. *Then for all* $N \geq 0$, $n \geq 1$ *and Lipschitz functions* $f : \mathbb{R}^d \to \mathbb{R}$, *we get*

$$\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} \leq 8\kappa^{-2} \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1} v_{N,n}(\gamma) \; ,$$

where

$$v_{N,n}(\gamma) \stackrel{def}{=} \left\{1 + \Gamma_{N+2,N+n+1}^{-1}(\kappa^{-1} + 2/(m+L))\right\} \; . \tag{6.23}$$

Proof. The proof is postponed to Section 6.8.8. □

It is worth to observe that this bound is independent from the dimension. We may now discuss the bounds on the MSE (obtained by combining the bounds for the squared bias Proposition 6.9 and the variance Theorem 6.10) for step sizes given for $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$ where $\alpha \in [0,1]$ and $\gamma_1 < 1/(m+L)$. Details of these calculations are included in the supplementary paper [DM15d, Section 5]. The order of the bounds (up to numerical constants) of the MSE are summarized in Table 6.5 as a function of γ_1, n and N. If the total number of iterations $n+N$ is held fixed (fixed horizon setting), as in Section 6.2, we may optimize the value of the step size γ_1 but also of the burn-in period N to minimize the upper bound of the MSE. The order (in n) for different values of $\alpha \in [0,1]$ are summarized in Table 6.8 and Table 6.7 (we display the order in n but not the constants, which are quite involved and not overly informative).

We observe two differents bounds based on Theorem 6.4 and Theorem 6.7. Let us discuss first, the bounds obtained by the last one. It appears that, for any $\alpha \in [0, 1/3)$,

	Bound for the MSE
$\alpha = 0$	$\gamma_1 + (\gamma_1 n)^{-1} \exp(-\kappa\gamma_1 N/2)$
$\alpha \in (0, 1/2)$	$\gamma_1 n^{-\alpha} + (\gamma_1 n^{1-\alpha})^{-1} \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))$
$\alpha = 1/2$	$\gamma_1 \log(n) n^{-1/2} + (\gamma_1 n^{1/2})^{-1} \exp(-\kappa\gamma_1 N^{1/2}/4)$
$\alpha \in (1/2, 1)$	$n^{\alpha-1} \left\{\gamma_1 + \gamma_1^{-1} \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))\right\}$
$\alpha = 1$	$\log(n)^{-1} \left\{\gamma_1 + \gamma_1^{-1} N^{-\gamma_1\kappa/2}\right\}$

Table 6.5: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**10 and **H**11

	Bound for the MSE
$\alpha = 0$	$\gamma_1^2 + (\gamma_1 n)^{-1} \exp(-\kappa\gamma_1 N/2)$
$\alpha \in (0, 1/3)$	$\gamma_1^2 n^{-2\alpha} + (\gamma_1 n^{1-\alpha})^{-1} \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))$
$\alpha = 1/3$	$\gamma_1^2 \log(n) n^{-2/3} + (\gamma_1 n^{2/3})^{-1} \exp(-\kappa\gamma_1 N^{1/2}/4)$
$\alpha \in (1/3, 1)$	$n^{\alpha-1} \left\{\gamma_1^2 + \gamma_1^{-1} \exp(-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha)))\right\}$
$\alpha = 1$	$\log(n)^{-1} \left\{\gamma_1^2 + \gamma_1^{-1} N^{-\gamma_1\kappa/2}\right\}$

Table 6.6: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**10, **H**11 and **H**12

we can always achieved the order $n^{-2/3}$ by choosing appropriately γ_1 and N (for $\alpha = 1/3$ we have only $\log^{1/3}(n)n^{-2/3}$). The worst case is for $\alpha \in (1/3, 1]$, where in fact the best strategy is to take $N = 0$ and the largest possible value for $\gamma_1 = 1/(m + L)$. Finally, we note that from the explicit expression of the bound in [DM15d, Section 5.2], that constant step sizes ($\alpha = 0$) are optimal. Finally, we mention that the bounds for $\alpha \in [0, 1/2)$ for a fixed number of iterations implies using the doubling trick (see [HK14]) an anytime algorithm which guarantees for all $n \geq 1$, a MSE of order $\mathcal{O}(n^{-2/3})$.

	Optimal choice of γ_1	Optimal choice of N	Bound for the MSE
$\alpha = 0$	$n^{-1/3}$	$n^{1/3}$	$n^{-2/3}$
$\alpha \in (0, 1/2)$	$n^{\alpha-1/3}$	$n^{(1/3-\alpha)/(1-\alpha)}$	$n^{-2/3}$
$\alpha = 1/2$	$(\log(n))^{-1/3}$	$\log^{1/2}(n)$	$\log^{1/3}(n)n^{-2/3}$
$\alpha \in (1/2, 1)$	$1/(m + L)$	0	$n^{1-\alpha}$
$\alpha = 1$	$1/(m + L)$	0	$\log(n)$

Table 6.7: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**10, **H**11 and **H**12

Now let us discuss the bounds based on Theorem 6.4. This time for any $\alpha \in [0, 1/2)$, we can always achieved the order $n^{-1/2}$ by choosing appropriately γ_1 and N (for $\alpha = 1/2$ we have only $\log(n)n^{-1/2}$). For $\alpha \in (1/2, 1]$, the best strategy is to take $N = 0$ and the largest possible value for $\gamma_1 = 1/(m + L)$. Finally, we note that from the explicit expression of the bound in [DM15d], that constant step sizes ($\alpha = 0$) are again optimal.

	Optimal choice of γ_1	Optimal choice of N	Bound for the MSE
$\alpha = 0$	$n^{-1/2}$	$n^{1/2}$	$n^{-1/2}$
$\alpha \in (0, 1/2)$	$n^{\alpha-1/2}$	$n^{(1/2-\alpha)/(1-\alpha)}$	$n^{-1/2}$
$\alpha = 1/2$	$(\log(n))^{-1/2}$	$\log(n)$	$\log(n) n^{-1/2}$
$\alpha \in (1/2, 1)$	$1/(m + L)$	0	$n^{1-\alpha}$
$\alpha = 1$	$1/(m + L)$	0	$\log(n)$

Table 6.8: Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**10 and **H**11

We can also follow the proof of [JO10, Theorem 5] to establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (6.21)

Theorem 6.11. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $(X_n)_{n\geq 0}$ be given by (6.2) and started at $x \in \mathbb{R}^d$. Then for all $N \geq 0$, $n \geq 1$, $r > 0$ and Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$:*

$$\mathbb{P}_x\left[\hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r\right] \leq \exp\left(-\frac{r^2\kappa^2\Gamma_{N+2,N+n+1}}{16\left\|f\right\|_{\mathrm{Lip}}^2 \upsilon_{N,n}(\gamma)}\right) .$$

Proof. Using the Markov inequality and Proposition 6.35, for all $\lambda > 0$, we have:

$$\mathbb{P}_x\left[\hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r\right] \leq \exp\left(-\lambda r + 4\kappa^{-2}\lambda^2 \left\|f\right\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1}\upsilon_{N,n}(\gamma)\right) .$$

Then the result follows from taking $\lambda = (r\kappa^2\Gamma_{N+2,N+n+1})/(8\left\|f\right\|_{\mathrm{Lip}}^2 \upsilon_{N,n}(\gamma))$. □

If we apply this result to the sequence $(\gamma_k)_{k\geq 1}$ defined for all $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$, for $\alpha \in [0, 1]$, we end up with a concentration of order $\exp(-n^{1-\alpha})$ for $\alpha \in [0, 1)$ and n^{-1} for $\alpha = 1$.

6.4 Quantitative bounds in total variation distance

We deal in this section with quantitative bounds in total variation distance. For Bayesian inference application, this kinds of bounds are of utmost interest for computing highest posterior density (HPD) credible regions and intervals. For computing such bounds we will use the results of Section 6.2 combined with the regularizing property of the semigroup $(P_t)_{t\geq 0}$.

Proposition 6.12 ([LR86, Example 5], [DM15a, Proposition 29])**.** *Assume **H**10 and **H** 11. Then for all $t > 0$, $x, y \in \mathbb{R}^d$ we have*

$$\left\|P_t(x, \cdot) - P_t(y, \cdot)\right\|_{\mathrm{TV}} \leq 1 - 2\boldsymbol{\Phi}\left(-\left\|x - y\right\|/(2\sqrt{2t})\right) .$$

Since for all $s > 0$, $1/2 - \boldsymbol{\Phi}(-s) \leq (2\pi)^{-1/2}s$, Proposition 6.12 implies that for all bounded measurable functions $f : \mathbb{R}^d \to \mathbb{R}$, and for all $t > 0$, $x, y \in \mathbb{R}^d$ we have

$$|P_t f(x) - P_t f(y)| \leq (4\pi t)^{-1/2}\|f\|_\infty \left\|x - y\right\| . \tag{6.24}$$

Therefore for all $t > 0$, $P_t f$ is a Lipschitz function with $\|P_t f\|_{\mathrm{Lip}} \leq (4\pi t)^{-1/2}\|f\|_\infty$. We can combine this result and Theorem 6.1 or Theorem 6.7 to get explicit bound in total variation between the Euler-Maruyama discretization and the target distribution π. For this we will always use the following decomposition, for all nonincreasing sequence $(\gamma_k)_{k\geq 0}$, initial point $x \in \mathbb{R}^d$ and $n \geq 0$,

$$\|\pi - Q_\gamma^n\|_{\mathrm{TV}} \leq \|\pi - P_{\gamma n}\|_{\mathrm{TV}} + \|P_{\gamma n} - Q_\gamma^n\|_{\mathrm{TV}} \ . \tag{6.25}$$

The first term is dealt with the following result.

Theorem 6.13. *Assume* ***H**10* *and* ***H**11*. *Then for all $x \in \mathbb{R}^d$ and $t \geq 0$,*

$$\|\pi - \delta_x P_t\|_{\mathrm{TV}} \leq \left\{5 + (d/m+1)^{1/2} + \|x - x^\star\|\right\} \exp\left(-\tau t\right) \ ,$$

where

$$\tau = m\log(2)\left\{2\left(\log\left[1 + \exp\left\{m/\left(2\mathbf{\Phi}^{-1}(3/4)\right)^2\right\}\right] + 2\log(2)\right)\right\}^{-1} \ . \tag{6.26}$$

Proof. The proof is an application of [DM15a, Theorem 18]. □

It remains to bound the second term in (6.25). Since we will use Theorem 6.1 and Theorem 6.7, we have two different results depending on the assumptions on U that we have again. Define for all $x \in \mathbb{R}^d$ and $n, p \in \mathbb{N}$,

$$\vartheta_{n,p}^{(1)}(x) = L^2 \sum_{i=1}^n \gamma_i^2 \prod_{k=i+1}^n (1 - \kappa\gamma_k/2)\left[\left\{\kappa^{-1} + \gamma_i\right\}(2d + dL^2\gamma_i^2/6) + L^2\delta_i\gamma_i\left\{\kappa^{-1} + \gamma_i\right\}\right] \tag{6.27}$$

$$\vartheta_{n,p}^{(2)}(x) = \sum_{i=1}^n \gamma_i^3 \prod_{k=i+1}^n (1 - \kappa\gamma_k/2)\left\{d(2L^2 + 4\kappa^{-1}(\tilde{L}^2/12 + \gamma_{n+1}L^4/4) + \gamma_{n+1}^2 L^4/6)\right. \tag{6.28}$$

$$\left. + \, L^4\delta_i(4\kappa^{-1}/3 + \gamma_{n+1})\right\} \ ,$$

where

$$\delta_i = \mathrm{e}^{-2m\Gamma_{i-1}}\varrho_{n,p}(x) + (1 - \mathrm{e}^{-2m\Gamma_{i-1}})(d/m) \ ,$$

and $\varrho_{n,p}(x)$ is given by (6.11). For $n \in \mathbb{N}_-$, we take the convention that for all $p \in \mathbb{N}$ and $x \in \mathbb{R}^d$, $\vartheta_{n,p}^{(1)}(x)$ and $\vartheta_{n,p}^{(2)}(x)$ are equal to 0.

Theorem 6.14. *Assume* ***H**10* *and* ***H**11*. *Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x \in \mathbb{R}^d$ and $p, n \in \mathbb{N}$, $p > n$,*

$$\|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\|_{\mathrm{TV}} \leq (\vartheta_n(x)/(4\pi\Gamma_{n+1,p}))^{1/2} + 2^{-3/2}L\left(\sum_{k=n+1}^p \left\{(\gamma_k^3 L^2/3)\varrho_{1,k-1}(x) + d\gamma_k^2\right\}\right)^{1/2} \ , \tag{6.29}$$

where $\varrho_{1,n}(x)$ is defined by (6.11), $\vartheta_n(x)$ is equal to $\vartheta^{(2)}_{n,0}(x)$ given by (6.28), if **H** *12 holds, and to $\vartheta^{(1)}_{n,0}(x)$ given by (6.27) otherwise.*

Proof. The proof is postponed to Section 6.8.10 □

Consider the case of decreasing step sizes defined for $k \geq 1$ by $\gamma_k = \gamma_1 k^{-\alpha}$ for $\alpha \in (0,1]$. If **H** 12 holds, choosing $n = p - \lfloor p^{\alpha/2} \rfloor$ in the bound given by (6.29) and using Table 6.2, (6.25) and Theorem 6.13 implies that $\lim_{p\to+\infty} \|\delta_x Q^p_\gamma - \pi\|_{\mathrm{TV}} = 0$ at a rate $\mathcal{O}(d^{1/2}p^{-3/4\alpha})$. If **H** 12 does not hold, choosing $n = p - \lfloor p^{\alpha} \rfloor$ in the bound given by Theorem 6.14 and using Table 6.1 implies that $\lim_{p\to+\infty} \|\delta_x Q^p_\gamma - \pi\|_{\mathrm{TV}} = 0$ at a rate $\mathcal{O}(dp^{-\alpha/2})$. Note that for $\alpha = 1$, this rate is true only for $\gamma_1 > 2\kappa^{-1}$. These conclusions and the dependency on the dimension are summarized in Table 6.9. At fixed

	H 10, **H** 11 and **H** 12	**H** 10, **H** 11 and **H** 12
Order for $\gamma_k = \gamma_1 k^{-\alpha}$, $\alpha \in (0,1]$	$\mathcal{O}(d^{1/2}n^{-\alpha/2})$	$\mathcal{O}(d^{1/2}n^{-3\alpha/4})$

Table 6.9: Order of convergence of $\|\delta_x Q^n_\gamma - \pi\|_{\mathrm{TV}}$ for $\gamma_k = \gamma_1 k^{-\alpha}$ based on Theorem 6.14

step size $\gamma_k = \gamma \in (0, 1/(m+L))$ for all $k \geq 1$, under **H** 12 choosing $n = p - \lfloor \gamma^{-1/2} \rfloor$ in (6.29), and combining this result with Theorem 6.13 in (6.25) implies a bound of order $\mathcal{O}(d^{1/2}\gamma^{3/4})$ for all $p > \lfloor \gamma^{-1/2} \rfloor$, since there exists $C \geq 0$ independent of d, γ and x such that $\vartheta^{(2)}_{n,0}(x) \leq Cd\gamma^2 \|x - x^\star\|^2$ for all $n \geq 0$. If **H** 12 does not hold, choosing $n = p - \lfloor \gamma^{-1} \rfloor$ implies a bound of order $\mathcal{O}((d\gamma)^{1/2})$ for all $p > \lfloor \gamma^{-1} \rfloor$, since there exists $C \geq 0$ independent of d, γ and x such that $\vartheta^{(1)}_{n,0}(x) \leq Cd\gamma \|x - x^\star\|^2$. For constant step sizes, we can improve this bound under **H** 12. Define for all $\gamma > 0$, the function $\mathsf{n} : \mathbb{R}^*_+ \to \mathbb{N}$ by

$$\mathsf{n}(\gamma) = \left\lceil \log\left(\left\lceil \gamma^{-1} \right\rceil \right) / \log(2) \right\rceil . \tag{6.30}$$

Theorem 6.15. *Assume* **H** *10 and* **H** *11. Let $\gamma \in (0, 1/(m+L))$. Then for all $x \in \mathbb{R}^d$ and $p \in \mathbb{N}^*$,*

$$\|\delta_x P_{p\gamma} - \delta_x R^p_\gamma\|_{\mathrm{TV}} \leq (\vartheta_{p-2^{\mathsf{n}(\gamma)},0}(x)/(\pi 2^{\mathsf{n}(\gamma)+2}\gamma))^{1/2}$$
$$+ 2^{-3/2} L \left\{ (\gamma^3 L^2/3)\varrho_{1,p-1}(x) + d\gamma^2 \right\}^{1/2} + \sum_{k=1}^{\mathsf{n}(\gamma)} (\vartheta_{2^{k-1},p-2^k}(x)/(\pi 2^{k+1}\gamma))^{1/2} . \tag{6.31}$$

where for all $n_1, n_2 \in \mathbb{N}$, ϑ_{n_1,n_2} is equal to $\vartheta^{(2)}_{n_1,n_2}$ given by (6.28), if **H** *12 holds, and to $\vartheta^{(1)}_{n_1,n_2}$ given by (6.27) otherwise.*

Proof. The proof is postponed to Section 6.8.11. □

If **H** 12 holds, (6.31) implies a bound of order $\mathcal{O}(d^{1/2}\gamma\log(\gamma^{-1}))$ for all $p \in \mathbb{N}^*$, since there exists $C \geq 0$ independent of d, γ and x such that $\vartheta^{(2)}_{n_1,n_2}(x) \leq CdgaStep^3 n_1 \|x - x^\star\|^2$

for all $n_1, n_2 \geq 0$. If **H**12 does not hold, (6.31) implies a bound of order $\mathcal{O}((d\gamma)^{1/2}\log(\gamma^{-1}))$, since there exists $C \geq 0$ independent of d, γ and x such that $\vartheta^{(1)}_{n_1,n_2}(x) \leq C d\gamma^2 n_1(1 + \|x - x^\star\|^2)$. Therefore in this case, the bound provided by Theorem 6.15 does not improve the dependency of the dimension provided by Theorem 6.14 and is even worst.

By the discussion on Theorem 6.14 and Theorem 6.15, we can conclude that there exists $C \geq 0$ independent of the dimension d and γ such that $\|\pi - \pi_\gamma\|_{\mathrm{TV}} \leq C d^{1/2}\gamma\log(\gamma^{-1})$ if **H** 12 holds and $\|\pi - \pi_\gamma\|_{\mathrm{TV}} \leq C(\gamma d)^{1/2}$ otherwise. We can also for a precision target $\varepsilon > 0$ say how to choose $\gamma > 0$ and the number of iterations $p > 0$ to get $\|\delta_x R^p_\gamma - \pi\|_{\mathrm{TV}} \leq \varepsilon$. If **H**12 holds, the number of iterations has to be of order $d^{1/2}\log^2(d)$ and $d\log(d)$ otherwise. This discussion is summarized in Table 6.10 and Table 6.11.

	H10, **H**11	**H**10, **H**11 and **H**12
$\|\pi - \pi_\gamma\|_{\mathrm{TV}}$	$\mathcal{O}((\gamma d)^{1/2})$	$\mathcal{O}(d^{1/2}\gamma\log(\gamma^{-1}))$

Table 6.10: Order of the bound between π and π_γ in total variation function of the step size $\gamma > 0$ and the dimension d.

	H10, **H**11	**H**10, **H**11 and **H**12
γ	$\mathcal{O}(d^{-1}\varepsilon^2)$	$\mathcal{O}(d^{-1/2}\log^{-1}(d)\varepsilon\log^{-1}(\varepsilon^{-1}))$
p	$\mathcal{O}(d\log(d))$	$\mathcal{O}(d^{1/2}\log^2(d)\varepsilon^{-1}\log^2(\varepsilon^{-1}))$

Table 6.11: Order of the step size $\gamma > 0$ and the number of iterations $p \in \mathbb{N}^*$ to get $\|\delta_x R^p_\gamma - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$

6.5 Mean square error and concentration for bounded measurable functions

The result of the previous section allows us to study the approximation of $\int_{\mathbb{R}^d} f(y)\pi(\mathrm{d}y)$ by the weighted average estimator $\hat{\pi}^N_n(f)$ defined by (6.21) for a measurable and bounded function $f : \mathbb{R}^d \to \mathbb{R}$. In all this section, $\mathbb{P}_x$ and $\mathbb{E}_x$ denote the probability and the expectation respectively, induced on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}})$ by the Markov chain $(X_n)_{n\geq 0}$ started at $x \in \mathbb{R}^d$, defined by the Euler discretization (6.2). As in Section 6.3. We first obtain an elementary bound for the bias term in (6.22). For all $k \in \{N+1, \ldots, N+n\}$, let ξ_k be the optimal transference plan between $\delta_x Q^k_\gamma$ and π for the total variation distance.

Then by the Jensen inequality and because f is bounded, we have:

$$
\begin{aligned}
\left(\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right)^2 &= \left(\sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \{f(z)-f(y)\}\xi_k(\mathrm{d}z,\mathrm{d}y)\right)^2 \\
&\leq 4\left\|f\right\|_\infty^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathbb{1}_{\mathsf{D}}(z,y)\xi_k(\mathrm{d}z,\mathrm{d}y) \\
&\leq 4\left\|f\right\|_\infty^2 \sum_{k=N+1}^{N+n} \omega_{k,n}^N \|\pi - Q_\gamma^k\|_{\mathrm{TV}} \;,
\end{aligned}
$$

where $\mathsf{D} = \{(x,y)\in\mathbb{R}^d\times\mathbb{R}^d \,|\, x=y\}$. Using the decomposition (6.25) and Theorem 6.13, Theorem 6.14 and Theorem 6.15, we can deduce different bounds for the bias, depending on the assumptions on U and the sequence of step sizes $(\gamma_k)_{k\geq 1}$.

The following result give a bound on the variance term. Define for all $n,\ell\geq 1$, $n<\ell$

$$
\Lambda_{n,\ell} = \kappa^{-1}\left\{\prod_{j=n}^{\ell}(1-\kappa\gamma_j)^{-1} - 1\right\} \;, \qquad \Lambda_\ell = \Lambda_{1,\ell} \;. \tag{6.32}
$$

Theorem 6.16. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N\geq 0$, $n\geq 1$, $x\in\mathbb{R}^d$ and $f\in\mathbb{F}_b(\mathbb{R}^d)$, we get*

$$
\begin{aligned}
\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} \leq 2\left\|f\right\|_\infty^2 \Gamma_{N+2,N+n+1}^{-1} &+ 8\left\|f\right\|_\infty^2 \sum_{k=N}^{N+n-1}\gamma_{k+1}\left\{\sum_{i=k+2}^{N+n}\omega_{i,n}^N/\Lambda_{k+2,i}^{1/2}\right\}^2 \\
&+ 8\kappa^{-1}\left\|f\right\|_\infty^2\left\{\sum_{i=N+1}^{N+n}\omega_{i,n}^N/\Lambda_{N+1,i}^{1/2}\right\}^2 \;.
\end{aligned}
$$

Proof. The proof is postponed to Section 6.8.12. □

We now establish an exponential deviation inequality for $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ given by (6.21) for a bounded measurable function f.

Theorem 6.17. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $(X_n)_{n\geq 0}$ be given by (6.2) and started at $x\in\mathbb{R}^d$. Then for all $N\geq 0$, $n\geq 1$, $r>0$, and functions $f\in\mathbb{F}_b(\mathbb{R}^d)$:*

$$
\mathbb{P}_x\left[\hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r\right] \leq \exp\left[-\left\{r - \left\|f\right\|_\infty(\Gamma_{N+2,N+n+1})^{-1}\right\}^2 / (16\left\|f\right\|_\infty^2 u_{N,n}^{(4)}(\gamma))\right] \;,
$$

where

$$
u_{N,n}^{(4)}(\gamma) = \sum_{k=N+1}^{N+n}\gamma_{k+1}\left(\sum_{i=k+2}^{N+n}\omega_{i,n}^N/\Lambda_{k+2,i}^{1/2}\right)^2 + \kappa^{-1}\left(\sum_{i=N+1}^{N+n}\omega_{i,n}^N/\Lambda_{N+1,i}^{1/2}\right)^2 \;. \tag{6.33}
$$

Proof. The proof is postponed to Section 6.8.13. □

6.6 Convergence in Wasserstein distance of order p

In this section, we extend the results of Section 6.2 to Wasserstein distance of order $p \in \mathbb{N}$, $p > 2$. In Section 6.2, we needed to bound the second moment of π. We begin with giving a bound on the $2p$-th moment of π for $p \geq 1$. For this define for all $p \geq 1$ and $k \in \{0, \cdots, p\}$,

$$a_{k,p} = m^{k-p} \prod_{i=k+1}^{p} \{d + 2(i-1)\} \; , \tag{6.34}$$

and $a_{0,0} = 1$.

Proposition 6.18. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11 and let* $2p \in \mathbb{N}$, $p \geq 1$. *The stationary distribution* π *belongs to* $\mathcal{P}_{2p}(\mathbb{R}^d)$ *and satisfies*

$$\int_{\mathbb{R}^d} \|x - x^\star\|^{2p} \, \pi(\mathrm{d}x) \leq a_{0,p} \; . \tag{6.35}$$

Proof. The proof is postponed to Section 6.8.14. □

For all $p \geq 1$, $k = 0, 1, 2$, define by induction the function $\tilde{a}_{k,p} : \mathbb{R}^d \to \mathbb{R}_+$ by for all $y \in \mathbb{R}^d$

$$\tilde{a}_{k,p}(x) = \tilde{a}_k(x) \prod_{i=2}^{p} \left\{ i \left(L^2/(2m) \|x - x^\star\|^2 + d + 2(i-1) \right) (k+i)^{-1} \right\} \; , \tag{6.36}$$

$$\tilde{a}_0(x) = 2d \; , \qquad \tilde{a}_1(x) = (3/2)L^2 \|x - x^\star\|^2 \; , \qquad \tilde{a}_2(x) = dL^2/3 \; .$$

Theorem 6.19. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 1/(m+L)$ *and* $p \in \mathbb{N}$, $p \geq 1$. *Then for all* $\mu_0 \in \mathcal{P}_{2p}(\mathbb{R}^d)$ *and* $n \geq 1$,

$$W_{2p}^{2p}(\mu_0 Q_\gamma^n, \pi) \leq u_n^{(1,p)}(\gamma) W_{2p}^{2p}(\mu_0, \pi) + u_n^{(2,p)}(\gamma) \; ,$$

where

$$u_n^{(1,p)}(\gamma) \stackrel{\text{def}}{=} \prod_{k=1}^{n} (1 - \kappa\gamma_k/4)^p \tag{6.37}$$

and

$$u_n^{(2,p)}(\gamma) \stackrel{\text{def}}{=} \sum_{i=1}^{n} L^{2p} \gamma_i^{p+1} (2\kappa^{-1})^{p-1} (\kappa^{-1} + \gamma_i)^p \left\{ \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/4)^p \right\} \sum_{l=0}^{2} U_{l,p} (l+p+1)^{-1} \gamma_i^l \; , \tag{6.38}$$

$$U_{l,p} \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \tilde{a}_{l,p}(y) \mathrm{d}\pi(y) \; , \tag{6.39}$$

κ *is defined in* (6.6), $\{\tilde{a}_{l,p} : \mathbb{R}^d \to \mathbb{R}, p \geq 1, l \in \{0,1,2\}\}$ *in* (6.36).

Proof. The proof is postponed to 6.8.15 □

Note that quantitative bounds on $U_{l,p}$ defined by (6.39) can be obtained by using Proposition 6.18. Following the proof of Corollary 6.5, if the sequence $(\gamma_k)_{k\geq 0}$ satisfies $\lim_{k\to+\infty}\gamma_k=0$, $\lim_{k\to+\infty}\Gamma_k=+\infty$, we get under the assumptions of Theorem 6.19 that $\lim_{n\to+\infty}W_{2p}(\mu_0 Q_\gamma^n,\pi)=0$. Another consequence of Theorem 6.19 is that under the assumptions of these Theorem, we have in the case of constant step size $\gamma_k=\gamma$ for all $k\geq 1$ with $\gamma\leq 1/(m+L)$, that $W_{2p}^{2p}(\pi_\gamma,\pi)\leq C(d\gamma)^p$, for some constant $C\geq 0$ independent of the dimension d and the step size γ.

We now give bounds on $W_{2p}^{2p}(\mu_0 Q_\gamma^n,\pi)$ under **H** 12 for all $\mu_0\in\mathcal{P}(\mathbb{R}^d)$, $p\geq 1$ and $n\geq 0$ which can be computed by induction on p.

Theorem 6.20. *Assume **H**10, **H**11 and **H**12. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1\leq 1/(m+L)$ and $p\in\mathbb{N}$, $p\geq 1$. Then for all $\mu_0\in\mathcal{P}_{2p}(\mathbb{R}^d)$ and $n\geq 1$, $W_{2p}^{2p}(\mu_0 Q_\gamma^n,\pi)\leq w_n^{(p)}(\mu_0,\gamma)$ where $\{w_n^{(p)}(\mu_0,\gamma)\mid ,n\in\mathbb{N},p\in\mathbb{N}^*\}$ is defined for all $n\geq 0$, $w_n^{(1)}=u_n^{(1)}W_2(\mu_0,\pi)+u_n^{(3)}$ for $(u_n^1)_{n\geq 0}$ and $(u_n^{(3)})_{n\geq 0}$ given by Theorem 6.7 and for all $p\geq 2$, $n\geq 0$, $w_n^{(p)}(\mu_0,\gamma)\leq u_n^{(1,p)}(\gamma)W_{2p}^{2p}(\mu_0,\pi)+u_n^{(3,p)}(\mu_0,\gamma)$ for $u_n^{(1,p)}(\gamma)$ given in (6.37),*

$$u_n^{(3,p)}(\mu_0,\gamma)\overset{\text{def}}{=}\sum_{i=1}^n\left\{\prod_{k=i+1}^n(1-\kappa\gamma_k/4)^p\right\}(6\kappa^{-1})^{p-1}\left(\gamma_k^{p/2+1}\frac{(2^{3/2}p^{1/2}Ld^{1/2})^p}{p/2+1}(w_{i-1}^{(p-1)}(\mu_0,\gamma))^{p/(2(p-1))}\right.$$
$$\left.+\gamma_i^{2p+1}\left[(2L^2)^p\sum_{l=0}^2\frac{U_{l,p}\gamma_k^l}{l+p+1}+(2p+1)^{-1}(2\kappa^{-1})^p\left\{2^{2p-1}L^{4p}a_{0,p}\sum_{l=0}^p a_{l,p}+(d\tilde{L}^2)^p/2\right\}\right]\right)\ ,$$

where κ, $\{\tilde{a}_{l,p}:\mathbb{R}^d\to\mathbb{R}_+\mid p\geq 1,l\in\{0,1,2\}\}$, $\{a_{k,p}\in\mathbb{R}_+\mid p\geq 1,k\in\{0,\cdots,p\}\}$, $\{U_{l,p}\in\mathbb{R}_+\mid p\geq 1,l\in\{0,1,2\}\}$ are given in (6.6)-(6.36)-(6.34)-(6.39) respectively.

Proof. The proof is postponed to 6.8.16. ☐

For the case $p=2$ and constant step sizes $\gamma_k=\gamma\in(0,1/(m+L))$ for all $k\geq 1$ we can see that the bounds obtained in Theorem 6.20 improve the one in Theorem 6.19 since it implies with (6.12) and Corollary 6.8, the bound $W_4(\pi_\gamma,\pi)\leq C\gamma^{3/4}d^{1/2}$.

6.7 Numerical experiments

Consider a binary regression set-up in which the binary observations (responses) $(Y_1,\ldots,Y_p)$ are conditionally independent Bernoulli random variables with success probability $\varrho(\boldsymbol{\beta}^T X_i)$, where ϱ is the logistic function defined for $z\in\mathbb{R}$ by $\varrho(z)=\mathrm{e}^z/(1+\mathrm{e}^z)$ and X_i and $\boldsymbol{\beta}$ are d dimensional vectors of known covariates and unknown regression coefficient, respectively. The prior distribution for the parameter $\boldsymbol{\beta}$ is a zero-mean Gaussian distribution with covariance matrix Σ. The posterior density distribution of $\boldsymbol{\beta}$ is up to a proportionality constant given by

$$\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}|((X_i,Y_i))_{1\leq i\leq p})\propto\exp\left(\sum_{i=1}^p Y_i\boldsymbol{\beta}^T X_i-\log(1+\mathrm{e}^{\boldsymbol{\beta}^T X_i})-(1/2)\boldsymbol{\beta}^T\Sigma^{-1}\boldsymbol{\beta}\right)\ .$$

Bayesian inference for the logistic regression model has long been recognized as a numerically involved problem, due to the analytically inconvenient form of the model's likelihood function. Several algorithms have been proposed, trying to mimick the data-augmentation (DA) approach of [AC93] for probit regression; see [HH06], [FF10] and [GP12]. Recently, a very promising DA algorithm has been proposed in [PSW13], using the Polya-Gamma distribution in the DA part. This algorithm has been shown to be uniformly ergodic for the total variation by [CH13, Proposition 1], which provides an explicit expression for the ergodicity constant. This constant is exponentially small in the dimension of the parameter space and the number of samples (it is likely however that this constant is very conservative). Moreover, the complexity of the augmentation step is cubic in the dimension, which prevents from using this algorithm when the dimension of the regressor is large.

We apply ULA to sample from the posterior distribution $\pi_{\boldsymbol{\beta}}(\cdot|(X_i, Y_i)_{1 \le i \le p})$. The gradient of its log-density may be expressed as

$$\nabla \log\{\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta}|(X_i, Y_i)_{1 \le i \le p})\} = \sum_{i=1}^{p} Y_i X_i - \frac{X_i}{1 + \mathrm{e}^{-\boldsymbol{\beta}^T X_i}} - \Sigma^{-1}\boldsymbol{\beta} \; ,$$

Therefore $-\log \pi_{\boldsymbol{\beta}}(\cdot|(X_i, Y_i)_{1 \le i \le p})$ is strongly convex **H**11 with $m = \lambda_{\max}^{-1}(\Sigma)$ and satisfies **H**10 with $L = (1/4) \max_{1 \le i \le p}\{\|X_i\| + \lambda_{\min}^{-1}(\Sigma)\}$, where $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ are the minimal and maximal eigenvalues of Σ, respectively. To assess the proposed algorithm, we first compare the histograms given by ULA and the Pòlya-Gamma Gibbs sampling from [PSW13]. For this, we take $d = 5$, $p = 100$, generate synthetic data $(Y_i)_{1 \le i \le p}$ and $(X_i)_{1 \le i \le p}$, and set $\Sigma_{\boldsymbol{\beta}} = (\sum_{i=1}^{p} \|X_i\|^2)(dp)^{-1} \mathrm{I}_d$. We produce 10^7 samples from the Pòlya-Gamma sampler using the R package BayesLogit [WPS13]. Next, we make 10^3 runs of the Euler approximation scheme with $n = 10^6$ effective iterations, with a constant sequence $(\gamma_k)_{k \ge 1}$, $\gamma_k = 10(\kappa n^{1/2})^{-1}$ for all $k \ge 0$ and a burn-in period $N = n^{1/2}$. The plot of the histogram of the Pólya-Gamma Gibbs sampler for one component, the corresponding mean of the obtained histograms for ULA and the quantiles at 95% can be found in Figure 6.1. The same procedure is also applied with the decreasing step size sequence $(\gamma_k)_{k \ge 1}$ defined by $\gamma_k = \gamma_1 k^{-1/2}$, with $\gamma_1 = 10(\kappa \log(n)^{1/2})^{-1}$, and for the burn in period $N = \log(n)$, see also Figure 6.1. In addition, we also compare the Pólya-Gamma Gibbs sampler, MALA and ULA on four real data sets, which are summarized in Table 6.12. Note that for the Australian credit data set, the ordinal covariates have been stratified by dummy variables. Furthermore, we normalized the data sets and consider the Zellner prior setting $\Sigma^{-1} = (\pi^2 p/3)\Sigma_X^{-1}$ where $\Sigma_X = p^{-1}\sum_{i=1}^{p} X_i X_i^T$; see [SH11], [HBJ14] and the references therein. Also, we apply a pre-conditioned version of MALA and ULA, targeting the probability density $\tilde{\pi}_{\boldsymbol{\beta}}(\cdot) \propto \pi_{\boldsymbol{\beta}}(\Sigma_X^{1/2} \cdot)$. Then, we obtain samples from $\pi_{\boldsymbol{\beta}}$ by post-multiplying the obtained draws by $\Sigma_X^{1/2}$. For each data sets, 100 runs of the Polya-Gamma Gibbs sampler (10^5 iterations per run), and 100 runs of MALA and ULA (10^6 iterations per run) have been performed. Despite the fact that longer runs are carried out, the computational time of ULA is still two orders of magnitude lower than the Pólya-Gamma simulator. For MALA, the step-size is chosen so that the acceptance probability in stationarity is approximately equal to 0.5. For ULA, we choose constant

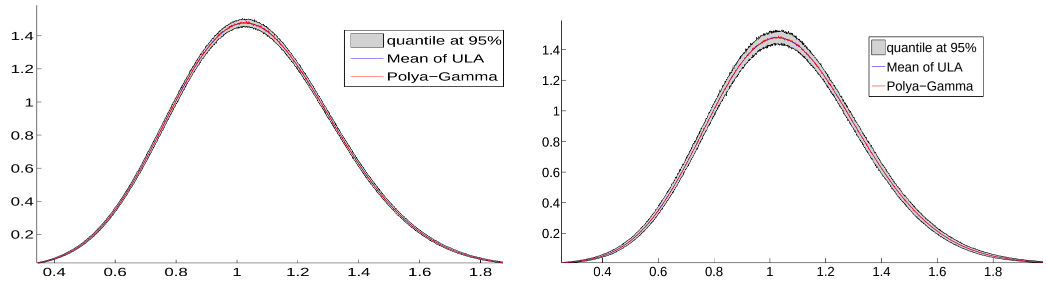


Figure 6.1: Empirical distribution comparison between the Polya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$

Data set \ Dimensions	Observations p	Covariates d
German credit [1]	1000	25
Heart disease [2]	270	14
Australian credit[3]	690	35
Prima indian diabetes[4]	768	9

Table 6.12: Dimension of the data sets

step-sizes $\gamma = 5 \times 10^{-3}$ for all the data sets. We display the boxplots of the estimators for the mean of one component of $\boldsymbol{\beta}$ in Figure 6.2. Note that there are some discrepancies between the posterior mean estimators obtained using either the DA, MALA and ULA. These differences are of order 10^{-3} and are likely to be due to accumulations of numerical errors. These differences are negligible compared to the posterior variance of these estimators, which is of order 10^{-1}. These results all imply that ULA is a much simpler and faster alternative to the Polya-Gamma Gibbs sampler and MALA algorithm.

[1] http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)
[2] http://archive.ics.uci.edu/ml/datasets/Statlog+(Heart)
[3] http://archive.ics.uci.edu/ml/datasets/Statlog+(Australian+Credit+Approval)
[4] http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

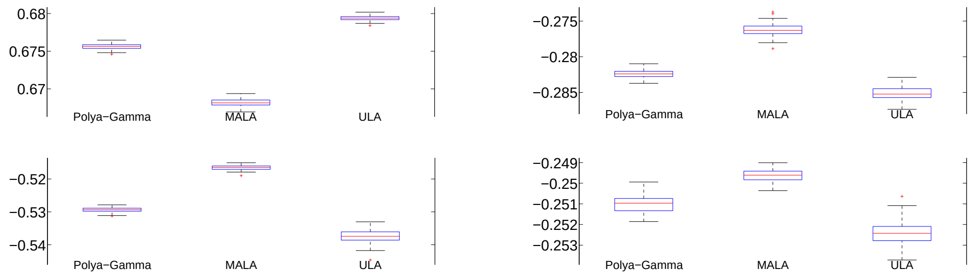


Figure 6.2: Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Prima Indian diabetes data set

6.8 Proofs

Let $(\mathcal{F}_t)_{t\geq 0}$ be the filtration associated with $(B_t)_{t\geq 0}$ with $\mathcal{F}_0$, the σ-field generated by $(Y_0, \overline{Y}_0)$.

6.8.1 Proof of Theorem 6.1

We preface the proof by a technical Lemma. Denote by $x^\star$ the unique minimizer of U. The generator $\mathscr{A}$ associated with $(P_t)_{t\geq 0}$ is given, for all $f \in C^2(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$, by:

$$\mathscr{A} f(x) = -\langle \nabla U(x), \nabla f(x) \rangle + \Delta f(x) \; . \tag{6.40}$$

Lemma 6.21. *Assume **H**10 and **H**11.*

(i) For all $t \geq 0$ and $x \in \mathbb{R}^d$,

$$\mathbb{E}_x\left[\|Y_t - x^\star\|^2\right] \leq \|x - x^\star\|^2 \,\mathrm{e}^{-2mt} + \frac{d}{m}(1-\mathrm{e}^{-2mt}) \; .$$

(ii) For all $t \geq 0$ and $x \in \mathbb{R}^d$,

$$\mathbb{E}_x\left[\|Y_t - x\|^2\right] \leq dt(2 + L^2t^2/3) + (3/2)t^2L^2\,\|x - x^\star\|^2 \; .$$

Proof. (i) Denote for all $x \in \mathbb{R}^d$ by $V(x) = \|x - x^\star\|^2$. Under **H**10 $\sup_{t\in[0,T]} \mathbb{E}_x[\|Y_t\|^2] < +\infty$ for all $T \geq 0$. Therefore, the process $\left(V(Y_t) - V(x) - \int_0^t \mathscr{A} V(Y_s)\mathrm{d}s\right)_{t\geq 0}$ is a $(\mathcal{F}_t)_{t\geq 0}$-martingale under $\mathbb{P}_x$. Since $\nabla U(x^\star) = 0$ and using (6.5), we have

$$\mathscr{A} V(x) = 2\left(-\langle \nabla U(x) - \nabla U(x^\star), x - x^\star\rangle + d\right) \leq 2\left(-mV(x) + d\right) \; . \tag{6.41}$$

Denote for all $t \geq 0$ and $x \in \mathbb{R}^d$ by $v(t,x) = P_tV(x)$. Then we have, $\partial v(t,x)/\partial t = P_t\mathscr{A} V(x)$. Using (6.41), we get

$$\frac{\partial v(t,x)}{\partial t} = P_t\mathscr{A} V(x) \leq -2mP_tV(x) + 2d = -2mv(t,x) + 2d \; ,$$

and the proof follows from the Grönwall inequality.

(ii) Denote for all $x, y \in \mathbb{R}^d$, $\tilde{V}_x(y) = \|y - x\|^2$. therefore the process $(\tilde{V}_x(Y_t) - \tilde{V}_x(x) - \int_0^t \mathscr{A}\tilde{V}_x(Y_s)\mathrm{d}s)_{t\geq 0}$, is a $(\mathcal{F}_t)_{t\geq 0}$-martingale under $\mathbb{P}_x$. Denote for all $t \geq 0$ and $x \in \mathbb{R}^d$ by $\tilde{v}(t,x) = P_t\tilde{V}_x(x)$. Then we get,

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A}\tilde{V}_x(x) \; . \tag{6.42}$$

By (6.5), we have for all $y \in \mathbb{R}^d$,

$$\mathscr{A}\tilde{V}_x(y) = 2\left(-\langle \nabla U(y), y - x\rangle + d\right) \leq 2\left(-m\tilde{V}_x(y) + d - \langle \nabla U(x), y - x\rangle\right) \; . \tag{6.43}$$

Using (6.42), this inequality and that $\tilde{V}_x$ is positive, we get

$$\frac{\partial \tilde{v}(t,x)}{\partial t} = P_t \mathscr{A}\tilde{V}_x(x) \leq 2\left(d - \int_{\mathbb{R}^d} \langle \nabla U(x), y - x\rangle \, P_t(x, \mathrm{d}y)\right) \; . \tag{6.44}$$

By the Cauchy-Schwarz inequality, $\nabla U(x^\star) = 0$, (6.1) and the Jensen inequality, we have,

$$\begin{aligned}
\left|\mathbb{E}_x\left[\langle \nabla U(x), Y_t - x\rangle\right]\right| &\leq \|\nabla U(x)\| \left\|\mathbb{E}_x\left[Y_t - x\right]\right\| \\
&\leq \|\nabla U(x)\| \left\|\mathbb{E}_x\left[\int_0^t \{\nabla U(Y_s) - \nabla U(x^\star)\}\,\mathrm{d}s\right]\right\| \\
&\leq \sqrt{t}\,\|\nabla U(x) - \nabla U(x^\star)\| \left(\int_0^t \mathbb{E}_x\left[\|\nabla U(Y_s) - \nabla U(x^\star)\|^2\right]\mathrm{d}s\right)^{1/2} .
\end{aligned}$$

Furthermore, by **H**10 and Lemma 6.21-(i), we have

$$\begin{aligned}
\left|\int_{\mathbb{R}^d} \langle \nabla U(x), y - x\rangle \, P_t(x, \mathrm{d}y)\right| &\leq \sqrt{t}L^2\,\|x - x^\star\| \left(\int_0^t \mathbb{E}_x\left[\|Y_s - x^\star\|^2\right]\mathrm{d}s\right)^{1/2} \\
&\leq \sqrt{t}L^2\,\|x - x^\star\| \left(\frac{1 - \mathrm{e}^{-2mt}}{2m}\,\|x - x^\star\|^2 + \frac{2tm + \mathrm{e}^{-2mt} - 1}{2m}(d/m)\right)^{1/2} \\
&\leq L^2\,\|x - x^\star\| \left(t\,\|x - x^\star\| + t^{3/2}d^{1/2}\right) ,
\end{aligned} \tag{6.45}$$

where we used for the last line that by the Taylor theorem with remainder term, for all $s \geq 0$, $(1 - \mathrm{e}^{-2ms})/(2m) \leq s$ and $(2ms + \mathrm{e}^{-2ms} - 1)/(2m) \leq ms^2$, and the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$. Plugging (6.45) in (6.44), and since $2\,\|x - x^\star\|\, t^{3/2}d^{1/2} \leq t\,\|x - x^\star\|^2 + t^2 d$, we get

$$\frac{\partial \tilde{v}(t,x)}{\partial t} \leq 2d + 3L^2 t\,\|x - x^\star\|^2 + L^2t^2 d$$

Since $\tilde{v}(0,x) = 0$, the proof is completed by integrating this result.

□

Proof of Theorem 6.1. (i) Consider the following SDE in $\mathbb{R}^d \times \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}Y_t &= -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \; , \\ \mathrm{d}\tilde{Y}_t &= -\nabla U(\tilde{Y}_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \; , \end{cases} \tag{6.46}$$

where $(Y_0, \tilde{Y}_0)$ is some coupling between μ and ν. Since μ and ν are in $\mathcal{P}_p(\mathbb{R}^d)$ and ∇U is Lipschitz, then by [KS91, Theorem 2.5, Theorem 2.9, Chapter 5], this SDE has a unique strong solution $(Y_t, \tilde{Y}_t)_{t\geq 0}$ associated with $(B_t)_{t\geq 0}$. Moreover since $(Y_t, \tilde{Y}_t)_{t\geq 0}$ is a solution of (6.46),

$$\left\| Y_t - \tilde{Y}_t \right\|^p = \left\| Y_0 - \tilde{Y}_0 \right\|^p - p \int_0^t \left\| Y_s - \tilde{Y}_s \right\|^{p-2} \left\langle \nabla U(Y_s) - \nabla U(\tilde{Y}_s), Y_s - \tilde{Y}_s \right\rangle \mathrm{d}s \; ,$$

which implies using (6.5) and Grönwall's inequality that

$$\left\| Y_t - \tilde{Y}_t \right\|^p \leq \left\| Y_0 - \tilde{Y}_0 \right\|^p - mp \int_0^t \left\| Y_s - \tilde{Y}_s \right\|^p \mathrm{d}s \leq \left\| Y_0 - \tilde{Y}_0 \right\|^p \mathrm{e}^{-mpt} \; .$$

For all $t \geq 0$, the law of $(Y_t, \tilde{Y}_t)$ is a coupling between μP_t and νP_t. Therefore by definition of W_p, $W_p(\mu P_t, \nu P_t) \leq \mathbb{E}[\|Y_t - \tilde{Y}_t\|^p]^{1/p}$ showing (i).

(ii) Set $V(x) = \|x - x^\star\|^2$. By Jensen's inequality and Lemma 6.21-(i), for all $c > 0$ and $t > 0$, we get

$$\begin{aligned} \pi(V \wedge c) = \pi P_t(V \wedge c) \leq \pi(P_t V \wedge c) &= \int \pi(\mathrm{d}x)\, c \wedge \left\{ \|x - x^*\|^2 \mathrm{e}^{-2mt} + \frac{d}{m}(1 - \mathrm{e}^{-2mt}) \right\} \\ &\leq \pi(V \wedge c)\mathrm{e}^{-2mt} + (1 - \mathrm{e}^{-2mt})d/m \; . \end{aligned}$$

Taking the limit as $t \to +\infty$, we get $\pi(V \wedge c) \leq d/m$. Using the monotone convergence theorem, taking the limit as $c \to +\infty$, we finally obtain (6.7).

$\square$

6.8.2 Proof of Theorem 6.2

Note that the proof is trivial if $p < n$. Therefore we only need to consider the case $p \geq n$. For any $\gamma \in (0, 2/(m + L))$, we have for all $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, R_\gamma(x, \mathrm{d}y) = \|x - \gamma \nabla U(x) - x^\star\|^2 + 2\gamma d \; .$$

Using that $\nabla U(x^\star) = 0$, and (6.4), we get from the previous inequality:

$$\begin{aligned} &\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma(x, \mathrm{d}y) \\ &\qquad \leq (1 - \kappa\gamma) \|x - x^\star\|^2 + \gamma \left(\gamma - \frac{2}{m + L} \right) \|\nabla U(x) - \nabla U(x^\star)\|^2 + 2\gamma d \\ &\qquad \leq (1 - \kappa\gamma) \|x - x^\star\|^2 + 2\gamma d \; , \end{aligned}$$

where we have used for the last inequality that $\gamma_1 \leq 2/(m + L)$ and $(\gamma_k)_{k\geq 1}$ is nonincreasing. Then by definition (6.10) of $Q_\gamma^{n,p}$ for $p, n \geq 1$, $p \leq n$, the proof follows from a straightforward induction.

6.8.3 Proof of Theorem 6.3

We preface the proof by a Lemma.

Lemma 6.22. *Assume* ***H*** *10 and* ***H*** *11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$ *and* $p \in \mathbb{N}$, $p \geq 1$. *Then for all* $\mu_0, \nu_0 \in \mathcal{P}_{2p}(\mathbb{R}^d)$ *and* $\ell \geq n \geq 1$,

$$W_{2p}^{2p}(\mu_0 Q_\gamma^{n,\ell}, \nu_0 Q_\gamma^{n,\ell}) \leq \left\{ \prod_{k=n}^{\ell} (1-\kappa\gamma_k) \right\}^p W_{2p}^{2p}(\mu_0, \nu_0) \; ;$$

Proof. Let ζ_0 be an optimal transference plan of μ_0 and ν_0 and $(Z_k)_{k\geq 1}$ be a sequence of i.i.d. d-dimensional Gaussian random variables. We consider the processes $(X_k^1, X_k^2)_{k\geq 0}$ with initial distributions equal to ζ_0 and defined for $k \geq 0$ by

$$X_{k+1}^j = X_k^j - \gamma_{k+n} \nabla U(X_k^j) + \sqrt{2\gamma_{k+n}} Z_{k+1} \qquad j = 1,2 \; . \tag{6.47}$$

Using (6.47), we get for any $p \geq n \geq 0$. $W_{2p}^{2p}(\mu_0 Q_\gamma^{n,\ell}, \nu_0 Q_\gamma^{n,\ell}) \leq \mathbb{E}\left[\|X_\ell^1 - X_\ell^2\|^{2p}\right]$ and (6.4) implies for $k \geq n-1$,

$$\begin{aligned} \left\|X_{k+1}^1 - X_{k+1}^2\right\|^2 &= \left\|X_k^1 - X_k^2\right\|^2 + \gamma_{n+k+1}^2 \left\|\nabla U(X_k^1) - \nabla U(X_k^2)\right\|^2 \\ &\quad - 2\gamma_{n+k} \left\langle X_k^1 - X_k^2, \nabla U(X_k^1) - \nabla U(X_k^2) \right\rangle \leq (1-\kappa\gamma_{n+k+1}) \left\|X_k^1 - X_k^2\right\|^2 \; . \end{aligned}$$

Therefore by a straightforward induction we get for all $\ell \geq n$,

$$\left\|X_\ell^1 - X_\ell^2\right\|^2 \leq \prod_{k=n}^{\ell} (1-\kappa\gamma_k) \left\|X_0^1 - X_0^2\right\|^2 \; .$$

□

Proof of Theorem 6.3. Let $\mu \in \mathcal{P}_{2p}(\mathbb{R}^d)$ and $p \geq 1$. It is straightforward that for all $n \geq 0$, $\mu R_\gamma^n \in \mathcal{P}_{2p}(\mathbb{R}^d)$. Then, by Lemma 6.22 is a strict contraction in $(\mathcal{P}_{2p}(\mathbb{R}^d), W_{2p})$ and there is a unique fixed point π_γ which is the unique invariant distribution. Equation (6.12) follows from Lemma 6.22. □

6.8.4 Proof of Theorem 6.4

We preface the proof by a technical Lemma.

Lemma 6.23. *Assume* ***H*** *10 and* ***H*** *11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 1/(m+L)$. *Let* $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, $(Y_t, \overline{Y}_t)_{t\geq 0}$ *such that* $(Y_0, \overline{Y}_0)$ *is distributed according to* ζ_0 *and given by* (6.13). *Then almost surely for all* $n \geq 0$ *and* $\epsilon > 0$,

$$\begin{aligned} \left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2 &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \\ &\quad + (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \, \mathrm{d}s \; , \end{aligned} \tag{6.48}$$

$$\begin{aligned} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right] &\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \\ &\quad + L^2\gamma_{n+1}^2 (1/(4\epsilon) + \gamma_{n+1}) \left(2d + L^2\gamma_{n+1} \|Y_{\Gamma_n} - x^\star\|^2 + dL^2\gamma_{n+1}^2/6\right) \; . \end{aligned} \tag{6.49}$$

Proof. Let $n \geq 0$ and $\epsilon > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. We first show (6.48).

By definition we have:

$$\|\Theta_{n+1}\|^2 = \|\Theta_n\|^2 + \left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \left\{ \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}) \right\} \mathrm{d}s \right\|^2 - 2\gamma_{n+1} \left\langle \Theta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n}) \right\rangle$$
$$- 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{ \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \} \rangle \, \mathrm{d}s \, . \quad (6.50)$$

Young's inequality and Jensen's inequality imply

$$\left\| \int_{\Gamma_n}^{\Gamma_{n+1}} \left\{ \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}) \right\} \mathrm{d}s \right\|^2 \leq 2\gamma_{n+1}^2 \left\| \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n}) \right\|^2$$
$$+ 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \|^2 \, \mathrm{d}s \, .$$

Using (6.4), $\gamma_1 \leq 1/(m+L)$ and $(\gamma_k)_{k \geq 1}$ is nonincreasing, (6.50) becomes

$$\|\Theta_{n+1}\|^2 \leq \{1 - \gamma_{n+1}\kappa\} \|\Theta_n\|^2 + 2\gamma_{n+1} \int_{\Gamma_n}^{\Gamma_{n+1}} \| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \|^2 \, \mathrm{d}s$$
$$- 2 \int_{\Gamma_n}^{\Gamma_{n+1}} \langle \Theta_n, \{ \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \} \rangle \, \mathrm{d}s \, . \quad (6.51)$$

Using the inequality $| \langle a, b \rangle | \leq \epsilon \|a\|^2 + (4\epsilon)^{-1} \|b\|^2$ concludes the proof of (6.48).

We now prove (6.49). Note that (6.48) implies that

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|\Theta_{n+1}\|^2 \right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon)\} \|\Theta_n\|^2$$
$$+ (2\gamma_{n+1} + (2\epsilon)^{-1}) \int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \|^2 \right] \mathrm{d}s \, . \quad (6.52)$$

By **H**10, the Markov property of $(Y_t)_{t \geq 0}$ and Lemma 6.21-(ii), we have

$$\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\| \nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) \|^2 \right] \mathrm{d}s$$
$$\leq L^2 \left(d\gamma_{n+1}^2 + dL^2 \gamma_{n+1}^4 / 12 + (1/2) L^2 \gamma_{n+1}^3 \|Y_{\Gamma_n} - x^\star\|^2 \right) \, .$$

The proof is then concluded plugging this bound in (6.52) . □

Proof of Theorem 6.4. Let ζ_0 be an optimal transference plan of μ_0 and π. Let $(Y_t, \overline{Y}_t)_{t \geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (6.13). By definition of W_2 and since for all $t \geq 0$, π is invariant for P_t, $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}_{\zeta_0}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right]$. Lemma 6.23-(6.49) with $\epsilon = \kappa/4$, a straightforward induction and Lemma 6.21-(i) imply for all $n \geq 0$

$$\mathbb{E}_{\zeta_0} \left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^2 \right] \leq u_n^{(1)}(\gamma) \mathbb{E}_{\zeta_0} \left[\left\| Y_0 - \overline{Y}_0 \right\|^2 \right] + A_n(\gamma) \, , \quad (6.53)$$

where $(u_n^{(1)}(\gamma))_{n\geq 1}$ is given by (6.14), and

$$A_n(\gamma) \stackrel{\text{def}}{=} L^2 \sum_{i=1}^{n} \gamma_i^2 \left\{\kappa^{-1} + \gamma_i\right\} (2d + dL^2\gamma_i^2/6) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) \\ + L^4 \sum_{i=1}^{n} \tilde{\delta}_i \gamma_i^3 \left\{\kappa^{-1} + \gamma_i\right\} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)$$

with

$$\tilde{\delta}_i = \mathrm{e}^{-2m\Gamma_{i-1}} \mathbb{E}_{\zeta_0}\left[\|Y_0 - x^\star\|^2\right] + (1 - \mathrm{e}^{-2m\Gamma_{i-1}})(d/m) \ . \tag{6.54}$$

Then the proof follows since Y_0 is distributed according to π and by (6.7), which shows that for all $i \in \{1, \cdots, n\}$, $\delta_i \leq d/m$. □

6.8.5 Proof of Corollary 6.5

Lemma 6.24. *Let $(\gamma_k)_{k\geq 1}$ be a sequence of nonincreasing real numbers, $\varpi > 0$ and $\gamma_1 < \varpi^{-1}$. Then for all $n \geq 0$, $j \geq 1$ and $\ell \in \{1, \ldots, n+1\}$,*

$$\sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi\gamma_k)\, \gamma_i^j \leq \prod_{k=\ell}^{n+1} (1 - \varpi\gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi} \ .$$

Proof. Let $\ell \in \{1, \ldots, n+1\}$. Since $(\gamma_k)_{k\geq 1}$ is non-increasing,

$$\begin{aligned}\sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi\gamma_k)\, \gamma_i^j &= \sum_{i=1}^{\ell-1} \prod_{k=i+1}^{n+1} (1 - \varpi\gamma_k)\, \gamma_i^j + \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi\gamma_k)\, \gamma_i^j \\ &\leq \prod_{k=\ell}^{n+1} (1 - \varpi\gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \gamma_\ell^{j-1} \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} (1 - \varpi\gamma_k)\, \gamma_i \\ &\leq \prod_{k=\ell}^{n+1} (1 - \varpi\gamma_k) \sum_{i=1}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi} \ .\end{aligned}$$

□

Proof of Corollary 6.5. By Theorem 6.4, it suffices to show that $u_n^{(1)}$ and $u_n^{(2)}$, defined by (6.14) and (6.15) respectively, goes to 0 as $n \to +\infty$. Using the bound $1 + t \leq \mathrm{e}^t$ for $t \in \mathbb{R}$, and $\lim_{n\to+\infty} \Gamma_n = +\infty$, we have $\lim_{n\to+\infty} u_n^{(1)} = 0$. Now to show that $\lim_{n\to+\infty} u_n^{(2)} = 0$, a sufficient condition since $(\gamma_k)_{k\geq 0}$ is nonincreasing, is that $\lim_{n\to+\infty} \sum_{i=1}^n \prod_{k=i+1}^n (1 - \kappa\gamma_k/2)\, \gamma_i^2 = 0$. But since $(\gamma_k)_{k\geq 1}$ is nonincreasing, there exists $c \geq 0$ such that $c\Gamma_n \leq n - 1$ and by Lemma 6.24 applied with $\ell = \lfloor c\Gamma_n \rfloor$ the integer part of $c\Gamma_n$:

$$\sum_{i=1}^{n} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)\, \gamma_i^2 \leq 2\gamma_{\lfloor c\Gamma_n \rfloor}/\kappa + \exp\left(-\kappa\Gamma_n(1 - \Gamma_n^{-1}\Gamma_{\lfloor c\Gamma_n \rfloor})/2\right) \sum_{i=1}^{\lfloor c\Gamma_n \rfloor - 1} \gamma_i \ . \tag{6.55}$$

Since $\lim_{k\to+\infty}\gamma_k = 0$, by the Cesáro theorem, $\lim_{n\to+\infty} n^{-1}\Gamma_n = 0$. Therefore since $\lim_{n\to+\infty}\Gamma_n = +\infty$, $\lim_{n\to+\infty}(\Gamma_n)^{-1}\Gamma_{\lfloor c\Gamma_n\rfloor} = 0$, and the conclusion follows from combining in (6.55), this limit, $\lim_{k\to+\infty}\gamma_k = 0$, $\lim_{n\to+\infty}\Gamma_n = +\infty$ and $\sum_{i=1}^{\lfloor c\Gamma_n\rfloor -1}\gamma_i \leq c\gamma_1\Gamma_n$. □

6.8.6 Proof of Corollary 6.6

Since by Theorem 6.3, for all $x\in\mathbb{R}^d$, $(\delta_x R_\gamma^n)_{n\geq 0}$ converges to π_γ as $n\to\infty$ in $(\mathcal{P}_2(\mathbb{R}^d), W_2)$, the proof then follows from Theorem 6.4 and Lemma 6.24 applied with $\ell = 1$.

6.8.7 Proofs of Theorem 6.7

Lemma 6.25. *Assume **H**10, **H**11 and **H**12. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. and $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$. Let $(Y_t, \overline{Y}_t)_{t\geq 0}$ be defined by (6.13) such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0. Then for all $n\geq 0$ and $\epsilon > 0$, almost surely*

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right] \leq \{1-\gamma_{n+1}(\kappa - 2\epsilon)\}\left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 \tag{6.56}$$
$$+\gamma_{n+1}^3\left\{d(2L^2 + \epsilon^{-1}(\tilde{L}^2/12 + \gamma_{n+1}L^4/4) + \gamma_{n+1}^2 L^4/6) + L^4(\epsilon^{-1}/3 + \gamma_{n+1})\left\|Y_{\Gamma_n} - x^\star\right\|^2\right\}\ .$$

Proof. Let $n\geq 0$ and $\epsilon > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. Using Itô's formula, we have for all $s\in[\Gamma_n, \Gamma_{n+1})$,

$$\nabla U(Y_s) - \nabla U(Y_{\Gamma_n}) = \int_{\Gamma_n}^s \left\{\nabla^2 U(Y_u)\nabla U(Y_u) + \vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u + \sqrt{2}\int_{\Gamma_n}^s \nabla^2 U(Y_u)\mathrm{d}B_u\ . \tag{6.57}$$

Since Θ_n is $\mathcal{F}_{\Gamma_n}$-measurable and $(\int_0^s \nabla^2 U(Y_u)\mathrm{d}B_u)_{s\in[0,\Gamma_{n+1}]}$ is a $(\mathcal{F}_s)_{s\in[0,\Gamma_{n+1}]}$-martingale under **H**10, by (6.57) we have:

$$\begin{aligned}&\left|\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\langle\Theta_n, \nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\rangle\right]\right| \\ &\qquad = \left|\left\langle\Theta_n, \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^s\left\{\nabla^2 U(Y_u)\nabla U(Y_u) + \vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right]\right\rangle\right|\end{aligned}$$

Combining this equality and $|\langle a,b\rangle| \leq \epsilon\|a\|^2 + (4\epsilon)^{-1}\|b\|^2$ in (6.51) we have

$$\begin{aligned}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Theta_{n+1}\|^2\right] &\leq \{1-\gamma_{n+1}(\kappa-2\epsilon)\}\|\Theta_n\|^2 + 2\gamma_{n+1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^{\Gamma_{n+1}}\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\mathrm{d}s\right] \\ &+ (2\epsilon)^{-1}\int_{\Gamma_n}^{\Gamma_{n+1}}\left\|\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^s\left\{\nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right]\right\|^2\mathrm{d}s\ .\end{aligned} \tag{6.58}$$

We now separately bound the two last terms of the right hand side. By **H**10, the Markov property of $(Y_t)_{t\geq 0}$ and Lemma 6.21-(ii), we have

$$\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\right] \mathrm{d}s \\ \leq L^2\left(d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (1/2)L^2\gamma_{n+1}^3 \|Y_{\Gamma_n} - x^\star\|^2\right) . \quad (6.59)$$

Also by (6.17), we get using $\nabla U(x^\star) = 0$, Jensen's inequality and Fubini's theorem,

$$\begin{aligned} A &\stackrel{\text{def}}{=} \int_{\Gamma_n}^{\Gamma_{n+1}} \left\| \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^{s} \nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\mathrm{d}u\right]\right\|^2 \mathrm{d}s \\ &\leq \int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\left\{\nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\right\}\right\|^2\right] \mathrm{d}u\,\mathrm{d}s \\ &\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\nabla^2 U(Y_u)\nabla U(Y_u)\right\|^2 + (1/4)\left\|\vec{\Delta}(\nabla U)(Y_u)\right\|^2\right] \mathrm{d}u\,\mathrm{d}s \\ &\leq 2\int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)L^4\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|Y_u - x^\star\|^2\right] \mathrm{d}u\,\mathrm{d}s + \gamma_{n+1}^3 d\tilde{L}^2/6 . \qquad (6.60) \end{aligned}$$

By Lemma 6.21-(i), the Markov property and for all $t \geq 0$, $1 - \mathrm{e}^{-t} \leq t$, we have for all $s \in [\Gamma_n, \Gamma_{n+1}]$,

$$\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|Y_u - x^\star\|^2\right] \mathrm{d}u \leq (2m)^{-1}(1 - \mathrm{e}^{-2m(s-\Gamma_n)}) \|Y_{\Gamma_n} - x^\star\|^2 + d(s-\Gamma_n)^2 .$$

Using this inequality in (6.60) and for all $t \geq 0$, $1 - \mathrm{e}^{-t} \leq t$, we get

$$A \leq (2L^4\gamma_{n+1}^3/3) \|Y_{\Gamma_n} - x^\star\|^2 + L^4 d\gamma_{n+1}^4/2 + \gamma_{n+1}^3 d\tilde{L}^2/6 .$$

Combining this bound and (6.59) in (6.58) concludes the proof. □

Proof of Theorem 6.7. The proof of the Theorem is the same as the one of Theorem 6.4, using Lemma 6.25 in place of Lemma 6.23, and is omitted. □

6.8.8 Proof of Theorem 6.10

Our main tool is the Gaussian Poincaré inequality [BLM13, Theorem 3.20] (see also [BGL14, Theorem 4.1.1]) which states that if $Z = (Z_1, \ldots, Z_d)$ is a Gaussian vector with identity covariance matrix, then $\mathrm{Var}\{g(Z)\} \leq \|g\|_{\mathrm{Lip}}^2$. The Gaussian Poincaré inequality may be applied to R_γ defined by (6.9) noticing that for all $y \in \mathbb{R}^d$, $R_\gamma(y, \cdot)$ is a Gaussian distribution with mean $y - \gamma\nabla U(y)$ and covariance matrix $2\gamma\,\mathrm{I}_d$.

Lemma 6.26. *Assume **H**10. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function. Then for all $\gamma > 0$, $y \in \mathbb{R}^d$,*

$$0 \leq R_\gamma \{g(\cdot) - R_\gamma g(y)\}^2 (y) = \int R_\gamma(y, \mathrm{d}z)\{g(z) - R_\gamma g(y)\}^2 \leq 2\gamma \|g\|_{\mathrm{Lip}}^2 .$$

To go further, we decompose $\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]$ as the sum of martingale increments,

$$\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)] = \sum_{k=N}^{N+n-1} \left\{ \mathbb{E}_x^{\mathcal{G}_{k+1}} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x^{\mathcal{G}_k} \left[\hat{\pi}_n^N(f) \right] \right\} + \mathbb{E}_x^{\mathcal{G}_N} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \,, \tag{6.61}$$

where $(\mathcal{G}_n)_{n \geq 0}$ here is the natural filtration associated with Euler approximation $(X_n)_{n\geq 0}$. This implies that the variance may be expressed as the following sum

$$\begin{aligned} \mathrm{Var}_x \left\{ \hat{\pi}_n^N(f) \right\} = \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[\left(\mathbb{E}_x^{\mathcal{G}_{k+1}} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x^{\mathcal{G}_k} \left[\hat{\pi}_n^N(f) \right] \right)^2 \right] \\ + \mathbb{E}_x \left[\left(\mathbb{E}_x^{\mathcal{G}_N} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x[\hat{\pi}_n^N(f)] \right)^2 \right] \,. \end{aligned} \tag{6.62}$$

Because $\hat{\pi}_n^N(f)$ is an additive functional, the martingale increment $\mathbb{E}_x^{\mathcal{G}_{k+1}} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x^{\mathcal{G}_k} \left[\hat{\pi}_n^N(f) \right]$ has a simple expression. For $k = N+n-1, \ldots, N+1$, define backward in time the function

$$\Phi_{n,k}^N : x_k \mapsto \omega_{k,n}^N f(x_k) + R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x_k) \,, \tag{6.63}$$

where $\Phi_{n,N+n}^N : x_{N+n} \mapsto \Phi_{n,N+n}^N(x_{N+n}) = \omega_{N+n,n}^N f(x_{N+n})$. Denote finally

$$\Psi_n^N : x_N \mapsto R_{\gamma_{N+1}} \Phi_{n,N+1}^N(x_N) \,. \tag{6.64}$$

Note that for $k \in \{N, \ldots, N+n-1\}$, by the Markov property,

$$\Phi_{n,k+1}^N(X_{k+1}) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) = \mathbb{E}_x^{\mathcal{G}_{k+1}} \left[\hat{\pi}_n^N(f) \right] - \mathbb{E}_x^{\mathcal{G}_k} \left[\hat{\pi}_n^N(f) \right] \,, \tag{6.65}$$

and $\Psi_n^N(X_N) = \mathbb{E}_x^{\mathcal{G}_N} \left[\hat{\pi}_n^N(f) \right]$. With these notations, (6.62) may be equivalently expressed as

$$\begin{aligned} \mathrm{Var}_x \left\{ \hat{\pi}_n^N(f) \right\} = \sum_{k=N}^{N+n-1} \mathbb{E}_x \left[R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(X_k) \right\}^2 (X_k) \right] \\ + \mathrm{Var}_x \left\{ \Psi_n^N(X_N) \right\} \,. \end{aligned} \tag{6.66}$$

Now for $k = N+n-1, \ldots, N$, we will use the Gaussian Poincaré inequality (Lemma 6.26) to the sequence of function $\Phi_{n,k+1}^N$ to prove that $x \mapsto R_{\gamma_{k+1}} \{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(x) \}^2(x)$ is uniformly bounded. It is required to bound the Lipschitz constant of $\Phi_{n,k}^N$. For $k \in \{N, \ldots, N+n-1\}$ and for all $y, z \in \mathbb{R}^d$, we have

$$\left| \Phi_{n,k+1}^N(y) - \Phi_{n,k+1}^N(z) \right| = \left| \omega_{k+1,n}^N \{ f(y) - f(z) \} + \sum_{i=k+2}^{N+n} \omega_{i,n}^N \left\{ Q_\gamma^{k+2,i} f(y) - Q_\gamma^{k+2,i} f(z) \right\} \right| \,. \tag{6.67}$$

Lemma 6.27. *Assume* $\boldsymbol{H}$10 *and* $\boldsymbol{H}$11*. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Then for all Lipschitz functions* $f : \mathbb{R}^d \to \mathbb{R}$ *and* $\ell \geq n \geq 1$*,* $Q_\gamma^{n,\ell} f$ *is a Lipschitz function with*

$$\left\| Q_\gamma^{n,\ell} f \right\|_{\mathrm{Lip}} \leq \prod_{k=n}^{\ell} (1-\kappa\gamma_k)^{1/2} \|f\|_{\mathrm{Lip}} \; .$$

Proof. Recall that for all μ, ν probability measures on $\mathbb{R}^d$ and $p \leq q$, $W_p(\mu,\nu) \leq W_q(\mu,\nu)$. Hence, for all $y, z \in \mathbb{R}^d$, the Monge-Kantorovich theorem (6.3):

$$\left| Q_\gamma^{n,\ell} f(y) - Q_\gamma^{n,\ell} f(z) \right| \leq \|f\|_{\mathrm{Lip}} W_1(\delta_y Q_\gamma^{n,\ell}, \delta_z Q_\gamma^{n,\ell}) \leq \|f\|_{\mathrm{Lip}} W_2(\delta_y Q_\gamma^{n,\ell}, \delta_z Q_\gamma^{n,\ell}) \; .$$

The proof then follows from Lemma 6.22 with $p = 1$. □

Lemma 6.28. *Assume* $\boldsymbol{H}$10 *and* $\boldsymbol{H}$11*. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Let* $N \geq 0$ *and* $n \geq 1$*. Then for all* $y \in \mathbb{R}^d$*, Lipschitz function* f *and* $k \in \{N, \ldots, N+n-1\}$*,*

$$R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2 (y) \leq 8\gamma_{k+1} \|f\|_{\mathrm{Lip}}^2 (\kappa \Gamma_{N+2,N+n+1})^{-2} \; ,$$

where $\Phi_{n,k+1}^N$ *is given by* (6.63)*.*

Proof. By (6.67), $\left\|\Phi_{n,k}^N\right\|_{\mathrm{Lip}} \leq \sum_{i=k+1}^{N+n} \omega_{i,n}^N \left\| Q_\gamma^{k+2,i} f \right\|_{\mathrm{Lip}}$. Using Lemma 6.27, the bound $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$ and the definition of $\omega_{i,n}^N$ given by (6.21), we have

$$\begin{aligned}
\left\|\Phi_{n,k}^N\right\|_{\mathrm{Lip}} &\leq \|f\|_{\mathrm{Lip}} \sum_{i=k+1}^{N+n} \omega_{i,n}^N \prod_{j=k+2}^{i} (1-\kappa\gamma_j/2) \\
&\leq 2\|f\|_{\mathrm{Lip}} (\kappa\Gamma_{N+2,N+n+1})^{-1} \sum_{i=k+1}^{N+n} \left\{ \prod_{j=k+2}^{i} (1-\kappa\gamma_j/2) - \prod_{j=k+2}^{i+1} (1-\kappa\gamma_j/2) \right\} \\
&\leq 2\|f\|_{\mathrm{Lip}} (\kappa\Gamma_{N+2,N+n+1})^{-1} \; .
\end{aligned}$$

Finally, the proof follows from Lemma 6.26. □

Also to control the last term in right hand side of (6.66), we need to control the variance of $\Psi_n^N(X_N)$ under $\delta_x Q_\gamma^N$. But similarly to the sequence of functions $\Phi_{n,k}^N$, Ψ_n^N is Lipschitz by Lemma 6.27 since for all $y, z \in \mathbb{R}^d$, we have

$$\left|\Psi_n^N(y) - \Psi_n^N(z)\right| = \left| \sum_{i=N+1}^{N+n} \omega_{i,n}^N \left\{ Q_\gamma^{N+1,i} f(y) - Q_\gamma^{N+1,i} f(z) \right\} \right| \; . \tag{6.68}$$

Therefore it suffices to find some bound for the variance of g under $\delta_y Q_\gamma^{n,p}$, for $g : \mathbb{R}^d \to \mathbb{R}$ a Lipschitz function, $y \in \mathbb{R}^d$ and $\gamma > 0$, which is done in the following Lemma.

Lemma 6.29. *Assume* ***H*** *10 and* ***H*** *11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Let* $g : \mathbb{R}^d \to \mathbb{R}$ *be a Lipschitz function. Then for all* $n, p \geq 1$, $n \leq p$ *and* $y \in \mathbb{R}^d$

$$0 \leq \int_{\mathbb{R}^d} Q_\gamma^{n,p}(y, \mathrm{d}z) \left\{ g(z) - Q_\gamma^{n,p} g(y) \right\}^2 \leq 2\kappa^{-1} \|g\|_{\mathrm{Lip}}^2 \ ,$$

where $Q_\gamma^{n,p}$ *is given by* (6.10).

Proof. By decomposing $g(X_p) - \mathbb{E}_y^{\mathcal{G}_n}[g(X_p)] = \sum_{k=n+1}^p \{\mathbb{E}_y^{\mathcal{G}_k}[g(X_p)] - \mathbb{E}_y^{\mathcal{G}_{k-1}}[g(X_p)]\}$, and using $\mathbb{E}_y^{\mathcal{G}_k}[g(X_p)] = Q_\gamma^{k+1,p} g(X_k)$, we get

$$\begin{aligned}
\mathrm{Var}_y^{\mathcal{G}_n}\{g(X_p)\} &= \sum_{k=n+1}^p \mathbb{E}_y^{\mathcal{G}_n}\left[\mathbb{E}_y^{\mathcal{G}_{k-1}}\left[\left(\mathbb{E}_y^{\mathcal{G}_k}[g(X_p)] - \mathbb{E}_y^{\mathcal{G}_{k-1}}[g(X_p)]\right)^2\right]\right] \\
&= \sum_{k=n+1}^p \mathbb{E}_y^{\mathcal{G}_n}\left[R_{\gamma_k}\left\{Q_\gamma^{k+1,p} g(\cdot) - R_{\gamma_k} Q_\gamma^{k+1,p} g(X_{k-1})\right\}^2 (X_{k-1})\right] .
\end{aligned}$$

Lemma 6.26 implies $\mathrm{Var}_y^{\mathcal{G}_n}\{g(X_p)\} \leq 2\sum_{k=n+1}^p \gamma_k \left\| Q_\gamma^{k+1,p} g \right\|_{\mathrm{Lip}}^2$. The proof follows from Lemma 6.27 and Lemma 6.24, using the bound $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$. □

Corollary 6.30. *Assume* ***H*** *10 and* ***H*** *11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Then for all Lipschitz function* f *and* $x \in \mathbb{R}^d$,

$$\mathrm{Var}_x\left\{\Psi_n^N(X_N)\right\} \leq 8\kappa^{-3} \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-2} \ ,$$

where Ψ_n^N *is given by* (6.64).

Proof. By (6.68) and Lemma 6.27, Ψ_n^N is Lipschitz function with

$$\left\|\Psi_n^N\right\|_{\mathrm{Lip}} \leq \sum_{i=N+1}^{N+n} \omega_{i,n}^N \left\| Q_\gamma^{N+1,i} f \right\|_{\mathrm{Lip}} .$$

Using Lemma 6.27, the bound $(1-t)^{1/2} \leq 1 - t/2$ for $t \in [0,1]$ and the definition of $\omega_{i,n}^N$ given by (6.21), we have

$$\begin{aligned}
\left\|\Psi_n^N\right\|_{\mathrm{Lip}} &\leq \|f\|_{\mathrm{Lip}} \sum_{i=N+1}^{N+n} \omega_{i,n}^N \prod_{j=N+2}^{i} (1 - \kappa\gamma_j/2) \\
&\leq 2\|f\|_{\mathrm{Lip}} (\kappa \Gamma_{N+2,N+n+1})^{-1} \sum_{i=N+1}^{N+n} \left\{ \prod_{j=N+2}^{i} (1-\kappa\gamma_j/2) - \prod_{j=N+2}^{i+1} (1-\kappa\gamma_j/2) \right\} \\
&\leq 2\|f\|_{\mathrm{Lip}} (\kappa \Gamma_{N+2,N+n+1})^{-1} .
\end{aligned}$$

The proof follows from Lemma 6.29. □

Proof of Theorem 6.10. Plugging the bounds given by Lemma 6.28 and Corollary 6.30 in (6.66), we have

$$\begin{aligned}\mathrm{Var}_x\left\{\hat{\pi}_n^N(f)\right\} &\leq 8\kappa^{-2}\left\|f\right\|_{\mathrm{Lip}}^2\left\{\Gamma_{N+2,N+n+1}^{-2}\Gamma_{N+1,N+n}+\kappa^{-1}\Gamma_{N+2,N+n+1}^{-2}\right\}\\ &\leq 8\kappa^{-2}\left\|f\right\|_{\mathrm{Lip}}^2\left\{\Gamma_{N+2,N+n+1}^{-1}+\Gamma_{N+2,N+n+1}^{-2}(\gamma_{N+1}+\kappa^{-1})\right\}\;.\end{aligned}$$

Using that $\gamma_{N+1}\leq 2/(m+L)$ concludes the proof. □

6.8.9 Proof of Theorem 6.11

Let $N\geq 0$, $n\geq 1$, $x\in\mathbb{R}^d$ and f be a Lipschitz function. To prove Theorem 6.11, we derive an upper bound of the Laplace transform of $\hat{\pi}_n^N(f)-\mathbb{E}_x[\hat{\pi}_n^N(f)]$. Using the decomposition by martingale increments (6.61)

$$\begin{aligned}&\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f)-\mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right]\\ &=\mathbb{E}_x\left[\exp\left(\lambda\{\mathbb{E}_x^{\mathcal{G}_N}\left[\hat{\pi}_n^N(f)\right]-\mathbb{E}_x[\hat{\pi}_n^N(f)]\}+\sum_{k=N}^{N+n-1}\lambda\{\mathbb{E}_x^{\mathcal{G}_{k+1}}\left[\hat{\pi}_n^N(f)\right]-\mathbb{E}_x^{\mathcal{G}_k}\left[\hat{\pi}_n^N(f)\right]\}\right)\right]\;.\end{aligned}$$

Now using (6.65) with the sequence of functions $(\Phi_{n,k}^N)$ and Ψ_n^N given by (6.63) and (6.64), respectively, we have by the Markov property

$$\begin{aligned}&\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f)-\mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right]\\ &=\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\Psi_n^N(X_n)-\mathbb{E}_x\left[\Psi_n^N(X_n)\right]\}}\prod_{k=N}^{N+n-1}R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\Phi_{n,k+1}^N(\cdot)-R_{\gamma_{k+1}}\Phi_{n,k+1}^N(X_k)\}\right)\right\}(X_k)\right]\end{aligned}\tag{6.69}$$

where R_γ is given by (6.9) for $\gamma>0$. We use the same strategy to get concentration inequalities than to bound the variance term in the previous section, replacing the Gaussian Poincaré inequality by the log-Sobolev inequality to get uniform bound for $R_{\gamma_{k+1}}\{\exp(\lambda$
$\{\Phi_{n,k+1}^N(\cdot)-R_{\gamma_{k+1}}\Phi_{n,k+1}^N(X_k)\})\}(X_k)$ w.r.t. X_k, for all $k\in\{N+1,\ldots,N+n\}$. Indeed for all $x\in\mathbb{R}^d$, recall that $R_\gamma(x,\cdot)$ is a Gaussian distribution with mean $x-\gamma\nabla U(x)$ and covariance matrix $2\gamma\,\mathrm{I}_d$. The log-Sobolev inequality provides a bound for the Laplace transform of Lipschitz function $g(Z)-R_\gamma g(x)$ where Z is distributed under $R_\gamma(x,\cdot)$.

Lemma 6.31 ([BLM13, Theorem 5.5])**.** *Assume **H**10. Then for all Lipschitz function g, $\gamma>0$, $x\in\mathbb{R}^d$ and $\lambda>0$,*

$$\int R_\gamma(x,\mathrm{d}y)\left\{\exp\left(\lambda\{g(y)-R_\gamma g(x)\}\right)\right\}\leq\exp\left(\gamma\lambda^2\left\|g\right\|_{\mathrm{Lip}}^2\right)\;.$$

where R_γ is given by (6.9).

We deduced from this lemma, (6.67) and Lemma 6.27, an equivalent of Lemma 6.28 for the Laplace transform of $\Phi^N_{n,k+1}$ under $\delta_y R_{\gamma_{k+1}}$ for $k \in \{N+1,\dots,N+n\}$ and all $y \in \mathbb{R}^d$.

Corollary 6.32. *Assume* ***H**10 and **H**11. Let $(\gamma_k)_{k\ge1}$ be a nonincreasing sequence with $\gamma_1 \le 2/(m+L)$. Let $N \ge 0$ and $n \ge 1$. Then for all $k \in \{N+1,\dots,N+n\}$, $y \in \mathbb{R}^d$ and $\lambda > 0$,*

$$R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\Phi^N_{n,k+1}(\cdot) - R_{\gamma_{k+1}}\Phi^N_{n,k+1}(y)\}\right)\right\}(y) \le \exp\left(4\gamma_{k+1}\lambda^2 \|f\|^2_{\mathrm{Lip}} (\kappa\Gamma_{N+2,N+n+1})^{-2}\right) ,$$

where $\Phi^N_{n,k}$ is given by (6.63).

It remains to control the Laplace transform of Ψ^N_n under $\delta_x Q^N_\gamma$, where $\delta_x Q^N_\gamma$ is defined by (6.10). For this, using again that by (6.68) and Lemma 6.27, Ψ^N_n is a Lipschitz function, we iterate Lemma 6.31 to get bounds on the Laplace transform of Lipschitz function g under $Q^{n,\ell}_\gamma(y,\cdot)$ for all $y \in \mathbb{R}^d$ and $n,\ell \ge 1$, since for all $n,\ell \ge 1$, $Q^{n,\ell}_\gamma g$ is a Lipschitz function by Lemma 6.27.

Lemma 6.33. *Assume* ***H**10 and **H**11. Let $(\gamma_k)_{k\ge1}$ be a nonincreasing sequence with $\gamma_1 \le 2/(m+L)$. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz function, then for all $n,p \ge 1$, $n \le p$, $y \in \mathbb{R}^d$ and $\lambda > 0$:*

$$Q^{n,p}_\gamma\left\{\exp\left(\lambda\{g(\cdot) - Q^{n,p}_\gamma g(y)\}\right)\right\}(y) \le \exp\left(\kappa^{-1}\lambda^2 \|g\|^2_{\mathrm{Lip}}\right) , \tag{6.70}$$

where $Q^\gamma_{n,p}$ is given by (6.10).

Proof. Let $(X_n)_{n\ge0}$ the Euler approximation given by (6.2) and started at $y \in \mathbb{R}^d$. By decomposing $g(X_p) - \mathbb{E}^{\mathcal{G}_n}_y[g(X_p)] = \sum_{k=n+1}^p\{\mathbb{E}^{\mathcal{G}_k}_y[g(X_p)] - \mathbb{E}^{\mathcal{G}_{k-1}}_y[g(X_p)]\}$, and using $\mathbb{E}^{\mathcal{G}_k}_y[g(X_p)] = Q^{k+1,p}_\gamma g(X_k)$, we get

$$\begin{aligned}
&\mathbb{E}^{\mathcal{G}_n}_y\left[\exp\left(\lambda\left\{g(X_p) - \mathbb{E}^{\mathcal{G}_n}_y[g(X_p)]\right\}\right)\right] \\
&= \mathbb{E}^{\mathcal{G}_n}_y\left[\prod_{k=n+1}^p \mathbb{E}^{\mathcal{G}_{k-1}}_y\left[\exp\left(\lambda\left\{\mathbb{E}^{\mathcal{G}_k}_y[g(X_p)] - \mathbb{E}^{\mathcal{G}_{k-1}}_y[g(X_p)]\right\}\right)\right]\right] \\
&= \mathbb{E}^{\mathcal{G}_n}_y\left[\prod_{k=n+1}^p R_{\gamma_k}\exp\left(\lambda\left\{Q^{k+1,p}_\gamma g(\cdot) - R_{\gamma_k}Q^{k+1,p}_\gamma g(X_{k-1})\right\}\right)(X_{k-1})\right] .
\end{aligned}$$

By the Gaussian log-Sobolev inequality Lemma 6.31, we get

$$\mathbb{E}^{\mathcal{G}_n}_y\left[\exp\left(\lambda\left\{g(X_p) - \mathbb{E}^{\mathcal{G}_n}_y[g(X_p)]\right\}\right)\right] \le \exp\left(\lambda^2\sum_{k=n+1}^p \gamma_k\left\|Q^{k+1,p}_\gamma g\right\|^2_{\mathrm{Lip}}\right) .$$

The proof follows from Lemma 6.27 and Lemma 6.24, using the bound $(1-t)^{1/2} \le 1-t/2$ for $t \in [0,1]$.

$\square$

Combining this result with (6.68) and Lemma 6.27, we get an analogue of Corollary 6.30 for the Laplace transform of Ψ_n^N:

Corollary 6.34. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $\lambda > 0$,*

$$\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\Psi_n^N(X_n)-\mathbb{E}_x[\Psi_n^N(X_n)]\}}\right] \leq \exp\left(4\kappa^{-3}\lambda^2 \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-2}\right) ,$$

where Ψ_n^N is given by (6.64).

The Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded using Corollary 6.32 and Corollary 6.34 in (6.69).

Proposition 6.35. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Then for all $N \geq 0$, $n \geq 1$, Lipschitz functions $f : \mathbb{R}^d \to \mathbb{R}$ and $\lambda > 0$:*

$$\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f)-\mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right] \leq \exp\left(4\kappa^{-2}\lambda^2 \|f\|_{\mathrm{Lip}}^2 \Gamma_{N+2,N+n+1}^{-1} u_{N,n}^{(3)}(\gamma)\right) ,$$

where $u_{N,n}^{(3)}(\gamma)$ is given by and (6.23).

6.8.10 Proof of Theorem 6.14

We preface the proof by the following preliminary result.

Lemma 6.36. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$. Then for all $x, y \in \mathbb{R}^d$ and $n \in \mathbb{N}$,*

$$W_2^2(\delta_x Q_\gamma^n, \delta_y P_{n\gamma}) \leq u_n^{(1)} \|x-y\|^2 + \vartheta_n(y)$$

*where $\vartheta_n(y)$ is equal to $\vartheta_{n,0}^{(2)}(y)$ given by (6.28), if **H**12 holds, and to $\vartheta_{n,0}^{(1)}(y)$ given by (6.27) otherwise.*

Proof. The proof is a straightforward application of Lemma 6.23 and Lemma 6.25, as appropriate. □

Proof of Theorem 6.14. We use the following decomposition

$$\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p = \delta_x P_{\Gamma_p} - \delta_x Q_\gamma^{1,n} P_{\Gamma_{n+1,p}} + \delta_x Q_\gamma^{1,n} P_{\Gamma_{n+1,p}} + \delta_x Q_\gamma^p .$$

By the triangle inequality, we get

$$\left\|\delta_x P_{\Gamma_p} - \delta_x Q_\gamma^p\right\|_{\mathrm{TV}} \leq \left\|\left\{\delta_x P_{\Gamma_{1,n}} - \delta_x Q_\gamma^{1,n}\right\} P_{\Gamma_{n+1,p}}\right\|_{\mathrm{TV}} + \left\|\delta_x Q_\gamma^{1,n}\left\{Q_\gamma^{n+1,p} - P_{\Gamma_{n+1,p}}\right\}\right\|_{\mathrm{TV}} . \tag{6.71}$$

We bound the two terms in (6.71) separately. For the first one, using (6.24), Lemma 6.36 and the Cauchy-Schwarz inequality, we have

$$\left\|\left\{\delta_x P_{\Gamma_{1,n}} - \delta_x Q_\gamma^{1,n}\right\} P_{\Gamma_{n+1,p}}\right\|_{\mathrm{TV}} \leq (\vartheta_n(x)/(4\pi\Gamma_{n+1,p}))^{1/2} , \tag{6.72}$$

where $\vartheta_n(x)$ is equal to $\vartheta^{(2)}_{n,0}(x)$ given by (6.28), if **H**12 holds, and to $\vartheta^{(1)}_{n,0}(x)$ given by (6.27) otherwise. For the second term, by [DM15a, Equation 15] (note that we have a different convention for the total variation distance) and the Pinsker inequality, we have

$$\left\|\delta_x Q_\gamma^{1,n}\left\{Q_\gamma^{n+1,p} - P_{\Gamma_{n+1,p}}\right\}\right\|_{\mathrm{TV}}^2 \leq 2^{-3}L^2 \sum_{k=n+1}^{p} \left\{(\gamma_k^3/3)\int_{\mathbb{R}^d} \|\nabla U(z)\|^2 Q_\gamma^{k-1}(x,\mathrm{d}z) + d\gamma_k^2\right\} .$$

By **H**10 and Theorem 6.2, we get

$$\begin{aligned}
&\left\|\delta_x Q_\gamma^{1,n}\left\{Q_\gamma^{n+1,p} - P_{\Gamma_{n+1,p}}\right\}\right\|_{\mathrm{TV}}^2 \\
&\qquad \leq 2^{-3}L^2 \sum_{k=n+1}^{p} \left\{(\gamma_k^3/3)\int_{\mathbb{R}^d} \|\nabla U(z) - \nabla U(x^\star)\|^2 Q_\gamma^{k-1}(x,\mathrm{d}z) + d\gamma_k^2\right\} \\
&\qquad \leq 2^{-3}L^2 \sum_{k=n+1}^{p} \left\{(\gamma_k^3 L^2/3)\varrho_{1,k-1}(x) + d\gamma_k^2\right\} .
\end{aligned}$$

Combining the last inequality and (6.72) in (6.71) concludes the proof. □

6.8.11 Proof of Theorem 6.15

Let $\gamma \in (0, 1/(m+L))$ and $p \in \mathbb{N}^*$. For ease of notation, let $n = \mathsf{n}(\gamma)$, and assume for the moment that $p > 2^n$. Consider the following decomposition

$$\begin{aligned}
\delta_x P_{p\gamma} - \delta_x R_\gamma^p = \left\{\delta_x P_{(p-2^n)\gamma} - \delta_x R_\gamma^{p-2^n}\right\} P_{2^n\gamma} + \delta_x R_\gamma^{p-1}\left\{P_\gamma - R_\gamma\right\} \\
+ \sum_{k=1}^{n} \delta_x R_\gamma^{p-2^k}\left\{P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}}\right\} P_{2^{k-1}\gamma} .
\end{aligned}$$

Therefore using the triangle inequality, we have

$$\begin{aligned}
\left\|\delta_x P_{p\gamma} - \delta_x R_\gamma^p\right\|_{\mathrm{TV}} \leq \left\|\left\{\delta_x P_{(p-2^n)\gamma} - \delta_x R_\gamma^{p-2^n}\right\} P_{2^n\gamma}\right\|_{\mathrm{TV}} + \left\|\delta_x R_\gamma^{p-1}\left\{P_\gamma - R_\gamma\right\}\right\|_{\mathrm{TV}} \\
+ \sum_{k=1}^{n} \left\|\delta_x R_\gamma^{p-2^k}\left\{P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}}\right\} P_{2^{k-1}\gamma}\right\|_{\mathrm{TV}} . \quad (6.73)
\end{aligned}$$

We bound each term in the inequality above. First by (6.24) and Lemma 6.36, we have

$$\left\|\left\{\delta_x P_{(p-2^n)\gamma} - \delta_x R_\gamma^{p-2^n}\right\} P_{2^n\gamma}\right\|_{\mathrm{TV}} \leq \left(\vartheta_{p-2^n,0}(x)/(\pi 2^{n+2}\gamma)\right)^{1/2} , \quad (6.74)$$

where $\vartheta_{p-2^n,0}(x)$ is equal to $\vartheta^{(2)}_{n,0}(x)$ given by (6.28), if **H**12 holds, and to $\vartheta^{(1)}_{n,0}(x)$ given by (6.27) otherwise. Similarly but using in addition Theorem 6.2 and the Cauchy-Schwarz inequality, we have for all $k \in \{1, \cdots, n\}$,

$$\left\|\delta_x R_\gamma^{p-2^k}\left\{P_{2^{k-1}\gamma} - R_\gamma^{2^{k-1}}\right\} P_{2^{k-1}\gamma}\right\|_{\mathrm{TV}} \leq \left(\vartheta_{2^{k-1},p-2^k}(x)/(\pi 2^{k+1}\gamma)\right)^{1/2} , \quad (6.75)$$

where $\vartheta_{2^{k-1},2^n}(x)$ is equal to $\vartheta^{(2)}_{2^{k-1},2^n}(x)$ given by (6.28), if **H**12 holds, and to $\vartheta^{(1)}_{2^{k-1},2^n}(x)$ given by (6.27) otherwise. For the last term, by [DM15a, Equation 15] and the Pinsker inequality, we have

$$\left\|\delta_x R_\gamma^{p-1}\left\{P_\gamma - R_\gamma\right\}\right\|_{\mathrm{TV}}^2 \le 2^{-3}L^2\left\{(\gamma^3/3)\int_{\mathbb{R}^d}\|\nabla U(z)\|^2 R_\gamma^{p-1}(x,\mathrm{d}z) + d\gamma^2\right\} .$$

By **H**10 and Theorem 6.2, we get

$$\left\|\delta_x R_\gamma^{2^n-1}\left\{R_\gamma - P_\gamma\right\}\right\|_{\mathrm{TV}}^2 \le 2^{-3}L^2\left\{(\gamma^3 L^2/3)\varrho_{1,p-1}(x) + d\gamma^2\right\} . \tag{6.76}$$

Combining (6.74),(6.75) and (6.76) in (6.73) concludes the proof. For $p \le 2^n$, the bound also holds using the same reasoning and the convention that for $n_1, n_2 \in \mathbb{N}$, $n_1 > n_2$, ϑ_{n_1,n_2} is equal to 0.

6.8.12 Proof of Theorem 6.16

The strategy is nearly the same as in the proof of Theorem 6.10. However as f is no longer Lipschitz, we cannot use Lemma 6.27. But using the following result we observe that $Q_\gamma^{n,\ell} f$ for $n,\ell \ge 1$, $n < \ell$, is still Lipschitz.

Theorem 6.37. *Assume **H** 10 and **H** 11. Let $(\gamma_k)_{k\ge 1}$ be a nonincreasing sequence satisfying $\gamma_1 < 2/(m+L)$. Then for all $x, y \in \mathbb{R}^d$ and $n,\ell \ge 1$, $n < \ell$, we have*

$$\|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \le 1 - 2\boldsymbol{\Phi}\left(-\|x-y\|/(8\,\Lambda_{n,\ell})^{1/2}\right) ,$$

where $\Lambda_{n,\ell}$ is given in (6.32).

Proof. By (6.4) for all x, y and for all $k \ge 1$, we have

$$\|x - \gamma_k \nabla U(x) - y + \gamma_k \nabla U(y)\| \le (1-\kappa\gamma_k)^{1/2}\|x-y\| .$$

Let $n,\ell \ge 1$, $n < \ell$, then applying Theorem 6.48, we get

$$\|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \le 1 - 2\boldsymbol{\Phi}\left(-\|x-y\|/(8\,\tilde{\Lambda}_{n,\ell})^{1/2}\right) ,$$

where

$$\tilde{\Lambda}_{n,\ell} = \sum_{i=n}^{\ell}\gamma_i\left\{\prod_{j=n}^{i}(1-\kappa\gamma_j)\right\}^{-1} .$$

To conclude the proof, we now show that $\tilde{\Lambda}_{n,\ell} = \Lambda_{n,\ell}$:

$$\begin{aligned}\tilde{\Lambda}_{n,\ell} &= \kappa^{-1}\sum_{i=n}^{\ell}(1-(1-\kappa\gamma_i))\left\{\prod_{j=n}^{i}(1-\kappa\gamma_j)\right\}^{-1} \\ &= \kappa^{-1}\left[\sum_{i=n}^{\ell}\left\{\prod_{j=n}^{i}(1-\kappa\gamma_j)\right\}^{-1} - \sum_{i=n}^{\ell}\left\{\prod_{j=n}^{i-1}(1-\kappa\gamma_j)\right\}^{-1}\right] = \Lambda_{n,\ell} .\end{aligned}$$

□

Corollary 6.38. *Assume **H** 10 and **H** 11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence satisfying $\gamma_1 < 2/(m+L)$. Then for all $f \in \mathbb{F}_b(\mathbb{R}^d)$ and $\ell, n \geq 1$, $n < \ell$, $Q_\gamma^{n,\ell} f$ is a Lipschitz function with*

$$\left\| Q_\gamma^{n,\ell} f \right\|_{\mathrm{Lip}} \leq 2\|f\|_\infty / \Lambda_{n,\ell}^{1/2} ,$$

where $\Lambda_{n,\ell}$ is given in (6.32).

Proof. Let $f \in \mathbb{F}_b(\mathbb{R}^d)$ and $\ell > n \geq 1$. For all $x, y \in \mathbb{R}^d$ by definition of the total variation distance and Theorem 6.37, we have

$$\begin{aligned} \left| Q_\gamma^{n,\ell} f(x) - Q_\gamma^{n,\ell} f(y) \right| &\leq 2\|f\|_\infty \|\delta_x Q_\gamma^{n,\ell} - \delta_y Q_\gamma^{n,\ell}\|_{\mathrm{TV}} \\ &\leq 2\|f\|_\infty \left\{ 1 - 2\boldsymbol{\Phi}\left(- \|x-y\| / (8\,\Lambda_{n,\ell})^{1/2} \right) \right\} , \end{aligned}$$

Using that for all $s > 0$, $1/2 - \boldsymbol{\Phi}(-s) \leq (2\pi)^{-1/2} s$ concludes the proof. □

Let $N \geq 0$, $n \geq 1$ and $f \in \mathbb{F}_b(\mathbb{R}^d)$. We will use again the decomposition of $\mathrm{Var}_x\left\{ \hat{\pi}_n^N(f) \right\}$ given in (6.66). We can not directly apply the Poincaré inequality since the functions $\Phi_{n,k}^N$, for $k \in \{N+1, \cdots, N+n\}$ defined in (6.63), have a component which is not Lipschitz continuous. But by Corollary 6.38 the other components are Lipschitz continuous and we just need to isolate the non-Lipschitz term to have a counterpart of Lemma 6.28 for $f \in \mathbb{F}_b(\mathbb{R}^d)$.

Lemma 6.39. *Assume **H** 10 and **H** 11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 2/(m+L)$. Let $N \geq 0$ and $n \geq 1$. Then for all $y \in \mathbb{R}^d$, $f \in \mathbb{F}_b(\mathbb{R}^d)$, and $k \in \{N, \ldots, N+n-1\}$,*

$$\begin{aligned} R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2 (y) &\leq 2\|f\|_\infty \gamma_{k+2} (\Gamma_{N+2,N+n+1})^{-2} \\ &\quad + 8\gamma_{k+1} \|f\|_\infty^2 \left\{ \sum_{i=k+2}^{N+n} \omega_{i,n}^N / \Lambda_{k+2,i}^{1/2} \right\}^2 , \end{aligned}$$

where $\Phi_{n,k+1}^N$ is given by (6.63) and $\Lambda_{k+2,i}$ in (6.32) for all $i \in \{k+2, \cdots, N+n\}$.

Proof. Let $k \in \{N, \ldots, N+n-1\}$. By definition (6.67), $\Phi_{n,k}^N = \omega_{k+1,n}^N f + \tilde{\Phi}_{n,k}^N$, where $\tilde{\Phi}_{n,k}^N = \sum_{i=k+2}^{N+n} \omega_{i,n}^N Q_\gamma^{k+2,i} f$. Using that f is bounded and the Young inequality, we get

$$\begin{aligned} R_{\gamma_{k+1}} \left\{ \Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \Phi_{n,k+1}^N(y) \right\}^2 (y) &\leq 2\gamma_{k+2} \|f\|_\infty (\Gamma_{N+2,N+n+1})^{-2} \\ &\quad + 2R_{\gamma_{k+1}} \left\{ \tilde{\Phi}_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}} \tilde{\Phi}_{n,k+1}^N(y) \right\}^2 (y) . \end{aligned}$$

In addition by Corollary 6.38, we get

$$\left\| \tilde{\Phi}_{n,k}^N \right\|_{\mathrm{Lip}} \leq \sum_{i=k+2}^{N+n} \omega_{i,n}^N \left\| Q_\gamma^{k+2,i} f \right\|_{\mathrm{Lip}} \leq 2\, \|f\|_\infty \sum_{i=k+2}^{N+n} \omega_{i,n}^N / \Lambda_{k+2,i}^{1/2} .$$

Finally, the proof follows from Lemma 6.26. □

It remains to control the variance of Ψ_n^N under $\delta_x Q_\gamma^N$, where $\delta_x Q_\gamma^N$ is defined by (6.10) and Φ_n^N is given by (6.64). But using again that by (6.68) and Corollary 6.38, Ψ_n^N is a Lipschitz function and we have the following result, which is the counterpart of Corollary 6.30.

Lemma 6.40. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Let* $N \geq 0$ *and* $n \geq 1$*. Then for all* $x \in \mathbb{R}^d$ *and* $f \in \mathbb{F}_b(\mathbb{R}^d)$*,*

$$\mathrm{Var}_x\left\{\Psi_n^N(X_N)\right\} \leq 8\kappa^{-1}\|f\|_\infty^2 \left\{\sum_{i=N+1}^{N+n} \omega_{i,n}^N/\Lambda_{N+1,i}^{1/2}\right\}^2 ,$$

where Ψ_n^N *is given by* (6.64) *and* $\Lambda_{N+1,i}$ *in* (6.32) *for all* $i \in \{N+1,\cdots,N+n\}$*.*

Proof. By (6.68) and Corollary 6.38, Ψ_n^N is Lipschitz function with

$$\left\|\Psi_n^N\right\|_{\mathrm{Lip}} \leq \sum_{i=N+1}^{N+n} \omega_{i,n}^N \left\|Q_\gamma^{N+1,i} f\right\|_{\mathrm{Lip}} \leq 2\|f\|_\infty \sum_{i=N+1}^{N+n} \omega_{i,n}^N/\Lambda_{N+1,i}^{1/2} .$$

Finally, the proof follows from Lemma 6.29. □

Proof of Theorem 6.16. The proof follows from combining the bounds given by Lemma 6.39 and Lemma 6.40 in (6.66). □

6.8.13 Proof of Theorem 6.17

For the proof, we will use the decomposition (6.69) again but combined with the decomposition of $\Phi_{n,k+1}^N$ into a Lipschitz component and a bounded measurable component as it is done in the proof of Lemma 6.39.

Lemma 6.41. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Let* $N \geq 0$ *and* $n \geq 1$*. Then for all* $k \in \{N+1,\ldots,N+n\}$*,* $y \in \mathbb{R}^d$*,* $f \in \mathbb{F}_b(\mathbb{R}^d)$ *and* $\lambda > 0$*,*

$$\begin{aligned}&R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\Phi_{n,k+1}^N(y)\}\right)\right\}(y)\\&\quad\leq \exp\left\{\lambda\|f\|_\infty\gamma_{k+2}(\Gamma_{N+2,N+n+1})^{-2} + 4\gamma_{k+1}\lambda^2\|f\|_\infty^2\left(\sum_{i=k+2}^{N+n}\omega_{i,n}^N/\Lambda_{k+2,i}^{1/2}\right)^2\right\} ,\end{aligned}$$

where $\Phi_{n,k+1}^N$ *is given by* (6.63) *and* $\Lambda_{k+2,i}$ *in* (6.32) *for all* $i \in \{k+2,\cdots,N+n\}$*.*

Proof. Let $k \in \{N,\ldots,N+n-1\}$. By definition (6.67), $\Phi_{n,k}^N = \omega_{k+1,n}^N f + \tilde{\Phi}_{n,k}^N$, where $\tilde{\Phi}_{n,k}^N = \sum_{i=k+2}^{N+n}\omega_{i,n}^N Q_\gamma^{k+2,i} f$. Using that f is bounded and the Young inequality, we get

$$\begin{aligned}&R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\Phi_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\Phi_{n,k+1}^N(y)\}\right)\right\}(y)\\&\quad\leq \mathrm{e}^{\lambda\|f\|_\infty\gamma_{k+2}(\Gamma_{N+2,N+n+1})^{-2}} R_{\gamma_{k+1}}\left\{\exp\left(\lambda\{\tilde{\Phi}_{n,k+1}^N(\cdot) - R_{\gamma_{k+1}}\tilde{\Phi}_{n,k+1}^N(y)\}\right)\right\}(y)\end{aligned}$$

In addition by Corollary 6.38, we get

$$\left\|\tilde{\Phi}_{n,k}^{N}\right\|_{\mathrm{Lip}} \leq \sum_{i=k+2}^{N+n} \omega_{i,n}^{N} \left\|Q_\gamma^{k+2,i} f\right\|_{\mathrm{Lip}} \leq 2 \left\|f\right\|_\infty \sum_{i=k+2}^{N+n} \omega_{i,n}^{N} / \Lambda_{k+2,i}^{1/2} .$$

Finally, Lemma 6.31 concludes the proof. □

It remains to control the Laplace transform of Ψ_n^N under $\delta_x Q_\gamma^N$, where $\delta_x Q_\gamma^N$ is defined by (6.10). For this, using again that by (6.68) and Corollary 6.38, Ψ_n^N is a Lipschitz function. Therefore Lemma 6.33 shows the following result which is the analogue of Corollary 6.34 for measurable bounded function f.

Lemma 6.42. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Let* $N \geq 0$ *and* $n \geq 1$*. Then for all* $x \in \mathbb{R}^d$*,* $f \in \mathbb{F}_b(\mathbb{R}^d)$ *and* $\lambda > 0$

$$\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\Psi_n^N(X_n) - \mathbb{E}_x[\Psi_n^N(X_n)]\}}\right] \leq \exp\left\{4\kappa^{-1}\lambda^2 \left\|f\right\|_\infty^2 \left(\sum_{i=N+1}^{N+n} \omega_{i,n}^N / \Lambda_{N+1,i}^{1/2}\right)^2\right\} ,$$

where Ψ_n^N *is given by* (6.64) *and* $\Lambda_{N+1,i}$ *in* (6.32) *for all* $i \in \{N+1, \cdots, N+n\}$*.*

The Laplace transform of $\hat{\pi}_n^N(f)$ can be explicitly bounded using Lemma 6.41 and Lemma 6.42 in (6.69).

Proposition 6.43. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11. Let* $(\gamma_k)_{k\geq 1}$ *be a nonincreasing sequence with* $\gamma_1 \leq 2/(m+L)$*. Then for all* $N \geq 0$*,* $n \geq 1$*,* $x \in \mathbb{R}^d$*,* $f \in \mathbb{F}_b(\mathbb{R}^d)$ *and* $\lambda > 0$*:*

$$\mathbb{E}_x\left[\mathrm{e}^{\lambda\{\hat{\pi}_n^N(f) - \mathbb{E}_x[\hat{\pi}_n^N(f)]\}}\right] \leq \exp\left(\lambda\|f\|_\infty(\Gamma_{N+2,N+n+1})^{-1} + 4\lambda^2 \left\|f\right\|_\infty^2 u_{N,n}^{(4)}(\gamma)\right) ,$$

where $u_{N,n}^{(4)}(\gamma)$ *is defined in* (6.33)*.*

Proof of Theorem 6.17. Using the Markov inequality and Proposition 6.43, for all $\lambda > 0$, we have:

$$\mathbb{P}_x\left[\hat{\pi}_n^N(f) \geq \mathbb{E}_x[\hat{\pi}_n^N(f)] + r\right] \leq \exp\left(-\lambda r + \lambda\|f\|_\infty(\Gamma_{N+2,N+n+1})^{-1} + 4\lambda^2 \left\|f\right\|_\infty^2 u_{N,n}^{(4)}(\gamma)\right) .$$

Then the result follows from taking $\lambda = (r - \left\|f\right\|_\infty (\Gamma_{N+2,N+n+1})^{-1})/(8 \left\|f\right\|_\infty^2 u_{N,n}^{(4)}(\gamma))$. □

6.8.14 Proof of Proposition 6.18

Lemma 6.44. *Assume* $\boldsymbol{H}$*10 and* $\boldsymbol{H}$*11 and let* $p \in \mathbb{N}$*,* $p \geq 1$*.*

(i) For all $t \geq 0$ *and* $x \in \mathbb{R}^d$*,* $\mathbb{E}_x\left[\left\|Y_t - x^\star\right\|^{2p}\right] \leq \sum_{k=0}^{p} a_{k,p} \mathrm{e}^{-2kmt} \left\|x - x^\star\right\|^{2k}$ *where for* $k \in \{0, \cdots, p\}$*,* $a_{k,p}$ *is given in* (6.34)*.*

(ii) *For all $t \geq 0$ and $x \in \mathbb{R}^d$, $\mathbb{E}_x\left[\|Y_t - x\|^{2p}\right] \leq \sum_{k=0}^{2} \tilde{a}_{k,p}(x) t^{k+p}$ where for $k = 0, 1, 2$, $\tilde{a}_{k,p} : \mathbb{R}^d \to \mathbb{R}$ are defined in (6.36).*

Proof. (i) Denote for all $x \in \mathbb{R}^d$ and $p \geq 1$, by $V_p(x) = \|x - x^\star\|^{2p}$. The proof is by induction on $p \geq 1$. Lemma 6.21-(i) shows that the result holds for $p = 1$. Assume it is true for $p - 1$, $p \geq 2$. We have using **H**11 and $\nabla U(x^\star) = 0$,

$$\begin{aligned}\mathscr{A} V_p(x) &= -2p \|x - x^\star\|^{2(p-1)} \langle \nabla U(x), x - x^\star \rangle + 2p(d + 2(p-1)) \|x - x^\star\|^{2(p-1)} \\ &\leq -2pm \|x - x^\star\|^{2p} + 2p \|x - x^\star\|^{2(p-1)} (d + 2(p-1)) \;.\end{aligned}$$

By [MT93b, Theorem 1.1], denoting by $v_p(t, x) = P_t V_p(x)$, we have

$$v_p(t, x) \leq V_p(x) - 2pm \int_0^t v_p(s, x) \mathrm{d}s + 2p(d + 2(p-1)) \int_0^t v_{p-1}(s, x) \mathrm{d}s \;.$$

By Grönwall's inequality, we get for all $x \in \mathbb{R}^d$ and $t \geq 0$,

$$v_p(t, x) \leq \mathrm{e}^{-2pmt} V_p(x) + 2p(d + 2(p-1)) \int_0^t \mathrm{e}^{-2pm(t-s)} v_{p-1}(s, x) \mathrm{d}s \;.$$

Using the induction hypothesis concludes the proof.

(ii) The proof is made by induction on p. For $p = 1$, the result holds by Lemma 6.21-(ii). Assume that it is true for $p-1$, $p \geq 2$. Denote for all $x, y \in \mathbb{R}^d$, $\tilde{V}_{x,p}(y) = \|y - x\|^{2p}$. By **H**11, the inequality for all $\epsilon > 0$, $z_1, z_2 \in \mathbb{R}^d$, $|\langle z_1, z_2 \rangle| \leq \epsilon \|z_1\|^2 + (4\epsilon)^{-1} \|z_2\|^2$, **H**10 and $\nabla U(x^\star) = 0$, for all $y \in \mathbb{R}^d$ we have:

$$\begin{aligned}\mathscr{A} \tilde{V}_{x,p}(y) &= -2p \|y - x\|^{2(p-1)} \langle \nabla U(y), y - x \rangle + 2p \left(d + 2(p-1)\right) \|y - x\|^{2(p-1)} \;, \\ &\leq -2pm \|y - x\|^{2p} - 2p \|y - x\|^{2(p-1)} \langle \nabla U(x), y - x \rangle + 2p \left(d + 2(p-1)\right) \|y - x\|^{2(p-1)} \;, \\ &\leq \left\{ pL^2/(2m) \|x - x^\star\|^2 + 2p \left(d + 2(p-1)\right) \right\} \|y - x\|^{2(p-1)} \;.\end{aligned}$$

Denote by $\tilde{v}_{x,p}(x, t) = P_t \tilde{V}_{x,p}(x)$. Using [MT93b, Theorem 1.1] and $\tilde{v}_{x,p}(x, 0) = 0$, we get

$$\tilde{v}_{x,p}(x, t) \leq \left\{ pL^2/(2m) \|x - x^\star\|^2 + 2p \left(d + 2(p-1)\right) \right\} \int_0^t \tilde{v}_{x,p-1}(x, s) \mathrm{d}s \;,$$

and the induction hypothesis concludes the proof.

□

Proof of Proposition 6.18. The proof follows the same line as the one of Theorem 6.1-(6.7) using Lemma 6.44-(i) in place of Lemma 6.21-(i) and is omitted. □

6.8.15 Proof of Theorem 6.19

Lemma 6.45. *Assume **H**10 and **H**11. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$ and $p \in \mathbb{N}$, $p \geq 1$. Let $\zeta_0 \in \mathcal{P}_{2p}(\mathbb{R}^d \times \mathbb{R}^d)$, $(Y_t, \overline{Y}_t)_{t\geq 0}$ such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0 and given by (6.13). Then almost surely for all $n \geq 0$ and $\epsilon_1, \epsilon_2 > 0$,*

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^{2p}\right] \leq \{(1+\epsilon_2)(1-\gamma_{n+1}(\kappa - 2\epsilon_1)\}^p \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^{2p} \\ + (L\gamma_{n+1})^{2p}(1+\epsilon_2^{-1})^{p-1}(1/(4\epsilon_1) + \gamma_{n+1})^p \sum_{k=0}^{2} \tilde{a}_{k,p}(Y_{\Gamma_n})(k+p+1)^{-1}\gamma_{n+1}^k .$$

Proof. Let $n \geq 0$ and $\epsilon_1, \epsilon_2 > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. By Lemma 6.23-(6.48) and the inequality for all $a, b \in \mathbb{R}$, $(a+b)^p \leq (1+\epsilon_2)^p a^p + (1+\epsilon_2^{-1})^{p-1} b^p$, we have

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Theta_{n+1}\|^{2p}\right] \leq \{(1+\epsilon_2)(1-\gamma_{n+1}(\kappa - 2\epsilon_1))\}^p \|\Theta_n\|^{2p} \\ + (1+\epsilon_2^{-1})^{p-1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\{(2\gamma_{n+1} + (2\epsilon)^{-1})\int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \mathrm{d}s\right\}^p\right] . \quad (6.77)$$

By **H**10, Jensen's inequality, the Markov property of $(Y_t)_{t\geq 0}$ and Lemma 6.21-(ii), we have

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\{\int_{\Gamma_n}^{\Gamma_{n+1}} \|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \mathrm{d}s\right\}^p\right] \leq \gamma_{n+1}^{p-1}\int_{\Gamma_n}^{\Gamma_{n+1}} \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^{2p}\right] \mathrm{d}s , \\ \leq L^{2p}\sum_{k=0}^{2} \tilde{a}_{k,p}(Y_{\Gamma_n})(k+p+1)^{-1}\gamma_{n+1}^{k+2p} .$$

The proof is then concluded plugging this bound in (6.52) . □

Proof of Theorem 6.19. Let ζ_0 be an optimal transference plan of μ_0 and π and $n \geq 0$. Let $(Y_t, \overline{Y}_t)_{t\geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (6.13). By definition of W_{2p} and since for all $t \geq 0$, π is invariant for P_t, $W_{2p}^{2p}(\mu_0 Q^n, \pi) \leq \mathbb{E}_{\zeta_0}\left[\left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^{2p}\right]$. The proof then follows from applying recursively Lemma 6.45 with $\epsilon_1 = \kappa/4$ and $\epsilon_2 = \kappa\gamma_i/2$ for $i = n, \cdots, 1$, and using that for all $k \in \mathbb{N}$, Y_{Γ_k} is distributed according to π. □

6.8.16 Proof of Theorem 6.20

Lemma 6.46. *Assume **H**10, **H**11 and **H**12. Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence with $\gamma_1 \leq 1/(m+L)$ and $p \in \mathbb{N}$, $p \geq 1$. and $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. Let $(Y_t, \overline{Y}_t)_{t\geq 0}$ be defined by (6.13) such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0. Then for all $n \geq 0$ and $\epsilon > 0$,* *almost surely*

$$
\begin{aligned}
&\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right] \leq \{(1+\epsilon_2)(1-\gamma_{n+1}(\kappa - 2\epsilon_1))\}^p \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^{2p} \\
&+ 3^{p-1}(1+\epsilon_2^{-1})^{p-1}\left((\gamma_{n+1}^{3p}/(2p+1))(2\epsilon_1)^{-p}\left\{2^{2p-1}L^{4p}\sum_{k=0}^{p} a_{k,p}\left\|Y_{\Gamma_n} - x^\star\right\|^{2k} + (d\tilde{L}^2)^p/2\right\}\right. \\
&\left.+(2\gamma_{n+1}^3 L^2)^p \sum_{k=0}^{2} \frac{\tilde{a}_{k,p}(Y_{\Gamma_n})\gamma_{n+1}^k}{k+p+1} + 2^{p/2}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left|\int_{\Gamma_n}^{\Gamma_{n+1}}\left\langle \Theta_n, \int_{\Gamma_n}^{s} \nabla^2 U(Y_u)\mathrm{d}B_u\right\rangle \mathrm{d}s\right|^p\right]\right) .
\end{aligned}
$$

Proof. Let $n \geq 0$ and $\epsilon_1, \epsilon_2 > 0$, and set $\Theta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$. Using (6.57)-(6.51), **H**12 and $|\langle a, b\rangle| \leq \epsilon_1\|a\|^2 + (4\epsilon_1)^{-1}\|b\|^2$, we have

$$
\begin{aligned}
\|\Theta_{n+1}\|^2 \leq \{1-\gamma_{n+1}(\kappa - 2\epsilon_1)\}\|\Theta_n\|^2 &+ 2\gamma_{n+1}\int_{\Gamma_n}^{\Gamma_{n+1}}\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \mathrm{d}s \\
&+ \sqrt{2}\left|\int_{\Gamma_n}^{\Gamma_{n+1}}\left\langle \Theta_n, \int_{\Gamma_n}^{s}\nabla^2 U(Y_u)\mathrm{d}B_u\right\rangle \mathrm{d}s\right| \\
&+ (2\epsilon_1)^{-1}\int_{\Gamma_n}^{\Gamma_{n+1}}\left\|\int_{\Gamma_n}^{s}\left\{\nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right\|^2 \mathrm{d}s .
\end{aligned}
$$

Therefore by the Hölder inequality and the inequality for all $a, b \geq 0$, $(a+b)^p \leq ((1+\epsilon_2)a)^p + (1+\epsilon_2^{-1})^{p-1}b^p$, we get

$$
\begin{aligned}
\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Theta_{n+1}\|^{2p}\right] \leq &\left\{(1+\epsilon_2)(1-\gamma_{n+1}(\kappa - 2\epsilon_1))\|\Theta_n\|^2\right\}^p \\
&+ 3^{p-1}(1+\epsilon_2^{-1})^{p-1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[A_1^p + (2\gamma_{n+1})^p A_2^p + (2\epsilon_1)^{-p}A_3^p\right] ,
\end{aligned}
\tag{6.78}
$$

where

$$
A_1 = \sqrt{2}\left|\int_{\Gamma_n}^{\Gamma_{n+1}}\left\langle \Theta_n, \int_{\Gamma_n}^{s}\nabla^2 U(Y_u)\mathrm{d}B_u\right\rangle \mathrm{d}s\right| \quad , \quad A_2 = \int_{\Gamma_n}^{\Gamma_{n+1}}\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2 \mathrm{d}s ,
$$
$$
A_3 = \int_{\Gamma_n}^{\Gamma_{n+1}}\left\|\int_{\Gamma_n}^{s}\left\{\nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u)\right\}\mathrm{d}u\right\|^2 \mathrm{d}s .
$$

We successively bound $\mathbb{E}^{\mathcal{F}_{\Gamma_n}}[A_i^p]$ for $i = 2, 3$. By **H**10, Jensen's inequality, the Markov property of $(Y_t)_{t\geq 0}$ and Lemma 6.44-(ii), we have

$$
\begin{aligned}
\mathbb{E}^{\mathcal{F}_{\Gamma_n}}[A_2^p] \leq L^{2p}\gamma_{n+1}^{p-1}\int_{\Gamma_n}^{\Gamma_{n+1}}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|Y_s - Y_{\Gamma_n}\|^{2p}\right]\mathrm{d}s \\
\leq L^{2p}\gamma_{n+1}^{2p}\sum_{k=0}^{2}\tilde{a}_{k,p}(Y_{\Gamma_n})(k+p+1)^{-1}\gamma_{n+1}^k .
\end{aligned}
\tag{6.79}
$$

Also by (6.17), we get using $\nabla U(x^\star)=0$, Jensen's inequality and Fubini's theorem,

$$
\begin{aligned}
A_3^p &\leq \gamma_{n+1}^{p-1} \int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)^{2p-1} \int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| \left\{ \nabla^2 U(Y_u)\nabla U(Y_u) + (1/2)\vec{\Delta}(\nabla U)(Y_u) \right\} \right\|^{2p} \right] \mathrm{d}u \, \mathrm{d}s \\
&\leq 2^{2p-1}\gamma_{n+1}^{p-1} \int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)^{2p-1} \int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\left\| \nabla^2 U(Y_u)\nabla U(Y_u) \right\|^{2p} + 2^{-2p} \left\| \vec{\Delta}(\nabla U)(Y_u) \right\|^{2p} \right] \mathrm{d}u \, \mathrm{d}s \\
&\leq 2^{2p-1}\gamma_{n+1}^{p-1} \int_{\Gamma_n}^{\Gamma_{n+1}} (s-\Gamma_n)^{2p-1} \int_{\Gamma_n}^{s} L^{4p} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|Y_u - x^\star\|^{2p} \right] \mathrm{d}u \, \mathrm{d}s + \gamma_{n+1}^{3p} (d\tilde{L}^2)^p / (2(2p+1)) \, .
\end{aligned}
\tag{6.80}
$$

By Lemma 6.44-(i), the Markov property and for all $t \geq 0$, $1 - \mathrm{e}^{-t} \leq t$, we have for all $s \in [\Gamma_n, \Gamma_{n+1}]$,

$$
\int_{\Gamma_n}^{s} \mathbb{E}^{\mathcal{F}_{\Gamma_n}} \left[\|Y_u - x^\star\|^{2p} \right] \mathrm{d}u \leq \sum_{k=0}^{p} a_{k,p} (2km)^{-1} (1 - \mathrm{e}^{-2km(s-\Gamma_n)}) \|Y_{\Gamma_n} - x^\star\|^{2k} \, .
$$

Using this inequality in (6.80) and for all $t \geq 0$, $1 - \mathrm{e}^{-t} \leq t$, we get

$$
A_3^p \leq (\gamma_{n+1}^{3p}/(2p+1)) \left\{ 2^{2p-1} L^{4p} \sum_{k=0}^{p} a_{k,p} \|Y_{\Gamma_n} - x^\star\|^{2k} + (d\tilde{L}^2)^p/2 \right\} \, .
$$

Combining this bound and (6.79) in (6.78) concludes the proof. ☐

Proof of Theorem 6.20. Let ζ_0 be an optimal transference plan of μ_0 and π and $n \geq 0$. Let $(Y_t, \overline{Y}_t)_{t \geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (6.13). By definition of W_{2p} and since for all $t \geq 0$, π is invariant for P_t, $W_{2p}^{2p}(\mu_0 Q^n, \pi) \leq \mathbb{E}_{\zeta_0}\left[\left\| Y_{\Gamma_n} - \overline{Y}_{\Gamma_n} \right\|^{2p} \right]$. Applying recursively Lemma 6.46 with $\epsilon_1 = \kappa/4$ and $\epsilon_2 = \kappa\gamma_i/2$ for $i = n, \cdots, 1$, and using that for all $k \in \mathbb{N}$, Y_{Γ_k} is distributed according to π.

$$
W_{2p}^{2p}(\mu_0 Q^n, \pi) \leq u_n^{(1,p)}(\gamma) W_{2p}^{2p}(\mu_0, \pi) + \sum_{k=1}^{n} \left\{ \prod_{i=k+1}^{n} (1 - \kappa\gamma_i/4)^p \right\} (6\kappa^{-1}\gamma_k^{-1})^{p-1} (A_1 + A_2 + A_3) \, ,
\tag{6.81}
$$

where

$$
A_1 = 2^{p/2} \mathbb{E}\left[\left| \int_{\Gamma_{k-1}}^{\Gamma_k} \left\langle Y_{\Gamma_{k-1}} - \overline{Y}_{\Gamma_{k-1}}, \int_{\Gamma_{k-1}}^{s} \nabla^2 U(Y_u) \mathrm{d}B_u \right\rangle \mathrm{d}s \right|^p \right] \, , \quad A_2 = (2\gamma_k^3 L^2)^p \sum_{l=0}^{2} \frac{U_{l,p}\gamma_k^l}{l+p+1} \, ,
$$
$$
A_3 = (\gamma_k^{3p}/(2p+1))(2\kappa^{-1})^p \left\{ 2^{2p-1} L^{4p} a_{0,p} \sum_{l=0}^{p} a_{l,p} + (d\tilde{L}^2)^p/2 \right\} \, .
$$

We now give a bound on A_1 which will concludes the proof. By Jensen's inequality, the

Burkholder-Davis-Gundy inequality [CK91, Theorem 1] and **H**10, we have

$$\begin{aligned}
A_1 &\leq 2^{p/2}\gamma_k^{p-1}\int_{\Gamma_{k-1}}^{\Gamma_k}\mathbb{E}\left[\left|\int_{\Gamma_{k-1}}^{s}\left\langle Y_{\Gamma_{k-1}}-\overline{Y}_{\Gamma_{k-1}},\nabla^2 U(Y_u)\mathrm{d}B_u\right\rangle\right|^p\right]\mathrm{d}s \\
&\leq 2^{3p/2}p^{p/2}\gamma_k^{p-1}\int_{\Gamma_{k-1}}^{\Gamma_k}\mathbb{E}\left[\left(\int_{\Gamma_{k-1}}^{s}L^2 d\left\|Y_{\Gamma_{k-1}}-\overline{Y}_{\Gamma_{k-1}}\right\|^2\mathrm{d}u\right)^{p/2}\right] \\
&\leq (2^{3/2}p^{1/2}\gamma_k^{3/2}Ld^{1/2})^p(p/2+1)^{-1}\mathbb{E}^{p/(2(p-1))}\left[\left\|Y_{\Gamma_{k-1}}-\overline{Y}_{\Gamma_{k-1}}\right\|^{2(p-1)}\right]
\end{aligned}$$

□

6.9 Contraction results in total variation for some functional autoregressive models

In this section, we consider functional autoregressive models of the form: for all $k \geq 0$

$$\mathrm{X}_{k+1} = h_{k+1}(\mathrm{X}_k) + \sigma_{k+1}\mathrm{Z}_{k+1}\ , \tag{6.82}$$

where $(Z_k)_{k\geq 1}$ is a sequence of i.i.d. d dimensional standard Gaussian random variables, $(\sigma_k)_{k\geq 1}$ is a sequence of positive real numbers and $(h_k)_{k\geq 1}$ are a sequence of measurable map from $\mathbb{R}^d$ to $\mathbb{R}^d$ which satisfies the following assumption:

AR3. *For all $k \geq 1$, there exists $\varpi_k \in [0,1]$ such that h_k is $1-\varpi_k$-Lipschitz, i.e. for all $x, y \in \mathbb{R}^d$, $\|h_k(x)-h_k(y)\| \leq (1-\varpi_k)\|x-y\|$.*

The sequence $(\mathrm{X}_k)_{k\in\mathbb{N}}$ defines an inhomogeneous Markov chain associated with the sequence of Markov kernel $(\mathsf{P}_k)_{k\geq 1}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given for all $x \in \mathbb{R}^d$ and $\mathtt{A} \in \mathbb{R}^d$ by

$$\mathsf{P}_k(x,\mathtt{A}) = \int_{\mathtt{A}} \exp\left(\|y-h_k(x)\|^2/(2\sigma_k^2)\right)\mathrm{d}y\ . \tag{6.83}$$

We denote for all $n \geq 1$ by Q^n the marginal laws of the sequence $(\mathrm{X}_k)_{k\geq 1}$ and given by

$$\mathsf{Q}^n = \mathsf{P}_1\cdots\mathsf{P}_n\ . \tag{6.84}$$

In this section we are interested in showing that for all $x, y \in \mathbb{R}^d$, the sequence $\{\|\delta_x\mathsf{Q}^n-\delta_y\mathsf{Q}^n\|_{\mathrm{TV}},\ k \in 1\}$ goes to 0 with an explicit rate depending on the assumption on the sequence $(h_k)_{k\geq 1}$, which in any cases does not depend on the dimension d. In addition these rates are optimal as we will see. For this we will consider an appropriate coupling which is based on a coupling for Gaussian random walks proposed in [BDJ98, Section 3.3]. Let $x, y \in \mathbb{R}^d$. We consider for all $k \geq 1$ the following coupling $(\mathrm{X}_1^{(1)}, \mathrm{X}_1^{(2)})$ between $\mathsf{P}_k(x,\cdot)$ and $\mathsf{P}_k(y,\cdot)$. Define the function E and e from $\mathbb{R}^d\times\mathbb{R}^d$ to $\mathbb{R}^d$ by

$$\mathrm{E}_k(x,y) = h_k(y)-h_k(x)\ , \qquad \mathrm{e}_k(x,y) = \begin{cases}\mathrm{E}_k(x,y)/\left\|\mathrm{E}_k(x,y)\right\| & \text{if } \mathrm{E}_k(x,y)\neq 0\\ 0 & \text{otherwise}\ ,\end{cases} \tag{6.85}$$

and let Z be a standard d dimensional Gaussian random variable. If $\mathrm{E}_k(x,y)=0$, then we set

$$\mathrm{X}_1^{(1)} = h_k(x) + \sigma_k \mathrm{Z}$$
$$\mathrm{X}_1^{(2)} = h_k(y) + \sigma_k \mathrm{Z}\ .$$

If $\mathrm{E}_k(x,y) \neq 0$, consider the following coupling

$$\begin{aligned}\mathrm{X}_1^{(1)} &= h_k(x) + \sigma_k Z \\ \mathrm{X}_1^{(2)} &= h_k(y) + \mathbb{1}_{\mathcal{M}}\left\{\sigma_k \mathrm{Z} - \mathrm{E}_k(x,y)\right\} + \mathbb{1}_{\mathcal{M}^c}\left\{\sigma_k(\mathrm{Id} - 2\mathrm{e}_k(x,y)\mathrm{e}_k(x,y)^{\mathrm{T}})\mathrm{Z}\right\}\ ,\end{aligned}$$

where

$$\begin{aligned}\mathcal{M} &= \{\mathrm{U} \leq \min(1, \alpha(x,y,\mathrm{Z}))\} \\ \alpha(\tilde{x},\tilde{y},\tilde{z}) &= \frac{\boldsymbol{\varphi}_{\sigma^2}(\|\mathrm{E}_k(\tilde{x},\tilde{y})\| - \sigma_k\langle \mathrm{e}_k(\tilde{x},\tilde{y}),\tilde{z}\rangle)}{\boldsymbol{\varphi}_{\sigma^2}(\sigma_k\langle \mathrm{e}_k(\tilde{x},\tilde{y}),\tilde{z}\rangle)}\ ,\ \text{for all } \tilde{x},\tilde{y},\tilde{z} \in \mathbb{R}^d\ ,\end{aligned} \tag{6.86}$$

U is uniform random variable on $[0,1]$ independent of Z and $\boldsymbol{\varphi}_{\sigma^2}$ is the standard one dimensional zero-mean Gaussian probability density with variance σ_k^2. In other words, when $\mathrm{E}_k(x,y)$ is not zero then with probability $\min(1,\alpha(x,y,\mathrm{Z}))$, $\mathrm{X}_1^{(1)} = \mathrm{X}_1^{(2)}$, and $\mathrm{X}_1^{(2)} = h_k(y) + \sigma_k(\mathrm{Id} - 2\mathrm{e}_k(x,y)\mathrm{e}_k(x,y)^{\mathrm{T}})\mathrm{Z}$ otherwise.

The laws of $(\mathrm{X}_1^{(1)},\mathrm{X}_1^{(2)})$ for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ defines the Markov kernel K_k on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$ given for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ and $\mathtt{A} \in \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned}\mathsf{K}((x,y),\mathtt{A}) &= \mathbb{1}_{\Delta}(h_k(x),h_k(y))(2\pi\sigma_k^2)^{-d/2}\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x})\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x} \\ &+ \frac{\mathbb{1}_{\Delta^c}(h_k(x),h_k(y))}{(2\pi\sigma_k^2)^{d/2}}\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\tilde{x},\tilde{x})\min\left[1,\alpha\left(x,y,\frac{\tilde{x}-h_k(x)}{\sigma_k}\right)\right]\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x} \\ &+ \frac{\mathbb{1}_{\Delta^c}(h_k(x),h_k(y))}{(2\pi\sigma_k^2)^{d/2}}\int_{\mathbb{R}^d}\mathbb{1}_{\mathtt{A}}(\tilde{x},\mathrm{F}_k(x,y,\tilde{x}))\left\{1-\min\left[1,\alpha\left(x,y,\frac{\tilde{x}-h_k(x)}{\sigma_k}\right)\right]\right\}\mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)}\mathrm{d}\tilde{x}\ ,\end{aligned} \tag{6.87}$$

where $\mathrm{F}_k : (\mathbb{R}^d)^3 \to \mathbb{R}^d$ is defined by for all $(\tilde{x},\tilde{y},\tilde{z})$, $\tilde{x} \neq \tilde{y}$,

$$\begin{aligned}\mathrm{F}_k(\tilde{x},\tilde{y},\tilde{z}) &= h_k(\tilde{y}) + \left(\mathrm{Id} - 2\mathrm{e}_k(\tilde{x},\tilde{y})\mathrm{e}_k(\tilde{x},\tilde{y})^{\mathrm{T}}\right)\tilde{z} \\ \Delta &= \left\{(\tilde{x},\tilde{y}) \in \mathbb{R}^d \times \mathbb{R}^d \,\middle|\, \tilde{x} = \tilde{y}\right\}\ .\end{aligned}$$

The following lemma shows that K is a Markovian coupling for P, in particular for all $x,y \in \mathbb{R}^d$, $(\mathrm{X}^{(1)},\mathrm{X}^{(2)})$ is a coupling of $\mathsf{P}(x,\cdot)$ and $\mathsf{P}(y,\cdot)$.

Lemma 6.47. *For all $x,y \in \mathbb{R}^d$ and $k \geq 1$, $\mathsf{K}_k((x,y),\cdot)$ is a transference plan of $\mathsf{P}_k(x,\cdot)$ and $\mathsf{P}_k(y,\cdot)$.*

Proof. The proof is postponed to Section 6.9.1 □

For all initial distribution μ_0 on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))$, $\widetilde{\mathbb{P}}_{\mu_0}$ and $\widetilde{\mathbb{E}}_{\mu_0}$ denote the probability and the expectation respectively, associated with the sequence of Markov kernels $(\mathsf{K}_k)_{k\geq 1}$ defined in (6.87) and μ_0 on the canonical space $((\mathbb{R}^d \times \mathbb{R}^d)^{\mathbb{N}}, (\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{N}})$, $\{(\mathrm{X}_i, \mathrm{Y}_i),\ i \in \mathbb{N}\}$ denotes the canonical process and $(\widetilde{\mathcal{F}}_i)_{i\in\mathbb{N}}$ the corresponding filtration. Then by Lemma 6.47 if $(\mathrm{X}_0, \mathrm{Y}_0) = (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, for all $k \geq 1$ $(\mathrm{X}_k, \mathrm{Y}_k)$ is coupling of $\delta_x \mathsf{Q}^k$ and $\delta_y \mathsf{Q}^k$. Therefore to have a bound on $\|\delta_x \mathsf{Q}^n - \delta_y \mathsf{Q}^n\|_{\mathrm{TV}}$ for $n \geq 1$, it suffices to have a bound on $\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n)$. It is the content of the proof of the main result of this section.

Theorem 6.48. *Assume* ***AR*** *3. Then for all $x, y \in \mathbb{R}^d$ and $n \geq 1$,*

$$\|\delta_x \mathsf{Q}^n - \delta_y \mathsf{Q}^n\|_{\mathrm{TV}} \leq \mathbb{1}_{\Delta^{\mathrm{c}}}((x,y)) \left\{ 1 - 2\boldsymbol{\Phi}\left(-\frac{\|x-y\|}{2\Xi_n^{1/2}} \right) \right\} ,$$

where the sequence $(\Xi_i)_{i\geq 1}$ is defined by for all $k \geq 1$

$$\Xi_k = \sum_{i=1}^{k} \sigma_i^2 \left\{ \prod_{j=1}^{i} (1-\varpi_j)^2 \right\}^{-1} .$$

Proof. The proof is postponed to Section 6.9.1. □

6.9.1 Proofs

Proof of Lemma 6.47. It is straightforward that $\mathsf{K}_k((x,y), \cdot \times \mathbb{R}^d) = \mathsf{P}_k(x, \cdot)$ for all $x, y \in \mathbb{R}^d$. In addition, for all $\mathtt{A} \in \mathcal{B}(\mathbb{R}^d)$, we have for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\begin{aligned}
&\mathsf{K}_k((x,y), \mathbb{R}^d \times \mathtt{A}) = \mathbb{1}_{\Delta}(h_k(x), h_k(y))(2\pi\sigma_k^2)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x}) \mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \qquad (6.88)\\
&+ \frac{\mathbb{1}_{\Delta^{\mathrm{c}}}(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{x}) \min\left[1, \alpha\left(x, y, \frac{\tilde{x}-h_k(x)}{\sigma_k}\right)\right] \mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x}\\
&+ \frac{\mathbb{1}_{\Delta^{\mathrm{c}}}(h_k(x), h_k(y))}{(2\pi\sigma_k^2)^{d/2}} \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\mathrm{F}_k(x,y,\tilde{x})) \left\{ 1 - \min\left[1, \alpha\left(x, y, \frac{\tilde{x}-h_k(x)}{\sigma_k}\right)\right]\right\} \mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} .
\end{aligned}$$

Therefore for all (x,y), $h_k(x) = h_k(y)$, we have that $\mathsf{K}((x,y), \mathbb{R}^d \times \cdot) = \mathsf{P}(y, \cdot)$. It remains to treat the case when $h_k(x) \neq h_k(y)$. Let $x, y \in \mathbb{R}^d$, $h_k(x) \neq h_k(y)$. By definition of α (6.86), we have for all $\tilde{x} \in \mathbb{R}^d$,

$$\alpha\left\{x, y, \frac{(\tilde{x}-h_k(x))}{\sigma_k^{-1}}\right\} = \frac{\boldsymbol{\varphi}_{\sigma^2}\left(\sigma_k\left\langle \mathrm{e}_k(x,y), \frac{h_k(y)-\tilde{x}}{\sigma_k^{-1}}\right\rangle\right)}{\boldsymbol{\varphi}_{\sigma^2}\left(\|\mathrm{E}_k(x,y)\| - \sigma_k\left\langle \mathrm{e}_k(x,y), \frac{h_k(y)-\tilde{x}}{\sigma_k^{-1}}\right\rangle\right)} = \alpha^{-1}\left\{x, y, \frac{\tilde{x}-h_k(y))}{\sigma_k^{-1}}\right\} . \qquad (6.89)$$

Since $(\mathrm{Id} - 2\mathrm{e}(x,y)\mathrm{e}(x,y)^{\mathrm{T}})$ is an orthogonal matrix, making the change of variable $\tilde{y} = h_k(y) + (\mathrm{Id} - 2\mathrm{e}(x,y)\mathrm{e}(x,y)^{\mathrm{T}})(\tilde{x} - h_k(x))$ and using that $\langle \mathrm{e}(x,y), \tilde{y} - h_k(y)\rangle =$

$-\langle \mathrm{e}(x,y), \tilde{x} - h_k(x) \rangle$ we get

$$\int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\mathrm{F}_k(x,y,\tilde{x})) \left\{ 1 - \min\left[1, \alpha\left(x, y, \frac{\tilde{x} - h_k(x)}{\sigma_k}\right)\right]\right\} \mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x}$$
$$= \int_{\mathbb{R}^d} \mathbb{1}_{\mathtt{A}}(\tilde{y}) \left\{ 1 - \min\left[1, \alpha\left(x, y, \frac{h_k(y) - \tilde{y}}{\sigma_k}\right)\right]\right\} \mathrm{e}^{-\|\tilde{y}-h_k(y)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{y} \; . \quad (6.90)$$

In addition using that $\|\tilde{x} - h_k(x)\|^2 = \|\tilde{x} - h_k(y)\|^2 - 2\langle h_k(y) - \tilde{x}, \mathrm{E}_k \rangle + \|\mathrm{E}_k\|^2$, we get

$$\min\left\{1, \alpha^{-1}\left(x, y, \frac{\tilde{x} - h_k(y)}{\sigma_k}\right)\right\} \mathrm{e}^{\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} =$$
$$\min\left\{1, \alpha\left(x, y, \frac{\tilde{x} - h_k(y)}{\sigma_k}\right)\right\} \mathrm{e}^{\|\tilde{x}-h_k(y)\|^2/(2\sigma_k^2)} \; . \quad (6.91)$$

Combining (6.89)-(6.90) and (6.91) in (6.87) implies that $\mathsf{K}_k((x,y), \mathbb{R}^d \times \mathtt{A}) = \mathsf{P}_k(y, \mathtt{A})$. □

Lemma 6.49. *For all $x, y \in \mathbb{R}^d$ and $k \geq 1$*

$$\mathsf{K}_k((x,y), \Delta) = 2\mathbf{\Phi}\left(-\frac{\|\mathrm{E}_k(x,y)\|}{2\sigma_k}\right) \; .$$

Proof. Let $x, y \in \mathbb{R}^d$. If $h_k(x) = h_k(y)$, then $\mathsf{K}_k((x,y), \Delta) = 1$. We now consider the case $h_k(x) \neq h_k(y)$ By (6.87), we have

$$\mathsf{K}_k((x,y), \Delta) = \left(2\pi\sigma_k^2\right)^{-d/2} \int_{\mathbb{R}^d} \min\left[1, \alpha\left(x, y, \frac{\tilde{x} - h_k(x)}{\sigma_k}\right)\right] \mathrm{e}^{-\|\tilde{x}-h_k(x)\|^2/(2\sigma_k^2)} \mathrm{d}\tilde{x} \; .$$

For all $\tilde{x} \in \mathbb{R}^d$, consider $\{\mathrm{w}_i, 1 \leq i \leq d\}$ the coordinates of $\tilde{x} - h_k(x)$ in an orthonormal basis whose first component is $\mathrm{e}_k(x,y)$, then we get

$$\mathsf{K}_k((x,y), \Delta) = \frac{1}{\left(2\pi\sigma_k^2\right)^{d/2}} \int_{\mathbb{R}^d} \min\left[1, \exp\left\{\left(\mathrm{w}_1^2 - (\|\mathrm{E}_k(x,y)\| - \mathrm{w}_1)^2\right)/(2\sigma_k^2)\right\}\right] \mathrm{e}^{-\sum_{i=1}^d \mathrm{w}_i^2/(2\sigma_k^2)} \mathrm{d}\mathrm{w}$$
$$= \left(2\pi\sigma_k^2\right)^{-1/2} \int_{\mathbb{R}} \min\left[\exp\left\{-\mathrm{w}_1^2/(2\sigma_k^2)\right\}, \exp\left\{-(\|\mathrm{E}_k(x,y)\| - \mathrm{w}_1)^2/(2\sigma_k^2)\right\}\right] \mathrm{d}\mathrm{w}_1$$
$$= 2\left(2\pi\sigma_k^2\right)^{-1/2} \int_{\|\mathrm{E}_k(x,y)\|/2}^{+\infty} \exp\left\{-\mathrm{w}_1^2/(2\sigma_k^2)\right\} \mathrm{d}\mathrm{w}_1 \; ,$$

which concludes the proof. □

Lemma 6.50. *For all $\varsigma, \mathrm{a} > 0$ and $t \in \mathbb{R}_+$, the following identity holds*

$$\int_{\mathbb{R}} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{1 - \min\left(1, \frac{\boldsymbol{\varphi}_{\varsigma^2}(t-y)}{\boldsymbol{\varphi}_{\varsigma^2}(y)}\right)\right\} \left\{1 - 2\mathbf{\Phi}\left(\frac{2y - t}{2\mathrm{a}}\right)\right\} \mathrm{d}y = 1 - 2\mathbf{\Phi}\left(-\frac{t}{2(\varsigma^2 + \mathrm{a}^2)^{1/2}}\right) \; .$$

Proof. Let $\varsigma, \mathrm{a} > 0$ and $t \in \mathbb{R}_+$. Denote by

$$I = \int_{\mathbb{R}} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 1 - \min\left(1, \frac{\boldsymbol{\varphi}_{\varsigma^2}(t-y)}{\boldsymbol{\varphi}_{\varsigma^2}(y)}\right) \right\} \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{2y-t}{2\mathrm{a}}\right) \right\} \mathrm{d}y \,.$$

Then,

$$\begin{aligned} I &= \int_0^{t/2} \{\boldsymbol{\varphi}_{\varsigma^2}(y) - \boldsymbol{\varphi}_{\varsigma^2}(t-y)\} \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{2y-t}{2\mathrm{a}}\right) \right\} \mathrm{d}y \\ &= \int_0^{t/2} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 2\boldsymbol{\Phi}\left(1 - \frac{2y-t}{2\mathrm{a}}\right) \right\} \mathrm{d}y - \int_{t/2}^{+\infty} \boldsymbol{\varphi}_{\varsigma^2}(y) \left\{ 1 - 2\boldsymbol{\Phi}\left(\frac{t-2y}{2\mathrm{a}}\right) \right\} \mathrm{d}y \,, \end{aligned} \tag{6.92}$$

where we use a simple change of variable for the last equality. Now to simplify the proof, we give a probabilistic interpretation of this two integrals. Let X and Y be two real Gaussian random variables with zero mean and variance a^2 and ς^2 respectively. Since for all $u \in \mathbb{R}_+$, $1 - 2\boldsymbol{\Phi}(-u/(2\mathrm{a})) = \mathbb{P}[|X| \le u]$, we have by (6.92)

$$I = \mathbb{P}\left(Y \le t/2, X+Y \le t/2, Y-X \le t/2\right) - \mathbb{P}\left(Y \ge t/2, X+Y \ge t/2, Y-X \ge t/2\right) \,.$$

Using that Y and $-Y$ have the same law in the second term, we get

$$\begin{aligned} I = \mathbb{P}\left(Y \le t/2, X+Y \le t/2, Y-X \le t/2\right) & \\ - \mathbb{P}\left(Y \le -t/2, X-Y \ge t/2, Y+X \le -t/2\right) & \qquad (6.93) \\ = I_1 + I_2 \,, & \qquad (6.94) \end{aligned}$$

where

$$\begin{aligned} I_1 &= \mathbb{P}\left(Y \le t/2, X+Y \le t/2, Y-X \le t/2, X \ge 0\right) \\ &\qquad - \mathbb{P}\left(Y \le -t/2, X-Y \ge t/2, Y+X \le -t/2, X \ge 0\right) \\ &= \mathbb{P}\left(|X+Y| \le t/2, X \ge 0\right) \,, \end{aligned} \tag{6.95}$$

and

$$\begin{aligned} I_2 = \mathbb{P}\left(Y \le t/2, X+Y \le t/2, Y-X \le t/2, X \le 0\right) & \\ - \mathbb{P}\left(Y \le -t/2, X-Y \ge t/2, Y+X \le -t/2, X \le 0\right) & \,. \end{aligned}$$

Using again that Y and $-Y$ have the same law in the two terms we have

$$\begin{aligned} I_2 &= \mathbb{P}\left(Y \ge -t/2, X-Y \le t/2, Y+X \ge -t/2, X \le 0\right) \\ &\qquad - \mathbb{P}\left(Y \ge t/2, X+Y \ge t/2, X-Y \le -t/2, X \le 0\right) \\ &= \mathbb{P}\left(|X+Y| \le t/2, X \le 0\right) \,. \end{aligned} \tag{6.96}$$

Combining (6.95), (6.96) in (6.94), we have $I = \mathbb{P}(|X+Y| \le t/2)$. The proof follows from the fact that $X+Y$ is a real Gaussian random variable with mean zero and variance $\mathrm{a}^2 + \varsigma^2$, since X and Y are independent. □

Proof of Theorem 6.48. Recall that by Lemma 6.47, for all $k \geq 1$, $(\mathrm{X}_k, \mathrm{Y}_k)$ is a coupling of $\delta_x \mathsf{Q}^k$ and $\delta_y \mathsf{Q}^k$, and therefore $\|\delta_x \mathsf{Q}^k - \delta_y \mathsf{Q}^k\|_{\mathrm{TV}} \leq \widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_k \neq \mathrm{Y}_k)$.

Define for all $k_1, k_2 \in \mathbb{N}$, $k_1, k_2 \geq 1$, $k_1 \leq k_2$,

$$\Xi_{k_1,k_2} = \sum_{i=k_1}^{k_2} \sigma_i^2 \left\{ \prod_{j=k_1}^{i} (1 - \varpi_j)^{-2} \right\} .$$

Let $n \geq 1$. We show by backward induction that for all $k \in \{0, \cdots, n-1\}$,

$$\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n) \leq \widetilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_k, \mathrm{Y}_k) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\left(\Xi_{k+1,n}\right)^{1/2}} \right\} \right] \right] , \tag{6.97}$$

Note that the inequality for $k = 0$ will conclude the proof.

Conditionning w.r.t. $\widetilde{\mathcal{F}}_{n-1}$, using that $\mathrm{X}_n \neq \mathrm{Y}_n$ implies that $\mathrm{X}_{n-1} \neq \mathrm{Y}_{n-1}$, the Markov property and Lemma 6.49, we get

$$\begin{aligned}
\widetilde{\mathbb{P}}_{(x,y)}(\mathrm{X}_n \neq \mathrm{Y}_n) &= \widetilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1}) \widetilde{\mathbb{E}}_{(\mathrm{X}_{n-1},\mathrm{Y}_{n-1})} \left[\mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_1, \mathrm{Y}_1) \right] \right] \\
&\leq \widetilde{\mathbb{E}}_{(x,y)} \left[\mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1}) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{E}_{n-1}(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})\|}{2\sigma_n} \right\} \right] \right]
\end{aligned}$$

Under **AR**3 and definition of E_n (6.85), $\|\mathrm{E}_n(\mathrm{X}_{n-1}, \mathrm{Y}_{n-1})\| \leq (1 - \varpi_n) \|\mathrm{X}_{n-1} - \mathrm{Y}_{n-1}\|$, and (6.97) holds for $k = n-1$.

Assume that (6.97) holds for $k \in \{1, \cdots, n-1\}$. By definition of the process $\{(\mathrm{X}_i, \mathrm{Y}_i),\ i \in \mathbb{N}\}$, if $\mathrm{X}_k \neq \mathrm{Y}_k$ necessarily $\mathrm{X}_{k-1} \neq \mathrm{Y}_{k-1}$ and

$$\mathrm{X}_k - \mathrm{Y}_k = \mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1}) + 2\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1}) \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k ,$$

where Z_k is a standard d-dimensional Gaussian random variable independent of $\widetilde{\mathcal{F}}_{k-1}$. Therefore, $\|\mathrm{X}_k - \mathrm{Y}_k\| = \|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\| + 2\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k$ and

$$\begin{aligned}
&\mathbb{1}_{\Delta}(\mathrm{X}_k, \mathrm{Y}_k) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\Xi_{k+1,n}^{1/2}} \right\} \right] \\
&\leq \mathbb{1}_{\Delta}(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1}) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\| + 2\sigma_k \mathrm{e}_{k-1}(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k}{2\Xi_{k+1,n}^{1/2}} \right\} \right] .
\end{aligned}$$

Since Z_k is independent of $\widetilde{\mathcal{F}}_{k-1}$, $\sigma_k \mathrm{e}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})^{\mathrm{T}} \mathrm{Z}_k$ is a one dimensional zero mean Gaussian random variable with variance σ_k^2. Therefore by Lemma 6.50, we get

$$\begin{aligned}
&\widetilde{\mathbb{E}}_{(x,y)}^{\widetilde{\mathcal{F}}_{k-1}} \left[\mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_k, \mathrm{Y}_k) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{X}_k - \mathrm{Y}_k\|}{2\Xi_{k+1,n}^{1/2}} \right\} \right] \right] \\
&\qquad \leq \mathbb{1}_{\Delta^{\mathrm{c}}}(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1}) \left[1 - 2\boldsymbol{\Phi} \left\{ -\frac{\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\|}{2\left(\sigma_k^2 + \Xi_{k+1,n}\right)^{1/2}} \right\} \right] .
\end{aligned}$$

Using by **AR**3 that $\|\mathrm{E}_k(\mathrm{X}_{k-1}, \mathrm{Y}_{k-1})\| \leq (1 - \varpi_k) \|\mathrm{X}_{k-1} - \mathrm{Y}_{k-1}\|$ concludes the induction. $\square$

Acknowledgements

The work of A.D. and E.M. is supported by the Agence Nationale de la Recherche, under grant ANR-14-CE23-0012 (COSMOS).

Part III

Optimal scaling of Metropolis-Hastings type algorithms

Chapter 7

Optimal scaling of the Random Walk Metropolis algorithm under L^p mean differentiability

Alain Durmus[1], Sylvain Le Corff[2], Eric Moulines[3] and Gareth O. Roberts[4]

Abstract

This paper considers the optimal scaling problem for high-dimensional random walk Metropolis algorithms for densities which are differentiable in L^p mean but which may be irregular at some points (like the Laplace density for example) and / or are supported on an interval. Our main result is the weak convergence of the Markov chain (appropriately rescaled in time and space) to a Langevin diffusion process as the dimension d goes to infinity. Because the log-density might be non-differentiable, the limiting diffusion could be singular. The scaling limit is established under assumptions which are much weaker than the one used in the original derivation of [RGG97]. This result has important practical implications for the use of random walk Metropolis algorithms in Bayesian frameworks based on sparsity inducing priors.

7.1 Introduction

A wealth of contributions have been devoted to the study of the behaviour of high-dimensional Markov chains. One of the most powerful approaches for that purpose is the scaling analysis, introduced by [RGG97]. Assume that the target distribution has a

[1]LTCI, CNRS and Télécom ParisTech.
[2]Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay.
[3]Centre de Mathématiques Appliquées, Ecole Polytechnique.
[4]University of Warwick, Department of Statistics.

density with respect to the d-dimensional Lebesgue measure given by:

$$\pi^d(x^d) = \prod_{i=1}^{d} \pi(x_i^d)\,. \tag{7.1}$$

The Random Walk Metropolis-Hastings (RWM) updating scheme was first applied in [Met+53] and proceeds as follows. Given the current state X_k^d, a new value $Y_{k+1}^d = (Y_{k+1,i}^d)_{i=1}^d$ is obtained by moving independently each coordinate, i.e. $Y_{k+1,i}^d = X_{k,i}^d + \ell d^{-1/2} Z_{k+1}^d$ where $\ell > 0$ is a scaling factor and $(Z_k)_{k\geq 1}$ is a sequence of independent and identically distributed (i.i.d.) Gaussian random variables. Here ℓ governs the overall size of the proposed jump and plays a crucial role in determining the efficiency of the algorithm. The proposal is then accepted or rejected according to the acceptance probability $\alpha(X_k^d, Y_{k+1}^d)$ where $\alpha(x^d, y^d) = 1 \wedge \pi^d(y^d)/\pi^d(x^d)$. If the proposed value is accepted it becomes the next current value, otherwise the current value is left unchanged:

$$X_{k+1}^d = X_k^d + \ell d^{-1/2} Z_{k+1}^d \mathbb{1}_{\mathcal{A}_{k+1}^d}\,, \tag{7.2}$$

$$\mathcal{A}_{k+1}^d = \left\{ U_{k+1} \leq \prod_{i=1}^{d} \pi(X_{k,i}^d + \ell d^{-1/2} Z_{k+1,i}^d)/\pi(X_{k,i}^d) \right\}\,. \tag{7.3}$$

where $(U_k)_{k\geq 1}$ of *i.i.d.* uniform random variables on $[0,1]$ independent of $(Z_k)_{k\geq 1}$.

Under certain regularity assumptions on π, it has been proved in [RGG97] that if the X_0^d is distributed under the stationary distribution π^d, then each component of $(X_k^d)_{k\geq 0}$ appropriately rescaled in time converges weakly to a Langevin diffusion process with invariant distribution π as $d \to +\infty$.

This result allows to compute the asymptotic mean acceptance rate and to derive a practical rule to tune the factor ℓ. It is shown in [RGG97] that the speed of the limiting diffusion has a function of ℓ has a unique maximum. The corresponding mean acceptance rate in stationarity is equal to 0.234.

These results have been derived for target distributions of the form (7.1) where $\pi(x) \propto \exp(-V(x))$ where V is three-times continuously differentiable. Therefore, they do not cover the cases where the target density is continuous but not smooth, for example the Laplace distribution which plays a key role as a sparsity-inducing prior in high-dimensional Bayesian inference.

The aim of this paper is to extend the scaling results for the RWM algorithm introduced in the seminal paper [RGG97, Theorem 7.7] to densities which are absolutely continuous densities differentiable in L^p mean (DLM) for some $p \geq 2$ but can be either non-differentiable at some points or are supported on an interval. As shown in [Le 86, Section 17.3], differentiability of the square root of the density in L^2 norm implies a quadratic approximation property for the log-likelihood known as local asymptotic normality. As shown below, the DLM permits the quadratic expansion of the log-likelihood without paying the twice-differentiability price usually demanded by such a Taylor expansion (such expansion of the log-likelihood plays a key role in [RGG97]).

The paper is organised as follows. In Section 7.2 the target density π is assumed to be positive on $\mathbb{R}$. Theorem 7.4 proves that under the DLM assumption of this paper, the average acceptance rate and the expected square jump distance are the same as in [RGG97]. Theorem 7.7 shows that under the same assumptions the rescaled in time Markov chain produced by the RWM algorithm converges weakly to a Langevin diffusion. We show that these results may be applied to a density of the form $\pi(x) \propto \exp(-\lambda |x| + U(x))$, where $\lambda \geq 0$ and U is a smooth function. In Section 7.3, we focus on the case where π is supported only on an open interval of $\mathbb{R}$. Under appropriate assumptions, Theorem 7.9 and Theorem 7.12 show that the same asymptotic results (limiting average acceptance rate and limiting Langevin diffusion associated with π) hold. We apply our results to Gamma and Beta distributions. The proofs are postponed to Section 7.4 and Section 7.5.

7.2 Positive Target density on the real line

The key of the proof of our main result will be to show that the acceptance ratio and the expected square jump distance converge to a finite and non trivial limit. In the original proof of [RGG97], the density of the product form (7.1) with

$$\pi(x) \propto \exp(-V(x)) \tag{7.4}$$

is three-times continuously differentiable and the acceptance ratio is expanded using the usual pointwise Taylor formula. More precisely, the log-ratio of the density evaluated at the proposed value and at the current state is given by $\sum_{i=1}^d \Delta V_i^d$ where

$$\Delta V_i^d = V\left(X_i^d\right) - V\left(X_i^d + \ell d^{-1/2} Z_i^d\right) , \tag{7.5}$$

and where X^d is distributed according to π^d and Z^d is a d-dimensional standard Gaussian random variable independent of X. Heuristically, the two leading terms are $\ell d^{-1/2} \sum_{i=1}^d \dot{V}(X_i^d) Z_i^d$ and $\ell^2 d^{-1} \sum_{i=1}^d \ddot{V}(X_i^d)(Z_i^d)^2/2$, where $\dot{V}$ and $\ddot{V}$ are the first and second derivatives of V, respectively. By the central limit theorem, this expression converges in distribution to a zero-mean Gaussian random variable with variance $\ell^2 I$ where

$$I = \int_{\mathbb{R}} \dot{V}^2(x)\pi(x)\mathrm{d}x\,. \tag{7.6}$$

Note that I is the Fisher information associated with the translation model $\theta \mapsto \pi(x+\theta)$ evaluated at $\theta = 0$. Under appropriate technical conditions, the second term converges almost surely to $-\ell^2 I/2$. Assuming that these limits exist, the acceptance ratio in the RWM algorithm converges to $\mathbb{E}[1 \wedge \exp(\mathsf{Z})]$ where Z is a Gaussian random variable with mean $-\ell^2 I/2$ and variance $\ell^2 I$; elementary computations show that $\mathbb{E}[1 \wedge \exp(\mathsf{Z})] = 2\Phi(-\ell/2\sqrt{I})$, where Φ stands for the cumulative distribution function of a standard normal distribution.

For $t \geq 0$, denote by Y_t^d the linear interpolation of the Markov chain $(X_k^d)_{k\geq 0}$ after time rescaling:

$$\begin{aligned} Y_t^d &= (\lceil d\,t\rceil - d\,t)\, X_{\lfloor d\,t\rfloor}^d + (d\,t - \lfloor d\,t\rfloor)\, X_{\lceil d\,t\rceil}^d && (7.7)\\ &= X_{\lfloor d\,t\rfloor}^d + (d\,t - \lfloor d\,t\rfloor)\, \ell d^{-1/2} Z_{\lceil d\,t\rceil}^d \mathbb{1}_{\mathcal{A}_{\lceil d\,t\rceil}^d}\,, && (7.8)\end{aligned}$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the lower and the upper integer part functions. Note that for all $k \geq 0$, $Y^d_{k/d} = X^d_k$. Denote by $(B_t, t \geq 0)$ the standard Brownian motion.

Theorem 7.1 ([RGG97]). *Suppose that the target π^d and the proposal distribution are given by (7.1)-(7.4) and (7.2) respectively. Assume that*

(i) V is twice continuously differentiable and $\dot{V}$ is Lipshitz continuous ;

(ii) $\mathbb{E}[(\dot{V}(X))^8] < \infty$ and $\mathbb{E}[(\ddot{V}(X))^4] < \infty$ where X is distributed according to π.

Then $(Y^d_{t,1}, t \geq 0)$, where $Y^d_{t,1}$ is the first component of the vector Y^d_t defined in (7.7), converges weakly in the Wiener space (equipped with the uniform topology) to the Langevin diffusion

$$\mathrm{d}Y_t = \sqrt{h(\ell)}\mathrm{d}B_t - \frac{1}{2}h(\ell)\dot{V}(Y_t)\mathrm{d}t\,, \tag{7.9}$$

where Y_0 is distributed according to π, $h(\ell)$ is given by

$$h(\ell) = 2\ell^2\Phi\left(-\frac{\ell}{2}\sqrt{I}\right)\,, \tag{7.10}$$

and I is defined in (7.6).

Whereas V is assumed to be twice continuously differentiable, the dual representation of the Fisher information $-\mathbb{E}[\ddot{V}(X)] = \mathbb{E}[(\dot{V}(X))^2] = I$ allows us to remove in the statement of the theorem all mention to the second derivative of V, which hints that two derivatives might not really be required. For all $\theta, x \in \mathbb{R}$, define

$$\xi_\theta(x) = \sqrt{\pi(x+\theta)}\,, \tag{7.11}$$

For $p \geq 1$, denote $\|f\|^p_{\pi,p} = \int |f(x)|^p\pi(x)\mathrm{d}x$. Consider the following assumptions:

H1. *There exists a measurable function $\dot{V} : \mathbb{R} \to \mathbb{R}$ such that:*

(i) There exist $p > 4$, $C > 0$ and $\beta > 1$ such that for all $\theta \in \mathbb{R}$,

$$\left\|V(\cdot+\theta) - V(\cdot) - \theta\dot{V}(\cdot)\right\|_{\pi,p} \leq C|\theta|^\beta\,.$$

(ii) The function $\dot{V}$ satisfies $\left\|\dot{V}\right\|_{\pi,6} < +\infty$.

Lemma 7.2. *Assume **H**1. Then, the family of densities $\theta \to \pi(\cdot+\theta)$ is Differentiable in Quadratic Mean (DQM) at $\theta = 0$ with derivative $\dot{V}$, i.e. there exists $C > 0$ such that for all $\theta \in \mathbb{R}$,*

$$\left(\int_{\mathbb{R}}\left(\xi_\theta(x) - \xi_0(x) + \theta\dot{V}(x)\xi_0(x)/2\right)^2\mathrm{d}x\right)^{1/2} \leq C|\theta|^\beta\,,$$

where ξ_θ is given by (7.11).

Proof. The proof is postponed to Section 7.4.1. □

The first step in the proof is to show that the acceptance ratio $\mathbb{P}\left(\mathcal{A}_1^d\right) = \mathbb{E}(1 \wedge \exp\{\sum_{i=1}^d \Delta V_i^d\})$, and the expected square jump distance $\mathbb{E}[(Z_1^d)^2\{1 \wedge \exp(\sum_{i=1}^d \Delta V_i^d)\}]$ both converge to a finite value. To that purpose, we consider

$$\mathrm{E}^d(q) = \mathbb{E}\left[\left(Z_1^d\right)^q \left|1 \wedge \exp\left(\sum_{i=1}^d \Delta V_i^d\right) - 1 \wedge \exp(\upsilon^d)\right|\right],$$

where ΔV_i^d is given by (7.5),

$$\upsilon^d = -\ell d^{-1/2} Z_1^d \dot{V}(X_1^d) + \sum_{i=2}^d b^d(X_i^d, Z_i^d) \tag{7.12}$$

$$b^d(x,z) = -\frac{\ell z}{\sqrt{d}}\dot{V}(x) + \mathbb{E}\left[2\zeta^d(X_1^d, Z_1^d)\right] - \frac{\ell^2}{4d}\dot{V}^2(x)\,, \tag{7.13}$$

$$\zeta^d(x,z) = \exp\left\{\left(V(x) - V\left(x + \ell d^{-1/2} z\right)\right)/2\right\} - 1\,. \tag{7.14}$$

Proposition 7.3. *Assume **H**1 holds. Let X^d be a random variable distributed according to π^d and Z^d be a zero-mean standard Gaussian random variable, independent of X^d. Then, for any $q \geq 0$, $\lim_{d\to+\infty} \mathrm{E}^d(q) = 0$.*

Proof. The proof is postponed to Section 7.4.2. □

Proposition 7.3 shows that it is enough to consider υ^d to analyse the asymptotic behaviour of the acceptance ratio and the expected square jump distance as $d \to +\infty$. By the central limit theorem, the term $-\ell \sum_{i=2}^d (Z_i^d/\sqrt{d})\dot{V}(X_i^d)$ in (7.12) converges in distribution to a zero-mean Gaussian random variable with variance $\ell^2 I$, where I is defined in (7.6). By Lemma 7.15 (Section 7.4.2), the second term, which is $d\,\mathbb{E}\left[2\zeta^d(X_1^d, Z_1^d)\right] = -d\,\mathbb{E}\left[(\zeta^d(X_1^d, Z_1^d))^2\right]$ converges to $-\ell^2 I/4$. The last term converges in probability to $-\ell^2 I/4$. Therefore, the two last terms plays a similar role in the expansion of the acceptance ratio as the second derivative of V in the regular case.

Theorem 7.4. *Assume **H**1 holds. Then, $\lim_{d\to+\infty} \mathbb{P}\left[\mathcal{A}_1^d\right] = a(\ell)$, where $a(\ell) = 2\Phi(-\sqrt{I}\ell/2)$.*

Proof. The proof is postponed to Section 7.4.2. □

The second result of this paper is that the sequence $\{(Y_{t,1}^d)_{t\geq 0}, d \in \mathbb{N}^\star\}$ defined by (7.7) converges weakly to a Langevin diffusion. Let $(\mu_d)_{d\geq 1}$ be the sequence of distributions of $\{(Y_{t,1}^d)_{t\geq 0}, d \in \mathbb{N}^\star\}$.

Proposition 7.5. *Assume **H**1 holds. Then, the sequence $(\mu_d)_{d\geq 1}$ is tight in* **W**.

Proof. The proof is adapted from [JLM15]; it is postponed to Section 7.4.4. □

By the Prohorov theorem, the tightness of $(\mu_d)_{d\geq 1}$ implies that this sequence has a weak limit point. We now prove that any limit point is the law of a solution to (7.9). For that purpose, we use the equivalence between the weak formulation of stochastic differential equations and martingale problems. The generator L of the Langevin diffusion (7.9) is given, for all $\phi \in C_c^2(\mathbb{R},\mathbb{R})$, by

$$\mathrm{L}\phi(x) = \frac{h(\ell)}{2}\left(-\dot{V}(x)\dot{\phi}(x) + \ddot{\phi}(x)\right) , \tag{7.15}$$

where for $k \in \mathbb{N}$ and I an open subset of $\mathbb{R}$, $C_c^k(I,\mathbb{R})$ is the space of k-times differentiable functions with compact support, endowed with the topology of uniform convergence of all derivatives up to order k. We set $C_c^\infty(I,\mathbb{R}) = \bigcap_{k=0}^\infty C_c^k(I,\mathbb{R})$ and $\mathbf{W} = C(\mathbb{R}_+,\mathbb{R})$. The canonical process is denoted by $(W_t)_{t\geq 0}$ and $(\mathscr{B}_t)_{t\geq 0}$ is the associated filtration. For any probability measure μ on $\mathbf{W}$, the expectation with respect to μ is denoted by $\mathbb{E}^\mu$. A probability measure μ on $\mathbf{W}$ is said to solve the martingale problem associated with (7.9) if the pushforward of μ by W_0 is π and if for all $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$, the process

$$\left(\phi(W_t) - \phi(W_0) - \int_0^t \mathrm{L}\phi(W_u)\mathrm{d}u\right)_{t\geq 0}$$

is a martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$, i.e. if for all $s,t \in \mathbb{R}_+$, $s \leq t$, $\mu -$ a.s.

$$\mathbb{E}^\mu\left[\phi(W_t) - \phi(W_0) - \int_0^t \mathrm{L}\phi(W_u)\mathrm{d}u \middle| \mathscr{B}_s\right] = \phi(W_s) - \phi(W_0) - \int_0^s \mathrm{L}\phi(W_u)\mathrm{d}s\,.$$

H 2. *The function $\dot{V}$ is continuous on $\mathbb{R}$ except on a Lebesgue-negligible set $\mathcal{D}_{\dot{V}}$ and is bounded on all compact sets of $\mathbb{R}$.*

If $\dot{V}$ satisfies **H**2, [RW00, Lemma 1.9, Theorem 20.1 Chapter 5] show that any solution to the martingale problem associated with (7.9) coincides with the law of a solution to the SDE (7.9), and conversely. Therefore, uniqueness in law of weak solutions to (7.9) implies uniqueness of the solution of the martingale problem.

Proposition 7.6. *Assume **H**2 holds. Assume also that for all $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$, $m \in \mathbb{N}^*$, $g : \mathbb{R}^m \to \mathbb{R}$ bounded and continuous, and $0 \leq t_1 \leq \cdots \leq t_m \leq s \leq t$:*

$$\lim_{d\to+\infty} \mathbb{E}^{\mu_d}\left[\left(\phi(W_t) - \phi(W_s) - \int_s^t \mathrm{L}\phi(W_u)\,\mathrm{d}u\right) g(W_{t_1},\ldots,W_{t_m})\right] = 0\,. \tag{7.16}$$

Then, every limit point of the sequence of probability measures $(\mu_d)_{d\geq 1}$ on $\mathbf{W}$ is a solution to the martingale problem associated with (7.9).

Proof. The proof is postponed to Section 7.4.5. □

Theorem 7.7. *Assume **H**1 and **H**2 hold. Assume also that (7.9) has a unique weak solution. Then, $\left\{(Y_{t,1}^d)_{t\geq 0}, d \in \mathbb{N}^*\right\}$ converges weakly to the solution $(Y_t)_{t\geq 0}$ of the Langevin equation defined by (7.9). Furthermore, $h(\ell)$ is maximized at the unique value of ℓ for which $a(\ell) = 0.234$, where a is defined in Theorem 7.4.*

Proof. The proof is postponed to Section 7.4.6. $\square$

Example 7.8 (Bayesian Lasso). *We apply the results obtained above to a target density π on $\mathbb{R}$ given by $x \mapsto \mathrm{e}^{-V(x)} / \int_{\mathbb{R}} \mathrm{e}^{-V(y)} \mathrm{d}y$ where V is given by*

$$V : x \mapsto U(x) + \lambda |x| \ ,$$

*where $\lambda \geq 0$ and U is twice continuously differentiable with bounded second derivative. Furthermore, $\int_{\mathbb{R}} |x|^6 \mathrm{e}^{-V(x)} \mathrm{d}x < +\infty$. Define $\dot{V} : x \mapsto U'(x) + \lambda \operatorname{sign}(x)$, with $\operatorname{sign}(x) = -1$ if $x \leq 0$ and $\operatorname{sign}(x) = 1$ otherwise. We first check that **H**1(i) holds. Note that for all $x, y \in \mathbb{R}$,*

$$||x+y| - |x| - \operatorname{sign}(x) y| \leq 2|y| \mathbb{1}_{\mathbb{R}_+}(|y| - |x|) \,, \tag{7.17}$$

which implies that, for any $p \geq 1$, there exists C_p such that

$$\begin{aligned}
&\left\| V(\cdot + \theta) - V(\cdot) - \theta \dot{V}(\cdot) \right\|_{\pi,p} \\
&\leq \left\| U(\cdot + \theta) - U(\cdot) - \theta U'(\cdot) \right\|_{\pi,p} + \lambda \left\| |\cdot + \theta| - |\cdot| - \theta \operatorname{sign}(\cdot) \right\|_{\pi,p} \\
&\leq \left\| U'' \right\|_{\infty} \theta^2 + 2 \, |\theta| \, \lambda \{\pi([-\theta, \theta])\}^{1/p} \leq C \, |\theta|^{p+1/p} \vee |\theta|^2 \ .
\end{aligned}$$

*Assumptions **H** 1(ii) and **H** 2 are easy to check. The uniqueness in law of (7.9) is established in [CE05, Theorem 4.5 (i)]. Therefore, Theorem 7.7 can be applied.*

7.3 Target density supported on an interval

We now prove that our results can be applied to densities of the form:

$$\pi(x) \propto \exp(-V(x)) \mathbb{1}_{\mathcal{I}}(x) \,,$$

where $\mathcal{I}$ is an open interval of $\mathbb{R}$ and $V : \mathcal{I} \to \mathbb{R}$ is a measurable function. Note that by convention $V(x) = -\infty$ for all $x \notin \mathcal{I}$. Denote by $\overline{\mathcal{I}}$ the closure of $\mathcal{I}$ in $\mathbb{R}$. The results of Section 7.2 may be extended to this setting but, as π is not positive on $\mathbb{R}$, this requires the following new assumptions.

G 4. *There exists a measurable function $\dot{V} : \mathbb{R} \to \mathbb{R}$ and $\mathrm{r} > 1$ such that:*

(i) There exist $p > 4$, $C > 0$ and $\beta > 1$ such that for all $\theta \in \mathbb{R}$,

$$\left\| \{V(\cdot + \theta) - V(\cdot)\} \mathbb{1}_{\mathcal{I}}(\cdot + \mathrm{r}\theta) \mathbb{1}_{\mathcal{I}}(\cdot + (1-\mathrm{r})\theta) - \theta \dot{V}(\cdot) \right\|_{\pi,p} \leq C |\theta|^{\beta} \,,$$

with the convention $0 \times \infty = 0$.

(ii) The function $\dot{V}$ satisfies $\int_{\mathcal{I}} \left| \dot{V}(x) \right|^6 \pi(x) \mathrm{d}x < +\infty$.

(iii) There exist $\gamma \geq 6$ and $C > 0$ such that, for all $\theta \in \mathbb{R}$,

$$\int_{\mathbb{R}} \mathbb{1}_{\mathcal{I}^c}(x + \theta) \pi(x) \mathrm{d}x \leq C |\theta|^{\gamma} \,.$$

In the examples that we consider $x \mapsto V(x)$ is not integrable on $\mathcal{I}$, and introducing $\mathrm{r} > 1$ in the assumption allows to circumvent this issue. Besides, note that since $\mathcal{I}$ is an interval and $\mathrm{r} > 1$, for all $\theta \in \mathbb{R}$, if $x \in \mathcal{I}$ and $x + \mathrm{r}\theta \in \mathcal{I}$, $x + \theta \in \mathcal{I}$ and $V(x+\theta)$ is finite for such an x. As an important consequence of **G**4(iii), if X is distributed according to π and is independent of the standard random variable Z, there exists a constant C such that

$$\mathbb{P}\left(X + \ell d^{-1/2} Z \in \mathcal{I}^c\right) \leq C d^{-\gamma/2}\,. \tag{7.18}$$

Theorem 7.9. *Assume **G**4 holds. Then,* $\lim_{d\to+\infty} \mathbb{P}\left[\mathcal{A}_1^d\right] = a(\ell)$*, where a is defined in Theorem 7.4.*

Proof. The proof is postponed to Section 7.5.1. $\square$

We now established the weak convergence of the sequence $\{(Y_{t,1}^d)_{t\geq 0}, d \in \mathbb{N}^\star\}$, following the same steps as for the proof of Theorem 7.7. Denote for all $d \geq 1$, μ_d the law of the process $(Y_{t,1}^d)_{t\geq 0}$.

Proposition 7.10. *Assume **G**4 holds. Then, the sequence $(\mu_d)_{d\geq 1}$ is tight in $\mathbf{W}$.*

Proof. The proof is postponed to Section 7.5.2. $\square$

Contrary to the case where π is positive on $\mathbb{R}$, we do not assume that $\dot{V}$ is bounded on all compact sets of $\mathbb{R}$. Therefore, the martingale problem associated with (7.9) does not characterize the law of a solution to (7.9). However, we can still consider the local martingale problem associated with (7.9). Using the same notations as in Section 7.2, a probability measure μ on $\mathbf{W}$ is said to solve the local martingale problem associated with (7.9) if the pushforward of μ by W_0 is π and if for all $\psi \in C^\infty(\mathbb{R}, \mathbb{R})$, the process

$$\left(\psi(W_t) - \psi(W_0) - \int_0^t \mathrm{L}\psi(W_u)\mathrm{d}u\right)_{t\geq 0}$$

is a local martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$. By [CE05, Theorem 1.27], any solution to the local martingale problem associated with (7.9) coincides with the law of a solution to the SDE (7.9) and conversely. If (7.9) admits a unique solution in law, this law is the unique solution to the local martingale problem associated with (7.9). In the following, we first prove that any limit point μ of $(\mu_d)_{d\geq 1}$ is a solution to the local martingale problem associated with (7.9).

G 5. *The function $\dot{V}$ is continuous on $\mathcal{I}$ except on a null-set $\mathcal{D}_{\dot{V}}$, with respect to the Lebesgue measure, and is bounded on all compact sets of $\mathcal{I}$.*

This condition does not preclude that $\dot{V}$ remains bounded at the boundary of $\mathcal{I}$.

Proposition 7.11. *Assume **G**4 and **G**5 hold. Assume also that for all $\phi \in C_c^\infty(\mathcal{I}, \mathbb{R})$, $m \in \mathbb{N}^*$, $g : \mathbb{R}^m \to \mathbb{R}$ bounded and continuous, and $0 \leq t_1 \leq \cdots \leq t_m \leq s \leq t$:*

$$\lim_{d\to+\infty} \mathbb{E}^{\mu_d}\left[\left(\phi(W_t) - \phi(W_s) - \int_s^t \mathrm{L}\phi(W_u)\,\mathrm{d}u\right) g(W_{t_1}, \ldots, W_{t_m})\right] = 0\,. \tag{7.19}$$

Then, every limit point of the sequence of probability measures $(\mu_d)_{d\geq 1}$ on $\mathbf{W}$ is a solution to the local martingale problem associated with (7.9).

Proof. The proof is postponed to Section 7.5.3. □

Theorem 7.12. *Assume **G**4 and **G**5 hold. Assume also that (7.9) has a unique weak solution. Then, $\left\{(Y^d_{t,1})_{t\geq 0}, d \in \mathbb{N}^*\right\}$ converges weakly to the solution $(Y_t)_{t\geq 0}$ of the Langevin equation defined by (7.9). Furthermore, $h(\ell)$ is maximized at the unique value of ℓ for which $a(\ell) = 0.234$, where a is defined in Theorem 7.4.*

Proof. The proof is postponed to Section 7.5.4. □

The conditions for uniqueness in law of singular one-dimensional stochastic differential equations are given in [CE05]. These conditions are rather involved and difficult to summarize in full generality. We rather illustrate Theorem 7.12 by two examples.

7.3.1 Application to the Gamma and the beta distributions

Define the class of the generalized Gamma distributions as the family of densities on $\mathbb{R}$ given by

$$\pi_\gamma : x \mapsto x^{\mathrm{a}_1 - 1} \exp(-x^{\mathrm{a}_2}) \mathbb{1}_{\mathbb{R}_+^\star}(x) / \int_{\mathbb{R}_+^\star} y^{\mathrm{a}_1 - 1} \exp(-y^{\mathrm{a}_2}) \mathrm{d}y \,,$$

with two parameters $\mathrm{a}_1 > 6$ and $\mathrm{a}_2 > 0$. Note that in this case $\mathcal{I} = \mathbb{R}_+^\star$, for all $x \in \mathcal{I}$,

$$V_\gamma : x \mapsto x^{\mathrm{a}_2} - (\mathrm{a}_1 - 1)\log x \quad \text{and} \quad \dot{V}_\gamma : x \mapsto \mathrm{a}_2 x^{\mathrm{a}_2 - 1} - (\mathrm{a}_1 - 1)/x \,,$$

We check that our results may be applied to this class of distributions for $\mathrm{r} = 3/2$. First, we show that **G**4(i) for $p = 5$. Write for all $\theta \in \mathbb{R}$ and $x \in \mathcal{I}$,

$$\{V_\gamma(x+\theta) - V_\gamma(x)\} \mathbb{1}_{\mathcal{I}}(x + (1-r)\theta) \mathbb{1}_{\mathcal{I}}(x + r\theta) - \theta \dot{V}_\gamma(x) = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 \,,$$

where

$$\begin{aligned}
\mathcal{E}_1 &\stackrel{\text{def}}{=} \theta \left\{(\mathrm{a}_1 - 1)/x - \mathrm{a}_2 x^{\mathrm{a}_2 - 1}\right\} \{1 - \mathbb{1}_{\mathcal{I}}(x - \theta/2)\mathbb{1}_{\mathcal{I}}(x + 3\theta/2)\} \,, \\
\mathcal{E}_2 &\stackrel{\text{def}}{=} (1 - \mathrm{a}_1)\{\log(1 + \theta/x) - \theta/x\} \mathbb{1}_{\mathcal{I}}(x - \theta/2)\mathbb{1}_{\mathcal{I}}(x + 3\theta/2) \,, \\
\mathcal{E}_3 &\stackrel{\text{def}}{=} ((x+\theta)^{\mathrm{a}_2} - x^{\mathrm{a}_2} - \mathrm{a}_2 \theta x^{\mathrm{a}_2 - 1}) \mathbb{1}_{\mathcal{I}}(x - \theta/2)\mathbb{1}_{\mathcal{I}}(x + 3\theta/2) \,.
\end{aligned}$$

It is enough to prove that there exists $q > p$ such that for all $i \in \{1,2,3\}$, $\int_{\mathcal{I}} |\mathcal{E}_i|^p \, \pi_\gamma(x)\mathrm{d}x \leq C|\theta|^q$. **G**4(i) is proved for $\theta < 0$ (the proof for $\theta > 0$ follows the same lines). For all

$\theta \in \mathbb{R}$ using $\mathrm{a}_1 > 6$,

$$\begin{aligned}
&\int_{\mathbb{R}_+^*} |\mathcal{E}_1|^5 \, \pi_\gamma(x)\mathrm{d}x \\
&\leq C|\theta|^5 \int_{\mathbb{R}_+^*} \left\{1/x^5 + x^{5(\mathrm{a}_2-1)}\right\} x^{\mathrm{a}_1-1}\mathrm{e}^{-x^{\mathrm{a}_2}} \{\mathbb{1}_{\mathbb{R}_-}(x+3\theta/2) + \mathbb{1}_{\mathbb{R}_-}(x-\theta/2)\}\mathrm{d}x\,, \\
&\leq C|\theta|^5 \int_0^{3|\theta|/2} \left\{1/x^5 + x^{5(\mathrm{a}_2-1)}\right\} x^{\mathrm{a}_1-1}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x\,, \\
&\leq C\left(|\theta|^{\mathrm{a}_1}\int_0^{3/2} x^{\mathrm{a}_1-6}\mathrm{e}^{-(|\theta|x)^{\mathrm{a}_2}}\mathrm{d}x + |\theta|^{5\mathrm{a}_2+\mathrm{a}_1}\int_0^{3/2} x^{5(\mathrm{a}_2-1)+\mathrm{a}_1-1}\mathrm{e}^{-(|\theta|x)^{\mathrm{a}_2}}\mathrm{d}x\right)\,, \\
&\leq C(|\theta|^{\mathrm{a}_1} + |\theta|^{5\mathrm{a}_2+\mathrm{a}_1})\,.
\end{aligned} \tag{7.20}$$

On the other hand, as for all $x > -1$, $x/(x+1) \leq \log(1+x) \leq x$, for all $\theta < 0$, and $x \geq 3|\theta|/2$,

$$|\log(1+\theta/x) - \theta/x| \leq \frac{|\theta|^2}{x^2(1+\theta/x)} \leq 3|\theta|^2/x^2\,,$$

where the last inequality come from $|\theta|/x \leq 2/3$. Then, it yields

$$\begin{aligned}
\int_{\mathbb{R}_+^*} |\mathcal{E}_2(x)|^5 \, \pi_\gamma(x)\mathrm{d}x &\leq C|\theta|^{10}\int_{3|\theta|/2}^{+\infty} x^{\mathrm{a}_1-11}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x\,, \\
&\leq C\left(|\theta|^{10}\int_{3|\theta|/2}^{1} x^{\mathrm{a}_1-11}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x + |\theta|^{10}\int_1^{+\infty} x^{\mathrm{a}_1-11}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x\right) \leq C(|\theta|^{\mathrm{a}_1} + |\theta|^{10})\,.
\end{aligned} \tag{7.21}$$

For the last term, for all $\theta < 0$ and all $x \geq 3|\theta|/2$, using a Taylor expansion of $x \mapsto x^{\mathrm{a}_2}$, there exists $\zeta \in [x+\theta, x]$ such that

$$\left|(x+\theta)^{\mathrm{a}_2} - x^{\mathrm{a}_2} - \mathrm{a}_2\theta x^{\mathrm{a}_2-1}\right| \leq C|\theta|^2|\zeta|^{\mathrm{a}_2-2} \leq C|\theta|^2\,|x|^{\mathrm{a}_2-2}\,.$$

Then,

$$\int_{\mathbb{R}_+^*} |\mathcal{E}_3(x)|^5 \, \pi_\gamma(x)\mathrm{d}x \leq C|\theta|^{10}\int_{3|\theta|/2}^{+\infty} x^{5(\mathrm{a}_2-2)+\mathrm{a}_1-1}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x \leq C(|\theta|^{5\mathrm{a}_2+\mathrm{a}_1}+|\theta|^{10})\,. \tag{7.22}$$

Combining (7.20), (7.21),(7.22) and using that $\mathrm{a}_1 > 6$ concludes the proof of **G**4(i) for $p = 5$. Consider now **G**4(ii). For all $\theta < 0$ (the case $\theta > 0$ is dealt along the same lines)

$$\begin{aligned}
\int_{\mathbb{R}_+^*} |\dot{V}_\gamma(x)|^6\pi_\gamma(x)\mathrm{d}x &\leq C\int_{\mathbb{R}_+^*} \left|\mathrm{a}_2x^{\mathrm{a}_2-1} + (\mathrm{a}_1-1)/x\right|^6 x^{\mathrm{a}_1-1}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x\,, \\
&\leq C\left(\int_{\mathbb{R}_+^*} x^{\mathrm{a}_1-1+6(\mathrm{a}_2-1)}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x + \int_{\mathbb{R}_+^*} x^{\mathrm{a}_1-7}\mathrm{e}^{-x^{\mathrm{a}_2}}\mathrm{d}x\right)\,,
\end{aligned}$$

where the right hand side is finite for $\mathrm{a}_1 > 6$. Distinguishing the cases $\theta < 0$ and $\theta \geq 0$, and using a change of variable, we have

$$\int_{\mathbb{R}} \{\mathbb{1}_{\mathcal{I}^c}(x + 3\theta/2) + \mathbb{1}_{\mathcal{I}^c}(x - \theta/2)\}\pi_\gamma(x)\mathrm{d}x \\ \leq C|\theta|^{\mathrm{a}_1} \int_{\mathbb{R}_+^*} (\mathbb{1}_{(0,1/2)} + \mathbb{1}_{(0,3/2)})x^{\mathrm{a}_1 - 1}\mathrm{d}x \leq C|\theta|^{\mathrm{a}_1}$$

and **G**4(iii) follows with $\gamma = \mathrm{a}_1$. Therefore, Theorem 7.9 can be applied. Now consider the Langevin equation associated with π_γ given by

$$\mathrm{d}Y_t = -\dot{V}_\gamma(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t\,, \tag{7.23}$$

with initial distribution π_γ. This stochastic differential equation has 0 as singular point, which has right type 3 according to the terminology of [CE05]. On the other hand ∞ has type A and the existence and uniqueness in law for (7.23) follows from [CE05, Theorem 4.6 (viii)]. Since **G**5 is straightforward, Theorem 7.12 can be applied.

Consider now the case of the beta distributions π_β with density $x \mapsto x^{a_1-1}(1-x)^{a_2-1}\mathbb{1}_{(0,1)}(x)$ with $a_1, a_2 > 6$. Here $\mathcal{I} = (0,1)$ and the log-density V_β and its derivative on $\mathcal{I}$ are defined by

$$V_\beta : -(a_1 - 1)\log x - (a_2 - 1)\log(1-x) \quad \text{and} \quad \dot{V}_\beta : x \mapsto -(a_1-1)/x - (a_2-1)/(1-x)\,.$$

Using the same calculations as for the gamma distributions, it follows that π_β satisfies **G**4 and Theorem 7.9 can be applied. Note also that **G**5 is straightforward. It remains to establish the uniqueness in law for the Langevin equation associated with π_β defined by

$$\mathrm{d}Y_t = -\dot{V}_\beta(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t\,, \tag{7.24}$$

with initial distribution π_β. But using again the terminology of [CE05], 0 has right type 3 and 1 has left type 3. Therefore by [CE05, Theorem 2.16 (i), (ii)], (7.24) has a global unique weak solution.

The remainder of this section presents a toy simulation study to assess our results for the beta distribution with parameter $a_1 = 10$ and $a_2 = 10$. Define the expected square distance by

$$\mathrm{ESJD}^d(\ell) \overset{\text{def}}{=} \mathbb{E}\left[\left\|X_1^d - X_0^d\right\|^2\right]\,,$$

where X_0^d has distribution π_β^d and X_1^d is the first iterate of the Markov chain defined by the Random Walk Metropolis algorithm given in (7.2). By Theorem 7.9 and Theorem 7.12, we have $\lim_{d\to+\infty} \mathrm{ESJD}^d(\ell) = h(\ell) = \ell^2 a(\ell)$. Figure 7.1 displays an empirical estimation for the ESJD^d for dimensions $d = 10, 50, 1000$ as a function of the empirical mean acceptance rate. We can observe that as expected, the ESJD^d converges to some limit function as d goes infinity, and this function has a maximum for a mean acceptance probability around 0.23.

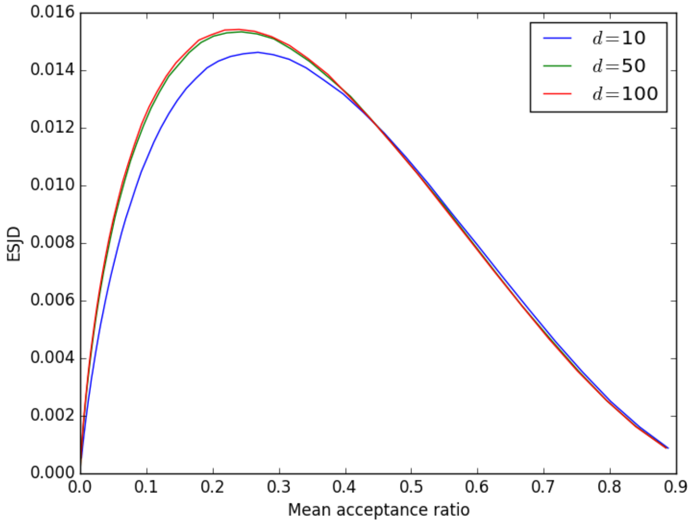


Figure 7.1: Expected square jumped distance for the beta distribution with parameters $a_1 = 10$ and $a_2 = 10$ as a function of the mean acceptance rate for $d = 10, 50, 100$.

7.4 Proofs of Section 7.2

For any real random variable Y and any $p \geq 1$, let $\|Y\|_p \stackrel{\text{def}}{=} \mathbb{E}[|Y|^p]^{1/p}$.

7.4.1 Proof of Lemma 7.2

Let $\Delta_\theta V(x) = V(x) - V(x+\theta)$. By definition of ξ_θ and π,

$$\left(\xi_\theta(x) - \xi_0(x) + \theta \dot{V}(x)\xi_0(x)/2\right)^2 \leq 2\left\{A_\theta(x) + B_\theta(x)\right\}\pi(x)\,,$$

where

$$A_\theta(x) = \left(\exp(\Delta_\theta V(x)/2) - 1 - \Delta_\theta V(x)/2\right)^2\,,$$
$$B_\theta(x) = \left(\Delta_\theta V(x) + \theta \dot{V}(x)\right)^2/4.$$

By **H** 1(i), $\|B_\theta\|_{\pi,p} \leq C|\theta|^\beta$. For A_θ, note that for all $x \in \mathbb{R}$, $(\exp(x) - 1 - x)^2 \leq 2x^4(\exp(2x)+1)$. Then,

$$\begin{aligned}\int_{\mathbb{R}} A_\theta(x)\pi(x)\mathrm{d}x &\leq C\int_{\mathbb{R}} \Delta_\theta V(x)^4\left(1 + \mathrm{e}^{\Delta_\theta V(x)}\right)\pi(x)\mathrm{d}x\\ &\leq C\int_{\mathbb{R}}\left(\Delta_\theta V(x)^4 + \Delta_{-\theta}V(x)^4\right)\pi(x)\mathrm{d}x\,.\end{aligned}$$

The proof is completed writing (the same inequality holds for $\Delta_{-\theta}V$):

$$\int_{\mathbb{R}} \Delta_\theta V(x)^4 \pi(x)\mathrm{d}x \leq C\left[\int_{\mathbb{R}}\left(\Delta_\theta V(x) - \theta\dot{V}(x)\right)^4\pi(x)\mathrm{d}x + \theta^4\int_{\mathbb{R}}\dot{V}^4(x)\pi(x)\mathrm{d}x\right]$$

and using **H**1(i)-(ii).

7.4.2 Proof of Proposition 7.3

Define

$$R(x) = \int_0^x \frac{(x-u)^2}{(1+u)^3} \mathrm{d}u \,. \tag{7.25}$$

R is the remainder term of the Taylor expansion of $x \mapsto \log(1+x)$:

$$\log(1+x) = x - x^2/2 + R(x) \,. \tag{7.26}$$

We preface the proof by the following Lemma.

Lemma 7.13. *Assume **H**1 holds. Then, if X is a random variable distributed according to π and Z is a standard Gaussian random variable independent of X,*

(i) $\lim_{d \to +\infty} d \, \left\| \zeta^d(X,Z) + \ell Z \dot{V}(X)/(2\sqrt{d}) \right\|_2^2 = 0.$

(ii) $\lim_{d \to +\infty} \sqrt{d} \left\| V(X) - V(X + \ell Z/\sqrt{d}) + \ell Z \dot{V}(X)/\sqrt{d} \right\|_p = 0.$

(iii) $\lim_{d \to \infty} d \left\| R\left(\zeta^d(X,Z)\right) \right\|_1 = 0,$

where ζ^d is given by (7.14).

Proof. Using the definitions (7.11) and (7.14) of ζ^d and ξ_θ ,

$$\zeta^d(x,z) = \xi_{\ell z d^{-1/2}}(x)/\xi_0(x) - 1 \,. \tag{7.27}$$

(i) The proof follows from Lemma 7.2 using that $\beta > 1$:

$$\left\| \zeta^d(X,Z) + \ell Z \dot{V}(X)/(2\sqrt{d}) \right\|_2^2 \leq C \ell^{2\beta} d^{-\beta} \mathbb{E}\left[|Z|^{2\beta} \right] \,.$$

(ii) Using **H**1(i), we get that

$$\left\| V(X) - V(X + \ell Z/\sqrt{d}) + \ell Z \dot{V}(X)/\sqrt{d} \right\|_p^p \leq C \ell^{\beta p} d^{-\beta p/2} \mathbb{E}\left[|Z|^{\beta p} \right]$$

and the proof follows since $\beta > 1$.

(iii) Note that for all $x > 0$, $u \in [0,x]$, $|(x-u)(1+u)^{-1}| \leq |x|$, and the same inequality holds for $x \in (-1,0]$ and $u \in [x,0]$. Then by (7.25) and (7.26), for all $x > -1$, $|R(x)| \leq x^2 \left|\log(1+x)\right|$.

Then by (7.27), setting $\Psi_d(x,z) = R(\zeta^d(x,z))$

$$|\Psi_d(x,z)| \leq (\xi_{\ell z d^{-1/2}}(x)/\xi_0(x) - 1)^2 \left| V(x + \ell z d^{-1/2}) - V(x) \right| /2 \,.$$

Since for all $x \in \mathbb{R}$, $|\exp(x) - 1| \leq |x|(\exp(x) + 1)$, this yields,

$$|\Psi_d(x,z)| \leq 4^{-1} \left| V(x + \ell z d^{-1/2}) - V(x) \right|^3 \left(\exp\left(V(x) - V(x + \ell z d^{-1/2}) \right) + 1 \right) ,$$

which implies that

$$\int_{\mathbb{R}} |\Psi_d(x,z)| \, \pi(x) \mathrm{d}x \leq 4^{-1} \int_{\mathbb{R}} \left| V(x + \ell z d^{-1/2}) - V(x) \right|^3 \{ \pi(x) + \pi(x + \ell z d^{-1/2}) \} \mathrm{d}x \, .$$

By Hölder's inequality and using **H**1(i),

$$\int_{\mathbb{R}} |\Psi_d(x,z)| \, \pi(x) \mathrm{d}x \leq C \left(\left| \ell z d^{-1/2} \right|^3 \left(\int_{\mathbb{R}} \left| \dot{V}(x) \right|^4 \pi(x) \mathrm{d}x \right)^{3/4} + \left| \ell z d^{-1/2} \right|^{3\beta} \right) .$$

The proof follows from **H**1(ii) since $\beta > 1$.

□

For all $d \geq 1$, let X^d be distributed according to π^d, and Z^d be d-dimensional Gaussian random variable independent of X^d, set

$$\mathrm{J}^d = \left\| \sum_{i=2}^{d} \left\{ \Delta V_i^d - b^d(X_i^d, Z_i^d) \right\} \right\|_1 ,$$

where ΔV_i^d and b^d are defined in (7.5) and (7.13), respectively.

Lemma 7.14. $\lim_{d \to +\infty} \mathrm{J}^d = 0$.

Proof. Noting that $\Delta V_i^d = 2 \log \left(1 + \zeta^d \left(X_i^d, Z_i^d \right) \right)$ and using (7.26), we get

$$\mathrm{J}^d \leq \sum_{i=1}^{3} \mathrm{J}_i^d = \left\| \sum_{i=2}^{d} 2\zeta^d \left(X_i^d, Z_i^d \right) + \frac{\ell Z_i^d}{\sqrt{d}} \dot{V}(X_i^d) - \mathbb{E}\left[2\zeta^d(X_i^d, Z_i^d) \right] \right\|_1 + \left\| \sum_{i=2}^{d} \zeta^d \left(X_i^d, Z_i^d \right)^2 - \frac{\ell^2}{4d} \dot{V}^2(X_i^d) \right\|_1 + 2 \left\| \sum_{i=2}^{d} R\left(\zeta^d \left(X_i^d, Z_i^d \right) \right) \right\|_1 ,$$

where R is defined by (7.25). By Lemma 7.13(i), the first term goes to 0 as d goes to $+\infty$ since

$$\mathrm{J}_1^d \leq \sqrt{d} \left\| 2\zeta^d \left(X_1^d, Z_1^d \right) + \frac{\ell Z_1^d}{\sqrt{d}} \dot{V}(X_1^d) \right\|_2 .$$

Consider now J_2^d. We use the following decomposition for all $2 \leq i \leq d$,

$$\zeta^d(X_i^d, Z_i^d)^2 - \frac{\ell^2}{4d} \dot{V}^2(X_i^d) = \left(\zeta^d(X_i^d, Z_i^d) + \frac{\ell}{2\sqrt{d}} Z_i^d \dot{V}(X_i^d) \right)^2 - \frac{\ell}{\sqrt{d}} Z_i^d \dot{V}(X_i^d) \left(\zeta^d(X_i^d, Z_i^d) + \frac{\ell}{2\sqrt{d}} Z_i^d \dot{V}(X_i^d) \right) + \frac{\ell^2}{4d} \left\{ (Z_i^d)^2 - 1 \right\} \dot{V}^2(X_i^d) .$$

Then,

$$\mathrm{J}_2^d \leq d \left\| \zeta^d(X_1^d, Z_1^d) + \frac{\ell}{2\sqrt{d}} Z_1^d \dot{V}(X_1^d) \right\|_2^2 + \frac{\ell^2}{4d} \left\| \sum_{i=2}^{d} \dot{V}^2(X_i^d) \left\{ (Z_i^d)^2 - 1 \right\} \right\|_1 + \ell\sqrt{d} \left\| \dot{V}(X_1^d) Z_1^d \left(\zeta^d(X_1^d, Z_1^d) + \frac{\ell}{2\sqrt{d}} Z_1^d \dot{V}(X_1^d) \right) \right\|_1 .$$

Using **H**1(ii), Lemma 7.13(i) and the Cauchy-Schwarz inequality show that the first and the last term converge to zero. For the second term note that $\mathbb{E}\left[(Z_i^d)^2 - 1\right] = 0$ so that

$$d^{-1} \left\| \sum_{i=2}^{d} \dot{V}^2(X_i^d) \left\{ (Z_i^d)^2 - 1 \right\} \right\|_1 \leq d^{-1/2} \operatorname{Var}\left[\dot{V}^2(X_1^d) \left\{ (Z_1^d)^2 - 1 \right\} \right]^{1/2} \to 0 .$$

Finally, $\lim_{d\to\infty} \mathrm{J}_3^d = 0$ by (7.26) and Lemma 7.13(iii). □

Proof of Proposition 7.3. Let $q > 0$ and $\Lambda^d = -\ell d^{-1/2} Z_1^d \dot{V}(X_1^d) + \sum_{i=2}^d \Delta V_i^d$. By the triangle inequality, $\mathrm{E}^d(q) \leq \mathrm{E}_1^d(q) + \mathrm{E}_2^d(q)$ where

$$\mathrm{E}_1^d(q) = \mathbb{E}\left[\left(Z_1^d\right)^q \left| 1 \wedge \exp\left\{ \sum_{i=1}^{d} \Delta V_i^d \right\} - 1 \wedge \exp\left\{ \Lambda^d \right\} \right| \right] ,$$
$$\mathrm{E}_2^d(q) = \mathbb{E}\left[\left(Z_1^d\right)^q \left| 1 \wedge \exp\left\{ \Lambda^d \right\} - 1 \wedge \exp\left\{ \upsilon^d \right\} \right| \right] .$$

Since $t \mapsto 1 \wedge \mathrm{e}^t$ is 1-Lipschitz, by the Cauchy-Schwarz inequality we get

$$\mathrm{E}_1^d(q) \leq \left\| Z_1^d \right\|_{2q}^q \left\| \Delta V_1^d + \ell d^{-1/2} Z_1^d \dot{V}(X_1^d) \right\|_2 .$$

By Lemma 7.13(ii), $\mathrm{E}_1^d(q)$ goes to 0 as d goes to $+\infty$. Consider now $\mathrm{E}_2^d(q)$. Using again that $t \mapsto 1 \wedge \mathrm{e}^t$ is 1-Lipschitz and Lemma 7.14, $\mathrm{E}_2^d(q)$ goes to 0. □

7.4.3 Proof of Theorem 7.4

Following [JLM15], we introduce the function $\mathcal{G}$ defined on $\overline{\mathbb{R}}_+ \times \mathbb{R}$ by:

$$\mathcal{G}(a,b) = \begin{cases} \exp\left(\frac{a-b}{2}\right) \Phi\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right) & \text{if } a \in (0, +\infty) , \\ 0 & \text{if } a = +\infty , \\ \exp\left(-\frac{b}{2}\right) \mathbb{1}_{\{b>0\}} & \text{if } a = 0 , \end{cases} \tag{7.28}$$

where Φ is the cumulative distribution function of a standard normal variable, and Γ:

$$\Gamma(a,b) = \begin{cases} \Phi\left(-\frac{b}{2\sqrt{a}}\right) + \exp\left(\frac{a-b}{2}\right) \Phi\left(\frac{b}{2\sqrt{a}} - \sqrt{a}\right) & \text{if } a \in (0, +\infty) , \\ \frac{1}{2} & \text{if } a = +\infty , \\ \exp\left(-\frac{b_+}{2}\right) & \text{if } a = 0 . \end{cases} \tag{7.29}$$

Note that $\mathcal{G}$ and Γ are bounded on $\overline{\mathbb{R}}_+ \times \mathbb{R}$. $\mathcal{G}$ and Γ are used throughout Section 7.4.

Lemma 7.15. *Assume **H1** holds. For all $d \in \mathbb{N}^*$, let X^d be a random variable distributed according to π^d and Z^d be a standard Gaussian random variable in $\mathbb{R}^d$, independent of X. Then,*

$$\lim_{d\to+\infty} d\,\mathbb{E}\left[2\zeta^d(X_1^d, Z_1^d)\right] = -\frac{\ell^2}{4} I\,,$$

where I is defined in (7.6) and ζ^d in (7.14).

Proof. By (7.14),

$$\begin{aligned} d\,\mathbb{E}\left[2\zeta^d(X_1^d, Z_1^d)\right] &= 2d\mathbb{E}\left[\int_{\mathbb{R}} \sqrt{\pi\left(x + \ell d^{-1/2} Z_1^d\right)}\sqrt{\pi\left(x\right)}\mathrm{d}x - 1\right]\,, \\ &= -d\mathbb{E}\left[\int_{\mathbb{R}} \left(\sqrt{\pi\left(x + \ell d^{-1/2}\, Z_1^d\right)} - \sqrt{\pi\left(x\right)}\right)^2 \mathrm{d}x\right] = -d\,\mathbb{E}\left[\{\zeta^d(X_1^d, Z_1^d)\}^2\right]\,. \end{aligned}$$

The proof is then completed by Lemma 7.13(i). □

Proof of Theorem 7.4. By definition of $\mathcal{A}_1^d$, see (7.3),

$$\mathbb{P}\left[\mathcal{A}_1^d\right] = \mathbb{E}\left[1 \wedge \exp\left\{\sum_{i=1}^d \Delta V_i^d\right\}\right]\,,$$

where $\Delta V_i^d = V(X_{0,i}^d) - V(X_{0,i}^d + \ell d^{-1/2} Z_{1,i}^d)$ and where X_0^d is distributed according to π^d and independent of the standard d-dimensional Gaussian random variable Z_1^d. Following the same steps as in the proof of Proposition 7.3 yields:

$$\lim_{d\to+\infty} \left|\mathbb{P}\left[\mathcal{A}_1^d\right] - \mathbb{E}\left[1 \wedge \exp\left\{\Theta^d\right\}\right]\right| = 0\,, \tag{7.30}$$

where

$$\Theta^d = -\ell d^{-1/2} \sum_{i=1}^d Z_{1,i}^d \dot{V}(X_{0,i}^d) - \ell^2 \sum_{i=2}^d \dot{V}(X_{0,i}^d)^2/(4d) + 2(d-1)\mathbb{E}\left[\zeta^d(X_{0,1}^d, Z_{1,1}^d)\right]\,.$$

Conditional on X_0^d, Θ^d is a one dimensional Gaussian random variable with mean μ_d and variance σ_d^2, defined by

$$\begin{aligned} \mu_d &= -\ell^2 \sum_{i=2}^d \dot{V}(X_{0,i}^d)^2/(4d) + 2(d-1)\mathbb{E}\left[\zeta^d(X_{0,1}^d, Z_{1,1}^d)\right] \\ \sigma_d^2 &= \ell^2 d^{-1} \sum_{i=1}^d \dot{V}(X_{0,i}^d)^2\,. \end{aligned}$$

Therefore, since for any $G \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{E}[1 \wedge \exp(G)] = \Phi(\mu/\sigma) + \exp(\mu + \sigma^2/2)\Phi(-\sigma - \mu/\sigma)$, taking the expectation conditional on X_0^d, we have

$$\begin{aligned} \mathbb{E}\left[1 \wedge \exp\left\{\Theta^d\right\}\right] &= \mathbb{E}\left[\Phi(\mu_d/\sigma_d) + \exp(\mu_d + \sigma_d^2/2)\Phi(-\sigma_d - \mu_d/\sigma_d)\right] \\ &= \mathbb{E}\left[\Gamma(\sigma_d^2, -2\mu_d)\right]\,, \end{aligned}$$

where the function Γ is defined in (7.29). By Lemma 7.15 and the law of large numbers, almost surely, $\lim_{d\to+\infty}\mu_d = -\ell^2 I/2$ and $\lim_{d\to+\infty}\sigma_d^2 = \ell^2 I$. Thus, as Γ is bounded, by Lebesgue's dominated convergence theorem:

$$\lim_{d\to+\infty}\mathbb{E}\left[1\wedge\exp\left\{\Theta^d\right\}\right] = 2\Phi\left(-\ell\sqrt{I}/2\right) .$$

The proof is then completed by (7.30). □

7.4.4 Proof of Proposition 7.5

By Kolmogorov's criterion it is enough to prove that there exists a non-decreasing function $\gamma:\mathbb{R}_+\to\mathbb{R}_+$ such that for all $d\geq 1$ and all $0\leq s\leq t$,

$$\mathbb{E}\left[\left(Y_{t,1}^d - Y_{s,1}^d\right)^4\right]\leq \gamma(t)(t-s)^2 .$$

The inequality is straightforward for all $0\leq s\leq t$ such that $\lfloor ds\rfloor = \lfloor dt\rfloor$. For all $0\leq s\leq t$ such that $\lceil ds\rceil\leq\lfloor dt\rfloor$,

$$Y_{t,1}^d - Y_{s,1}^d = X_{\lfloor dt\rfloor,1}^d - X_{\lceil ds\rceil,1}^d + \frac{dt-\lfloor dt\rfloor}{\sqrt{d}}\ell Z_{\lceil dt\rceil,1}^d\mathbb{1}_{\mathcal{A}_{\lceil dt\rceil}^d} + \frac{\lceil ds\rceil - ds}{\sqrt{d}}\ell Z_{\lceil ds\rceil,1}^d\mathbb{1}_{\mathcal{A}_{\lceil ds\rceil}^d} .$$

Then by the Hölder inequality,

$$\mathbb{E}\left[\left(Y_{t,1}^d - Y_{s,1}^d\right)^4\right]\leq C\left((t-s)^2 + \mathbb{E}\left[\left(X_{\lfloor dt\rfloor,1}^d - X_{\lceil ds\rceil,1}^d\right)^4\right]\right) ,$$

where we have used

$$\frac{(dt-\lfloor dt\rfloor)^2}{d^2} + \frac{(\lceil ds\rceil - ds)^2}{d^2}\leq\frac{(dt-ds)^2 + (\lceil ds\rceil - \lfloor dt\rfloor)^2}{d^2}\leq 2(t-s)^2 .$$

The proof is completed using Lemma 7.16.

Lemma 7.16. *Assume* ***H****1. Then, there exists $C>0$ such that, for all $0\leq k_1<k_2$,*

$$\mathbb{E}\left[\left(X_{k_2,1}^d - X_{k_1,1}^d\right)^4\right]\leq C\sum_{p=2}^4\frac{(k_2-k_1)^p}{d^p} .$$

Proof. For all $0\leq k_1<k_2$,

$$\mathbb{E}\left[\left(X_{k_2,1}^d - X_{k_1,1}^d\right)^4\right] = \frac{\ell^4}{d^2}\mathbb{E}\left[\left(\sum_{k=k_1+1}^{k_2} Z_{k,1}^d - \sum_{k=k_1+1}^{k_2} Z_{k,1}^d\mathbb{1}_{(\mathcal{A}_k^d)^c}\right)^4\right] .$$

Therefore by the Hölder inequality,

$$\mathbb{E}\left[\left(X_{k_2,1}^d - X_{k_1,1}^d\right)^4\right]\leq\frac{24\ell^4}{d^2}(k_2-k_1)^2 + \frac{8\ell^4}{d^2}\mathbb{E}\left[\left(\sum_{k=k_1+1}^{k_2} Z_{k,1}^d\mathbb{1}_{(\mathcal{A}_k^d)^c}\right)^4\right] . \tag{7.31}$$

The second term can be written:

$$\mathbb{E}\left[\left(\sum_{k=k_1+1}^{k_2} Z^d_{k,1}\mathbb{1}_{(\mathcal{A}^d_k)^c}\right)^4\right] = \sum \mathbb{E}\left[\prod_{i=1}^{4} Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right] ,$$

where the sum is over all the quadruplets $(m_i)_{i=1}^4$ satisfying $m_i \in \{k_1+1,\dots,k_2\}$, $i=1,\dots,4$. The expectation on the right hand side can be upper bounded depending on the cardinality of $\{m_1,\dots,m_4\}$. For all $1 \le j \le 4$, define

$$\mathcal{I}_j = \{(m_1,\dots,m_4) \in \{k_1+1,\dots,k_2\} \ ; \ \#\{m_1,\dots,m_4\} = j\} \ . \tag{7.32}$$

Let $(m_1,m_2,m_3,m_4) \in \{k_1+1,\dots,k_2\}^4$ and $(\tilde{X}^d_k)_{k\ge 0}$ be defined as:

$$\tilde{X}^d_0 = X^d_0 \quad \text{and} \quad \tilde{X}^d_{k+1} = \tilde{X}^d_k + \mathbb{1}_{k\notin\{m_1-1,m_2-1,m_3-1,m_4-1\}}\frac{\ell}{\sqrt{d}} Z^d_{k+1}1_{\tilde{\mathcal{A}}^d_{k+1}} ,$$

with $\tilde{\mathcal{A}}^d_{k+1} = \left\{U_{k+1} \le \exp\left(\sum_{i=1}^d \Delta\tilde{V}^d_{k,i}\right)\right\}$, where for all $k \ge 0$ and all $1 \le i \le d$, $\Delta\tilde{V}_{k,i}$ is defined by

$$\Delta\tilde{V}^d_{k,i} = V\left(\tilde{X}^d_{k,i}\right) - V\left(\tilde{X}^d_{k,i} + \frac{\ell}{\sqrt{d}} Z^d_{k+1,i}\right) .$$

Note that on the event $\bigcap_{j=1}^4 \left\{\mathcal{A}^d_{m_j}\right\}^c$, the two processes $(X_k)_{k\ge 0}$ and $(\tilde{X}_k)_{k\ge 0}$ are equal. Let $\mathcal{F}$ be the σ-field generated by $\left(\tilde{X}^d_k\right)_{k\ge 0}$.

(a) $\#\{m_1,\dots,m_4\} = 4$, as the $\left\{\left(U_{m_j}, Z^d_{m_j,1},\dots,Z^d_{m_j,d}\right)\right\}_{1\le j\le 4}$ are independent conditionally to $\mathcal{F}$,

$$\begin{aligned}\mathbb{E}\left[\prod_{j=1}^{4} Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\middle|\mathcal{F}\right] &= \prod_{j=1}^{4}\mathbb{E}\left[Z^d_{m_j,1}\mathbb{1}_{\left(\tilde{\mathcal{A}}^d_{m_j}\right)^c}\middle|\mathcal{F}\right] , \\ &= \prod_{j=1}^{4}\mathbb{E}\left[Z^d_{m_j,1}\varphi\left(\sum_{i=1}^{d}\Delta\tilde{V}^d_{m_j-1,i}\right)\middle|\mathcal{F}\right] .\end{aligned}$$

where $\varphi(x) = (1-\mathrm{e}^x)_+$. Since the function φ is 1-Lipschitz, we get

$$\begin{aligned}&\left|\varphi\left(\sum_{i=1}^{d}\Delta\tilde{V}^d_{m_j-1,i}\right) - \varphi\left(-\frac{\ell}{\sqrt{d}}\dot{V}(\tilde{X}^d_{m_j-1,1})Z^d_{m_j,1} + \sum_{i=2}^{d}\Delta\tilde{V}^d_{m_j-1,i}\right)\right| \\ &\qquad\qquad \le \left|\Delta\tilde{V}^d_{m_j-1,1} + \frac{\ell}{\sqrt{d}}\dot{V}(\tilde{X}^d_{m_j-1,1})Z^d_{m_j,1}\right| .\end{aligned}$$

Then,

$$\left|\mathbb{E}\left[\prod_{j=1}^{4} Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\right]\right| \le \mathbb{E}\left[\prod_{j=1}^{4}\left\{A^d_{m_j} + B^d_{m_j}\right\}\right] ,$$

where

$$A^d_{m_j} = \mathbb{E}\left[\left|Z^d_{m_j,1}\right|\left|\Delta\tilde{V}^d_{m_j-1,1} + \frac{\ell}{\sqrt{d}}\dot{V}(\tilde{X}^d_{m_j-1,1})Z^d_{m_j,1}\right|\middle|\mathcal{F}\right],$$

$$B^d_{m_j} = \left|\mathbb{E}\left[Z^d_{m_j,1}\left(1-\exp\left\{-\frac{\ell}{\sqrt{d}}\dot{V}(\tilde{X}^d_{m_j-1,1})Z^d_{m_j,1} + \sum_{i=2}^{d}\Delta\tilde{V}^d_{m_j-1,i}\right\}\right)_+\middle|\mathcal{F}\right]\right|.$$

By the inequality of arithmetic and geometric means and convex inequalities,

$$\left|\mathbb{E}\left[\prod_{j=1}^{4} Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\right]\right| \le 8\mathbb{E}\left[\sum_{j=1}^{4}\left(A^d_{m_j}\right)^4 + \left(B^d_{m_j}\right)^4\right].$$

By Lemma 7.13(ii) and the Hölder inequality, there exists $C > 0$ such that $\mathbb{E}\left[\left(A^d_{m_j}\right)^4\right] \le Cd^{-2}$. On the other hand, by [JLM15, Lemma 6] since $Z^d_{m_j,1}$ is independent of $\mathcal{F}$,

$$B^d_{m_j} = \left|\mathbb{E}\left[\frac{\ell}{\sqrt{d}}\dot{V}(\tilde{X}^d_{m_j-1,1})\mathcal{G}\left(\frac{\ell^2}{d}\dot{V}(\tilde{X}^d_{m_j-1,1})^2, -2\sum_{i=2}^{d}\Delta\tilde{V}^d_{m_j-1,i}\right)\middle|\mathcal{F}\right]\right|,$$

where the function $\mathcal{G}$ is defined in (7.28). By **H**1(ii) and since $\mathcal{G}$ is bounded, $\mathbb{E}[(B^d_{m_j})^4] \le Cd^{-2}$. Therefore $|\mathbb{E}[\prod_{j=1}^4 Z^d_{m_j,1}\mathbb{1}_{(\mathcal{A}^d_{m_j})^c}]| \le Cd^{-2}$, showing that

$$\sum_{(m_1,m_2,m_3,m_4)\in\mathcal{I}_4}\left|\mathbb{E}\left[\prod_{i=1}^{4} Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right]\right| \le \frac{C}{d^2}\binom{k_2-k_1}{4}. \tag{7.33}$$

(b) $\#\{m_1,\dots,m_4\} = 3$, as the $\left\{\left(U_{m_j}, Z^d_{m_j,1},\dots,Z^d_{m_j,d}\right)\right\}_{1\le j\le 3}$ are independent conditionally to $\mathcal{F}$,

$$\begin{aligned}&\left|\mathbb{E}\left[\left(Z^d_{m_1,1}\right)^2\mathbb{1}_{(\mathcal{A}^d_{m_1})^c}\prod_{j=2}^{3} Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\middle|\mathcal{F}\right]\right|\\ &\qquad\le \mathbb{E}\left[\left(Z^d_{m_1,1}\right)^2\middle|\mathcal{F}\right]\left|\prod_{j=2}^{3}\mathbb{E}\left[Z^d_{m_j,1}\mathbb{1}_{\left(\tilde{\mathcal{A}}^d_{m_j}\right)^c}\middle|\mathcal{F}\right]\right| \le \left|\prod_{j=2}^{3}\mathbb{E}\left[Z^d_{m_j,1}\mathbb{1}_{\left(\tilde{\mathcal{A}}^d_{m_j}\right)^c}\middle|\mathcal{F}\right]\right|.\end{aligned}$$

Then, following the same steps as above, and using Holder's inequality yields

$$\left|\mathbb{E}\left[\prod_{j=2}^{3} Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\right]\right| \le C\mathbb{E}\left[\sum_{j=2}^{3}\left(A^d_{m_j}\right)^2 + \left(B^d_{m_j}\right)^2\right] \le Cd^{-1}$$

and

$$\sum_{(m_1,m_2,m_3,m_4)\in\mathcal{I}_3}\left|\mathbb{E}\left[\prod_{i=1}^{4} Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right]\right| \le \frac{C}{d}\binom{k_2-k_1}{3} \le \frac{C}{d}(k_2-k_1)^3. \tag{7.34}$$

(c) If $\#\{m_1, \ldots, m_4\} = 2$ two cases have to be considered:

$$\mathbb{E}\left[\left(Z^d_{m_1,1}\right)^2 \mathbb{1}_{(\tilde{\mathcal{A}}^d_{m_1})^c} \left(Z^d_{m_2,1}\right)^2 \mathbb{1}_{(\mathcal{A}^d_{m_2})^c}\right] \le \mathbb{E}\left[\left(Z^d_{m_1,1}\right)^2\right] \mathbb{E}\left[\left(Z^d_{m_2,1}\right)^2\right] \le 1\,,$$
$$\mathbb{E}\left[\left(Z^d_{m_1,1}\right)^3 \mathbb{1}_{(\mathcal{A}^d_{m_1})^c} Z^d_{m_2,1} \mathbb{1}_{(\mathcal{A}^d_{m_2})^c}\right] \le \mathbb{E}\left[\left|Z^d_{m_1,1}\right|^3\right] \mathbb{E}\left[\left|Z^d_{m_2,1}\right|\right] \le \frac{4}{\pi}\,.$$

This yields

$$\sum_{(m_1,m_2,m_3,m_4)\in\mathcal{I}_2} \left|\mathbb{E}\left[\prod_{i=1}^4 Z^d_{m_i,1} \mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right]\right| \le \left(3 + 4\cdot\frac{4}{\pi}\right)(k_2-k_1)(k_2-k_1-1) \le C(k_2-k_1)^2\,. \quad (7.35)$$

(d) If $\#\{m_1, \ldots, m_4\} = 1$: $\mathbb{E}\left[\left(Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right)^4\right] \le \mathbb{E}\left[\left(Z^d_{m_1,1}\right)^4\right] \le 3$, then

$$\sum_{(m_1,m_2,m_3,m_4)\in\mathcal{I}_1} \left|\mathbb{E}\left[\prod_{i=1}^4 Z^d_{m_i,1} \mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right]\right| \le 3(k_2-k_1)\,. \quad (7.36)$$

The proof is completed by combining (7.31) with (7.33), (7.34), (7.35) and (7.36). □

7.4.5 Proof of Proposition 7.6

We preface the proof by a preliminary lemma.

Lemma 7.17. *Assume that* ***H****1 holds. Let μ be a limit point of the sequence of laws $(\mu_d)_{d\ge1}$ of $\left\{(Y^d_{t,1})_{t\ge0},\ d\in\mathbb{N}^*\right\}$. Then for all $t\ge0$, the pushforward measure of μ by W_t is π.*

Proof. By (7.7),

$$\lim_{d\to+\infty} \mathbb{E}\left[\left|Y^d_{t,1} - X^d_{\lfloor dt\rfloor,1}\right|\right] = 0\,.$$

Since $(\mu_d)_{d\ge1}$ converges weakly to μ, for all bounded Lipschitz function $\psi : \mathbb{R}\to\mathbb{R}$, $\mathbb{E}^\mu[\psi(W_t)] = \lim_{d\to+\infty}\mathbb{E}[\psi(Y^d_{t,1})] = \lim_{d\to+\infty}\mathbb{E}[\psi(X^d_{\lfloor dt\rfloor,1})]$. The proof is completed upon noting that for all $d\in\mathbb{N}^*$ and all $t\ge0$, $X^d_{\lfloor dt\rfloor,1}$ is distributed according to π . □

Proof of Proposition 7.6. Let μ be a limit point of $(\mu_d)_{d\ge1}$. It is straightforward to show that μ is a solution to the martingale problem associated with L if for all $\phi\in C_c^\infty(\mathbb{R},\mathbb{R})$, $m\in\mathbb{N}^*$, $g:\mathbb{R}^m\to\mathbb{R}$ bounded and continuous, and $0\le t_1\le\cdots\le t_m\le s\le t$:

$$\mathbb{E}^\mu\left[\left(\phi(W_t) - \phi(W_s) - \int_s^t \mathrm{L}\phi(W_u)\,\mathrm{d}u\right) g(W_{t_1},\ldots,W_{t_m})\right] = 0\,. \quad (7.37)$$

Let $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$, $m \in \mathbb{N}^*$, $g : \mathbb{R}^m \to \mathbb{R}$ continuous and bounded, $0 \leq t_1 \leq \cdots \leq t_m \leq s \leq t$ and $\mathbf{W}_{\dot{V}} = \{w \in \mathbf{W} | w_u \notin \mathcal{D}_{\dot{V}}$ for almost every $u \in [s,t]\}$. Note first that $w \in \mathbf{W}_{\dot{V}}^c$ if and only if $\int_s^t \mathbb{1}_{\mathcal{D}_{\dot{V}}}(w_u)\mathrm{d}u > 0$. Therefore, by **H**2 and Fubini's theorem:

$$\mathbb{E}^\mu\left[\int_s^t \mathbb{1}_{\mathcal{D}_{\dot{V}}}(W_u)\mathrm{d}u\right] = \int_s^t \mathbb{E}^\mu\left[\mathbb{1}_{\mathcal{D}_{\dot{V}}}(W_u)\right]\mathrm{d}u = 0\,,$$

showing that $\mu(\mathbf{W}_{\dot{V}}^c) = 0$. We now prove that on $\mathbf{W}_{\dot{V}}$,

$$\Psi_{s,t} :\; w \mapsto \left\{\phi(w_t) - \phi(w_s) - \int_s^t \mathrm{L}\phi(w_u)\mathrm{d}u\right\} g(w_{t_1},\ldots,w_{t_m}) \tag{7.38}$$

is continuous. It is clear that it is enough to show that $w \mapsto \int_s^t \mathrm{L}\phi(w_u)\mathrm{d}u$ is continuous on $\mathbf{W}_{\dot{V}}$. So let $w \in \mathbf{W}_{\dot{V}}$ and $(w^n)_{n\geq 0}$ be a sequence in $\mathbf{W}$ which converges to w in the uniform topology on compact sets. Then by **H**2, for any u such that $w_u \notin \mathcal{D}_{\dot{V}}$, $\mathrm{L}\phi(w_u^n)$ converges to $\mathrm{L}\phi(w_u)$ when n goes to infinity and $\mathrm{L}\phi$ is bounded. Therefore by Lebesgue's dominated convergence theorem, $\int_s^t \mathrm{L}\phi(w_u^n)\mathrm{d}u$ converges to $\int_s^t \mathrm{L}\phi(w_u)\mathrm{d}u$. Hence, the map defined by (7.38) is continuous on $\mathbf{W}_{\dot{V}}$. Since $(\mu_d)_{d\geq 1}$ converges weakly to μ, by (7.16):

$$\mu\left(\Psi_{s,t}\right) = \lim_{d\to+\infty} \mu^d\left(\Psi_{s,t}\right) = 0\,,$$

which is precisely (7.37). □

7.4.6 Proof of Theorem 7.7

By Proposition 7.6, it is enough to check (7.16) to prove that μ is a solution to the martingale problem. The core of the proof of Theorem 7.7 is Proposition 7.20, for which we need two technical lemmata.

Lemma 7.18. *Let X, Y and U be $\mathbb{R}$-valued random variables and $\epsilon > 0$. Assume that U is nonnegative and bounded by 1. Let $g : \mathbb{R} \to \mathbb{R}$ be a bounded function on $\mathbb{R}$ such that for all $(x,y) \in (-\infty,-\epsilon]^2 \cup [\epsilon,+\infty)^2$, $|g(x) - g(y)| \leq C_g |x-y|$.*

(i) For all $a > 0$,

$$\begin{aligned}\mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X}) - g(Y)\right|\right] &\leq C_g \mathbb{E}\left[\mathsf{U}\left|\mathsf{X} - Y\right|\right] \\ &\quad + \operatorname{osc}(g)\left\{\mathbb{P}\left(|\mathsf{X}| \leq \epsilon\right) + a^{-1}\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-Y\right|\right] + \mathbb{P}\left(\epsilon < |\mathsf{X}| < \epsilon + a\right)\right\}\,,\end{aligned}$$

where $\operatorname{osc}(g) = \sup(g) - \inf(g)$.

(ii) If there exist $\mu \in \mathbb{R}$ and $\sigma, C_{\mathsf{X}} \in \mathbb{R}_+$ such that

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\mathsf{X} \leq x\right) - \Phi((x-\mu)/\sigma)\right| \leq C_{\mathsf{X}}\,,$$

then

$$\mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X})-g(Y)\right|\right] \leq C_g \mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-Y\right|\right] + 2\operatorname{osc}(g)\left\{C_{\mathsf{X}} + \sqrt{2\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-Y\right|\right](2\pi\sigma^2)^{-1/2}} + \epsilon(2\pi\sigma^2)^{-1/2}\right\} .$$

Proof. (i) Consider the following decomposition

$$\mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X})-g(\mathsf{Y})\right|\right] = \mathbb{E}\left[\mathsf{U}\left|(g(\mathsf{X})-g(\mathsf{Y}))\right| \mathbb{1}_{\left\{(\mathsf{X},\mathsf{Y})\in(-\infty,-\epsilon]^2\right\}\cup\left\{(\mathsf{X},\mathsf{Y})\in[\epsilon,+\infty)^2\right\}}\right] + \mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X})-g(\mathsf{Y})\right|\left(\mathbb{1}_{\{\mathsf{X}\in[-\epsilon,\epsilon]\}} + \mathbb{1}_{(\{\mathsf{X}<-\epsilon\}\cap\{\mathsf{Y}\geq-\epsilon\})\cup(\{\mathsf{X}>\epsilon\}\cap\{\mathsf{Y}\leq\epsilon\})}\right)\right] .$$

In addition, for all $a > 0$,

$$(\{\mathsf{X} < -\epsilon\} \cap \{\mathsf{Y} \geq -\epsilon\}) \cup (\{\mathsf{X} > \epsilon\} \cap \{\mathsf{Y} \leq \epsilon\}) \subset \{\epsilon < |\mathsf{X}| < \epsilon + a\} \cup (\{|\mathsf{X}| \geq \epsilon + a\} \cap \{|\mathsf{X}-\mathsf{Y}| \geq a\}) .$$

Then using that $\mathsf{U} \in [0,1)$, we get

$$\mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X})-g(\mathsf{Y})\right|\right] \leq C_g \mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right] + \operatorname{osc}(g)\left(\mathbb{P}\left(|\mathsf{X}| < \epsilon + a\right) + a^{-1}\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right]\right) .$$

(ii) The result is straightforward if $\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right] = 0$. Assume $\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right] > 0$. Combining the additional assumption and the previous result,

$$\mathbb{E}\left[\mathsf{U}\left|g(\mathsf{X})-g(\mathsf{Y})\right|\right] \leq C_g \mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right] + \operatorname{osc}(g)\left\{2C_{\mathsf{X}} + 2(\epsilon+a)(2\pi\sigma^2)^{-1/2} + a^{-1}\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right]\right\} .$$

As this result holds for all $a > 0$, the proof is concluded by setting $a = \sqrt{\mathbb{E}\left[\mathsf{U}\left|\mathsf{X}-\mathsf{Y}\right|\right](2\pi\sigma^2)^{1/2}/2}$. □

Lemma 7.19. *Assume **H**1 holds. Let X^d be distributed according to π^d and Z^d be a d-dimensional standard Gaussian random variable, independent of X^d. Then,* $\lim_{d\to+\infty} \mathrm{E}^d = 0$*, where*

$$\mathrm{E}^d = \mathbb{E}\left[\left|\dot{V}(X_1^d)\left\{\mathcal{G}\left(\frac{\ell^2}{d}\dot{V}(X_1^d)^2, 2\sum_{i=2}^d \Delta V_i^d\right) - \mathcal{G}\left(\frac{\ell^2}{d}\dot{V}(X_1^d)^2, 2\sum_{i=2}^d b_i^d\right)\right\}\right|\right] ,$$

ΔV_i^d and b_i^d are resp. given by (7.5) and (7.13).

Proof. Set for all $d \geq 1$, $\bar{Y}_d = \sum_{i=2}^d \Delta V_i^d$ and $\bar{X}_d = \sum_{i=2}^d b_i^d$. By (7.28), $\partial_b \mathcal{G}(a,b) = -\mathcal{G}(a,b)/2 + \exp(-b^2/8a)/(2\sqrt{2\pi a})$. As $\mathcal{G}$ is bounded and $x \mapsto x\exp(-x)$ is bounded on $\mathbb{R}_+$, we get $\sup_{a\in\mathbb{R}_+;|b|\geq a^{1/4}} \partial_b \mathcal{G}(a,b) < +\infty$. Therefore, there exists $C \geq 0$ such that, for all $a \in \mathbb{R}_+$and $(b_1,b_2) \in \left(-\infty,-a^{1/4}\right)^2 \cup \left(a^{1/4},+\infty\right)^2$,

$$|\mathcal{G}(a,b_1) - \mathcal{G}(a,b_2)| \leq C\,|b_1 - b_2| . \tag{7.39}$$

By definition of b_i^d (7.13), $\bar{X}_d$ may be expressed as $\bar{X}_d = \sigma_d \bar{S}_d + \mu_d$, where

$$\mu_d = 2(d-1)\mathbb{E}\left[\zeta^d(X_1^d, Z_1^d)\right] - \frac{\ell^2(d-1)}{4d}\mathbb{E}\left[\dot{V}(X_1^d)^2\right] ,$$
$$\sigma_d^2 = \ell^2 \mathbb{E}\left[\dot{V}(X_1^d)^2\right] + \frac{\ell^4}{16d}\mathbb{E}\left[\left(\dot{V}(X_1^d)^2 - \mathbb{E}\left[\dot{V}(X_1^d)^2\right]\right)^2\right] ,$$
$$\bar{S}_d = (\sqrt{d}\sigma_d)^{-1}\sum_{i=2}^{d} \beta_i^d ,$$
$$\beta_i^d = -\ell Z_i^d \dot{V}(X_i^d) - \frac{\ell^2}{4\sqrt{d}}\left(\dot{V}(X_i^d)^2 - \mathbb{E}\left[\dot{V}(X_i^d)^2\right]\right) .$$

By **H**1(ii) the Berry-Essen Theorem [Pet95, Theorem 5.7] can be applied to $\bar{S}_d$. Then, there exists a universal constant C such that for all $d > 0$,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\left(\frac{d}{d-1}\right)^{1/2} \bar{S}_d \leq x\right) - \Phi(x)\right| \leq C/\sqrt{d} .$$

It follows that

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\bar{X}_d \leq x\right) - \Phi((x-\mu_d)/\tilde{\sigma}_d)\right| \leq C/\sqrt{d} ,$$

where $\tilde{\sigma}_d^2 = (d-1)\sigma_d^2/d$. By this result and (7.39), Lemma 7.18 can be applied to obtain a constant $C \geq 0$, independent of d, such that:

$$\mathbb{E}\left[\left|\mathcal{G}\left(\ell^2\dot{V}(X_1^d)^2/d, 2\bar{Y}_d\right) - \mathcal{G}\left(\ell^2\dot{V}(X_1^d)^2/d, 2\bar{X}_d\right)\right| \middle| X_1^d\right]$$
$$\leq C\left(\varepsilon_d + d^{-1/2} + \sqrt{2\varepsilon_d(2\pi\tilde{\sigma}_d^2)^{-1/2}} + \sqrt{\ell|\dot{V}(X_1^d)|/(2\pi d^{1/2}\tilde{\sigma}_d^2)}\right) ,$$

where $\varepsilon_d = \mathbb{E}\left[\left|\bar{X}_d - \bar{Y}_d\right|\right]$. Using this result, we have

$$\mathrm{E}^d \leq C\left\{\left(\varepsilon_d + d^{-1/2} + \sqrt{2\varepsilon_d(2\pi\tilde{\sigma}_d^2)^{-1/2}}\right)\mathbb{E}\left[|\dot{V}(X_1^d)|\right]\right.$$
$$\left. + \ell^{1/2}\mathbb{E}\left[|\dot{V}(X_1^d)|^{3/2}\right](2\pi d^{1/2}\tilde{\sigma}_d^2)^{-1/2}\right\} . \quad (7.40)$$

By Lemma 7.14, ε_d goes to 0 as d goes to infinity, and by **H** 1(ii) $\lim_{d\to+\infty}\sigma_d^2 = \ell^2\mathbb{E}\left[\dot{V}(X)^2\right]$. Combining these results with (7.40), it follows that E^d goes to 0 when d goes to infinity. □

For all $n \geq 0$, define $\mathcal{F}_n^d = \sigma(\{X_k^d, k \leq n\})$ and for all $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$,

$$M_n^d(\phi) = \frac{\ell}{\sqrt{d}}\sum_{k=0}^{n-1}\phi'(X_{k,1}^d)\left\{Z_{k+1,1}^d \mathbb{1}_{\mathcal{A}_{k+1}^d} - \mathbb{E}^{\mathcal{F}_k^d}\left[Z_{k+1,1}^d\mathbb{1}_{\mathcal{A}_{k+1}^d}\right]\right\}$$
$$+ \frac{\ell^2}{2d}\sum_{k=0}^{n-1}\phi''(X_{k,1}^d)\left\{(Z_{k+1,1}^d)^2\mathbb{1}_{\mathcal{A}_{k+1}^d} - \mathbb{E}^{\mathcal{F}_k^d}\left[(Z_{k+1,1}^d)^2\mathbb{1}_{\mathcal{A}_{k+1}^d}\right]\right\} . \quad (7.41)$$

Proposition 7.20. *Assume* ***H****1 and* ***H****2 hold. Then, for all $s \le t$ and all $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$,*

$$\lim_{d\to+\infty} \mathbb{E}\left[\left|\phi(Y^d_{t,1}) - \phi(Y^d_{s,1}) - \int_s^t \mathrm{L}\phi(Y^d_{r,1})\mathrm{d}r - \left(M^d_{\lceil dt\rceil}(\phi) - M^d_{\lceil ds\rceil}(\phi)\right)\right|\right] = 0\,.$$

Proof. First, since $\mathrm{d}Y^d_{r,1} = \ell\sqrt{d} Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r$,

$$\phi(Y^d_{t,1}) - \phi(Y^d_{s,1}) = \ell\sqrt{d}\int_s^t \phi'(Y^d_{r,1}) Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r\,. \tag{7.42}$$

As ϕ is C^3, using (7.7) and a Taylor expansion, for all $r \in [s,t]$ there exists $\chi_r \in \left[X^d_{\lfloor dr\rfloor,1}, Y^d_{r,1}\right]$ such that:

$$\begin{aligned}\phi'(Y^d_{r,1}) = \phi'(X^d_{\lfloor dr\rfloor,1}) &+ \frac{\ell}{\sqrt{d}}(dr - \lfloor dr\rfloor)\phi''(X^d_{\lfloor dr\rfloor,1})Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}} \\ &+ \frac{\ell^2}{2d}(dr-\lfloor dr\rfloor)^2\phi^{(3)}(\chi_r)\left(Z^d_{\lceil dr\rceil,1}\right)^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\,.\end{aligned}$$

Plugging this expression into (7.42) yields:

$$\begin{aligned}\phi(Y^d_{t,1}) - \phi(Y^d_{s,1}) = \ell\sqrt{d}&\int_s^t \phi'(X^d_{\lfloor dr\rfloor,1})Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r \\ &+ \ell^2\int_s^t (dr-\lfloor dr\rfloor)\phi''(X^d_{\lfloor dr\rfloor,1})(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r \\ &+ \frac{\ell^3}{2\sqrt{d}}\int_s^t (dr-\lfloor dr\rfloor)^2\phi^{(3)}(\chi_r)(Z^d_{\lceil dr\rceil,1})^3\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r\,.\end{aligned}$$

As $\phi^{(3)}$ is bounded,

$$\lim_{d\to+\infty}\mathbb{E}\left[\left|d^{-1/2}\int_s^t (dr-\lfloor dr\rfloor)^2\phi^{(3)}(\chi_r)(Z^d_{\lceil dr\rceil,1})^3\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r\right|\right] = 0\,.$$

On the other hand, $I = \int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})(dr-\lfloor dr\rfloor)(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r = I_1 + I_2$ with

$$\begin{aligned}I_1 &= \int_s^{\lceil ds\rceil/d} + \int_{\lfloor dt\rfloor/d}^t \phi''(X^d_{\lfloor dr\rfloor,1})(dr - \lfloor dr\rfloor - 1/2)(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r \\ I_2 &= \frac{1}{2}\int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r\,.\end{aligned}$$

Note that

$$\begin{aligned}I_1 = \frac{1}{2d}(\lceil ds\rceil - ds)(ds-\lfloor ds\rfloor)\phi''(X^d_{\lfloor ds\rfloor,1})(Z^d_{\lceil ds\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}} \\ + \frac{1}{2d}(\lceil dt\rceil - dt)(dt-\lfloor dt\rfloor)\phi''(X^d_{\lfloor dt\rfloor,1})(Z^d_{\lceil dt\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}\end{aligned}$$

showing, as ϕ'' is bounded, that $\lim_{d\to+\infty}\mathbb{E}[|I_1|]=0$. Therefore,

$$\lim_{d\to+\infty}\mathbb{E}\left[\left|\phi(Y^d_{t,1})-\phi(Y^d_{s,1})-I_{s,t}\right|\right]=0\,,$$

where

$$I_{s,t}=\int_s^t\left\{\ell\sqrt{d}\phi'(X^d_{\lfloor dr\rfloor,1})Z^d_{\lceil dr\rceil,1}+\ell^2\phi''(X^d_{\lfloor dr\rfloor,1})(Z^d_{\lceil dr\rceil,1})^2/2\right\}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\mathrm{d}r\,.$$

Write

$$I_{s,t}-\int_s^t\mathrm{L}\phi(Y^d_{r,1})\mathrm{d}r-\left(M^d_{\lceil dt\rceil}(\phi)-M^d_{\lceil ds\rceil}(\phi)\right)=T^d_1+T^d_2+T^d_3-T^d_4+T^d_5\,,$$

where

$$\begin{aligned}
T^d_1&=\int_s^t\phi'(X^d_{\lfloor dr\rfloor,1})\left(\ell\sqrt{d}\,\mathbb{E}^{\mathcal{F}^d_{\lfloor dr\rfloor}}\left[Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\right]+\frac{h(\ell)}{2}\dot{V}(X^d_{\lfloor dr\rfloor,1})\right)\mathrm{d}r\,,\\
T^d_2&=\int_s^t\phi''(X^d_{\lfloor dr\rfloor,1})\left(\frac{\ell^2}{2}\,\mathbb{E}\left[(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right]-\frac{h(\ell)}{2}\right)\mathrm{d}r\,,\\
T^d_3&=\int_s^t\left(\mathrm{L}\phi(Y^d_{\lfloor dr\rfloor/d,1})-\mathrm{L}\phi(Y^d_{r,1})\right)\mathrm{d}r\,,\\
T^d_4&=\frac{\ell(\lceil dt\rceil-dt)}{\sqrt{d}}\phi'(X^d_{\lfloor dt\rfloor,1})\left(Z^d_{\lceil dt\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}-\mathbb{E}\left[Z^d_{\lceil dt\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}\middle|\mathcal{F}^d_{\lfloor dt\rfloor}\right]\right)\\
&\quad+\frac{\ell^2(\lceil dt\rceil-dt)}{2d}\phi''(X^d_{\lfloor dt\rfloor,1})\left((Z^d_{\lceil dt\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}-\mathbb{E}\left[(Z^d_{\lceil dt\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}\middle|\mathcal{F}^d_{\lfloor dt\rfloor}\right]\right)\,,\\
T^d_5&=\frac{\ell(\lceil ds\rceil-ds)}{\sqrt{d}}\phi'(X^d_{\lfloor ds\rfloor,1})\left(Z^d_{\lceil ds\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}-\mathbb{E}\left[Z^d_{\lceil ds\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}\middle|\mathcal{F}^d_{\lfloor ds\rfloor}\right]\right)\\
&\quad+\frac{\ell^2(\lceil ds\rceil-ds)}{2d}\phi''(X^d_{\lfloor ds\rfloor,1})\left((Z^d_{\lceil ds\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}-\mathbb{E}\left[(Z^d_{\lceil ds\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}\middle|\mathcal{F}^d_{\lfloor ds\rfloor}\right]\right)\,.
\end{aligned}$$

It is now proved that for all $1\leq i\leq 5$, $\lim_{d\to+\infty}\mathbb{E}[|T^d_i|]=0$. First, as ϕ' and ϕ'' are bounded,

$$\mathbb{E}\left[\left|T^d_4\right|+\left|T^d_5\right|\right]\leq Cd^{-1/2}\,. \tag{7.43}$$

Denote for all $r\in[s,t]$ and $d\geq 1$,

$$\begin{aligned}
\Delta V^d_{r,i}&=V\left(X^d_{\lfloor dr\rfloor,i}\right)-V\left(X^d_{\lfloor dr\rfloor,i}+\ell d^{-1/2}Z^d_{\lceil dr\rceil,i}\right)\\
\Xi^d_r&=1\wedge\exp\left\{-\ell Z^d_{\lceil dr\rceil,1}\dot{V}(X^d_{\lfloor dr\rfloor,1})/\sqrt{d}+\sum_{i=2}^d b^d_{\lfloor dr\rfloor,i}\right\}\,,\\
\Upsilon^d_r&=1\wedge\exp\left\{-\ell Z^d_{\lceil dr\rceil,1}\dot{V}(X^d_{\lfloor dr\rfloor,1})/\sqrt{d}+\sum_{i=2}^d\Delta V^d_{r,i}\right\}\,,
\end{aligned}$$

where for all $k, i \geq 0$, $b^d_{k,i} = b^d(X^d_{k,i}, Z^d_{k+1,i})$, and for all $x, z \in \mathbb{R}$, $b^d(x, y)$ is given by (7.13). By the triangle inequality,

$$\left|T^d_1\right| \leq \int_s^t \left|\phi'(X^d_{\lfloor dr\rfloor,1})\right| (A_{1,r} + A_{2,r} + A_{3,r})\mathrm{d}r\,, \tag{7.44}$$

where

$$\begin{aligned} A_{1,r} &= \left|\ell\sqrt{d}\;\mathbb{E}\left[Z^d_{\lceil dr\rceil,1}\left(\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}} - \Upsilon^d_r\right)\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right]\right|\,, \\ A_{2,r} &= \left|\ell\sqrt{d}\;\mathbb{E}\left[Z^d_{\lceil dr\rceil,1}\left(\Upsilon^d_r - \Xi^d_r\right)\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right]\right|\,, \\ A_{3,r} &= \left|\ell\sqrt{d}\;\mathbb{E}\left[Z^d_{\lceil dr\rceil,1}\Xi^d_r\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right] + \dot{V}(X^d_{\lfloor dr\rfloor,1})h(\ell)/2\right|\,. \end{aligned}$$

Since $t \mapsto 1 \wedge \exp(t)$ is 1-Lipschitz, by Lemma 7.13(ii) $\mathbb{E}[\left|A^d_{1,r}\right|]$ goes to 0 as $d \to +\infty$ for almost all r. So by the Fubini theorem, the first term in (7.44) goes to 0 as $d \to +\infty$. For $A^d_{2,r}$, by [JLM15, Lemma 6],

$$\mathbb{E}\left[\left|A^d_{2,r}\right|\right] \leq \mathbb{E}\left[\left|\ell^2\dot{V}(X^d_{\lfloor dr\rfloor,1})\left\{\mathcal{G}\left(\frac{\ell^2\dot{V}(X^d_{\lfloor dr\rfloor,1})^2}{d}, 2\sum_{i=2}^d \Delta V^d_{r,i}\right)\right.\right.\right.$$
$$\left.\left.\left. -\mathcal{G}\left(\frac{\ell^2\dot{V}(X^d_{\lfloor dr\rfloor,1})^2}{d}, 2\sum_{i=2}^d b^d_{\lfloor dr\rfloor,i}\right)\right\}\right|\right]\,,$$

where $\mathcal{G}$ is defined in (7.28). By Lemma 7.19, this expectation goes to zero when d goes to infinity. Then by the Fubini theorem and the Lebesgue dominated convergence theorem, the second term of (7.44) goes 0 as $d \to +\infty$. For the last term, by [JLM15, Lemma 6] again:

$$\begin{aligned} \ell\sqrt{d}\;\mathbb{E}\left[Z^d_{\lceil dr\rceil,1}\Xi^d_r\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right] &= -\ell^2\dot{V}(X^d_{\lfloor dr\rfloor,1}) \\ &\times \mathcal{G}\left(\frac{\ell^2}{d}\sum_{i=1}^d \dot{V}(X^d_{\lfloor dr\rfloor,i})^2, \frac{\ell^2}{2d}\sum_{i=2}^d \dot{V}(X^d_{\lfloor dr\rfloor,i})^2 - 4(d-1)\mathbb{E}\left[\zeta^d(X,Z)\right]\right)\,, \end{aligned} \tag{7.45}$$

where X is distributed according to π and Z is a standard Gaussian random variable independent of X. As $\mathcal{G}$ is continuous on $\mathbb{R}_+ \times \mathbb{R} \setminus \{0, 0\}$ (see [JLM15, Lemma 2]), by **H**1(ii), Lemma 7.15 and the law of large numbers, almost surely,

$$\begin{aligned} \lim_{d\to+\infty} \ell^2\mathcal{G}&\left(\frac{\ell^2}{d}\sum_{i=1}^d \dot{V}(X^d_{\lfloor dr\rfloor,i})^2, \frac{\ell^2}{2d}\sum_{i=2}^d \dot{V}(X^d_{\lfloor dr\rfloor,i})^2 - 4(d-1)\mathbb{E}\left[\zeta^d(X,Z)\right]\right) \\ &= \ell^2\mathcal{G}\left(\ell^2\mathbb{E}[\dot{V}(X)^2], \ell^2\mathbb{E}[\dot{V}(X)^2]\right) = h(\ell)/2\,, \end{aligned} \tag{7.46}$$

where $h(\ell)$ is defined in (7.10). Therefore by Fubini's Theorem, (7.45) and Lebesgue's dominated convergence theorem, the last term of (7.44) goes to 0 as d goes to infinity.

The proof for T_2^d follows the same lines. By the triangle inequality,

$$\left|T_2^d\right| \le \left|\int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})(\ell^2/2)\ \mathbb{E}\left[(Z^d_{\lceil dr\rceil,1})^2\left(\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}} - \Xi^d_r\right)\Big|\mathcal{F}^d_{\lfloor dr\rfloor}\right]\mathrm{d}r\right| \\ + \left|\int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})\left((\ell^2/2)\ \mathbb{E}\left[(Z^d_{\lceil dr\rceil,1})^2\Xi^d_r\Big|\mathcal{F}^d_{\lfloor dr\rfloor}\right] - h(\ell)/2\right)\mathrm{d}r\right|. \quad (7.47)$$

By Fubini's Theorem, Lebesgue's dominated convergence theorem and Proposition 7.3, the expectation of the first term goes to zero when d goes to infinity. For the second term, by [JLM15, Lemma 6 (A.5)],

$$(\ell^2/2)\mathbb{E}\left[(Z^d_{\lceil dr\rceil,1})^2 1\wedge\exp\left\{-\frac{\ell Z^d_{\lceil dr\rceil,1}}{\sqrt{d}}\dot V(X^d_{\lfloor dr\rfloor,1}) + \sum_{i=2}^d b^d_{\lfloor dr\rfloor,i}\right\}\Big|\mathcal{F}^d_{\lfloor dr\rfloor}\right] \\ = (B_1+B_2-B_3)/2\,, \quad (7.48)$$

where

$$B_1 = \ell^2\Gamma\left(\frac{\ell^2}{d}\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2, \frac{\ell^2}{2d}\sum_{i=2}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2 - 4(d-1)\mathbb{E}\left[\zeta^d(X,Z)\right]\right),$$

$$B_2 = \frac{\ell^4\dot V(X^d_{\lfloor dr\rfloor,1})^2}{d}\mathcal{G}\left(\frac{\ell^2}{d}\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2, \frac{\ell^2}{2d}\sum_{i=2}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2 - 4(d-1)\mathbb{E}\left[\zeta^d(X,Z)\right]\right),$$

$$B_3 = \frac{\ell^4\dot V(X^d_{\lfloor dr\rfloor,1})^2}{d}\left(2\pi\ell^2\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2/d\right)^{-1/2} \\ \times\exp\left\{-\frac{\left[-(d-1)\mathbb{E}[2\zeta^d(X,Z)] + (\ell^2/(4d))\sum_{i=2}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2\right]^2}{2\ell^2\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2/d}\right\},$$

where Γ is defined in (7.29). As Γ is continuous on $\mathbb{R}_+\times\mathbb{R}\setminus\{0,0\}$ (see [JLM15, Lemma 2]), by **H**1(ii), Lemma 7.15 and the law of large numbers, almost surely,

$$\lim_{d\to+\infty}\ell^2\Gamma\left(\frac{\ell^2}{d}\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2, \frac{\ell^2}{2d}\sum_{i=2}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2 - 4(d-1)\mathbb{E}\left[\zeta^d(X,Z)\right]\right) \\ = \ell^2\Gamma\left(\ell^2\mathbb{E}[\dot V(X)^2], \ell^2\mathbb{E}[\dot V(X)^2]\right) = h(\ell)\,. \quad (7.49)$$

By Lemma 7.15, by **H**1(ii) and the law of large numbers, almost surely,

$$\lim_{d\to+\infty}\exp\left\{-\frac{\left[-(d-1)\mathbb{E}[2\zeta^d(X,Z)] + (\ell^2/(4d))\sum_{i=2}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2\right]^2}{2\ell^2\sum_{i=1}^d \dot V(X^d_{\lfloor dr\rfloor,i})^2/d}\right\} \\ = \exp\left\{-\frac{\ell^2}{8}\mathbb{E}[\dot V(X)^2]\right\}.$$

Then, as $\mathcal{G}$ is bounded on $\mathbb{R}_+ \times \mathbb{R}$,

$$\lim_{d\to+\infty} \mathbb{E}\left[\left|\int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})\,(B_2 - B_3)\,\mathrm{d}r\right|\right] = 0\,. \tag{7.50}$$

Therefore, by Fubini's Theorem, (7.48), (7.49), (7.50) and Lebesgue's dominated convergence theorem, the second term of (7.47) goes to 0 as d goes to infinity. Write $T_3^d = (h(\ell)/2)(T^d_{3,1} - T^d_{3,2})$ where

$$T^d_{3,1} = \int_s^t \left\{\phi''\left(X^d_{\lfloor dr\rfloor,1}\right) - \phi''\left(Y^d_{r,1}\right)\right\}\mathrm{d}r\,,$$
$$T^d_{3,2} = \int_s^t \left\{\dot V\left(X^d_{\lfloor dr\rfloor,1}\right)\phi'\left(X^d_{\lfloor dr\rfloor,1}\right) - \dot V\left(Y^d_{r,1}\right)\phi'\left(Y^d_{r,1}\right)\right\}\mathrm{d}r\,.$$

It is enough to show that $\mathbb{E}[\left|T^d_{3,1}\right|]$ and $\mathbb{E}[\left|T^d_{3,2}\right|]$ go to 0 when d goes to infinity to conclude the proof. By (7.7) and the mean value theorem, for all $r \in [s,t]$ there exists $\chi_r \in \left[X^d_{\lfloor dr\rfloor,1}, Y^d_{r,1}\right]$ such that

$$\phi''\left(X^d_{\lfloor dr\rfloor,1}\right) - \phi''\left(Y^d_{r,1}\right) = \phi^{(3)}\left(\chi_r\right)(dr - \lfloor dr\rfloor)(\ell/\sqrt{d})Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\,.$$

Since $\phi^{(3)}$ is bounded, it follows that $\lim_{d\to+\infty}\mathbb{E}[|T^d_{3,1}|] = 0$. On the other hand,

$$T^d_{3,2} = \int_s^t \left\{\dot V\left(X^d_{\lfloor dr\rfloor,1}\right) - \dot V\left(Y^d_{r,1}\right)\right\}\phi'\left(X^d_{\lfloor dr\rfloor,1}\right)\mathrm{d}r + \int_s^t \left\{\phi'\left(X^d_{\lfloor dr\rfloor,1}\right) - \phi'\left(Y^d_{r,1}\right)\right\}\dot V\left(Y^d_{r,1}\right)\mathrm{d}r\,.$$

Since ϕ' has a bounded support, by **H**2, Fubini's theorem, and Lebesgue's dominated convergence theorem, the expectation of the absolute value of the first term goes to 0 as d goes to infinity. The second term is dealt with following the same steps as for $T^d_{3,1}$ and using **H**1(ii). □

Proof of Theorem 7.7. By Proposition 7.5, Proposition 7.6 and Proposition 7.20, it is enough to prove that for all $\phi \in C_c^\infty(\mathbb{R},\mathbb{R})$, $p \geq 1$, all $0 \leq t_1 \leq \cdots \leq t_p \leq s \leq t$ and $g : \mathbb{R}^p \to \mathbb{R}$ bounded and continuous function,

$$\lim_{d\to+\infty} \mathbb{E}\left[(M^d_{\lceil dt\rceil}(\phi) - M^d_{\lceil ds\rceil}(\phi))g(Y^d_{t_1},\ldots,Y^d_{t_p})\right] = 0\,,$$

where for $n \geq 1$, $M^d_n(\phi)$ is defined in (7.41). But this result is straightforward taking successively the conditional expectations with respect to $\mathcal{F}_k$, for $k = \lceil dt\rceil, \ldots, \lceil ds\rceil$. □

7.5 Proofs of Section 7.3

7.5.1 Proof of Theorem 7.9

The proof of this theorem follows the same steps as the the proof of Theorem 7.4. Note that ξ_θ and ξ_0, given by (7.11), are well defined on $\mathcal{I} \cap \{x \in \mathbb{R} \mid x + \mathrm{r}\theta \in \mathcal{I}\}$. Let the function $\upsilon : \mathbb{R}^2 \to \mathbb{R}$ be defined for $x, \theta \in \mathbb{R}$ by

$$\upsilon(x,\theta) = \mathbb{1}_{\mathcal{I}}(x + \mathrm{r}\theta)\mathbb{1}_{\mathcal{I}}(x + (1-\mathrm{r})\theta) \,. \tag{7.51}$$

Lemma 7.21. *Assume* ***G****4 holds. Then, there exists $C > 0$ such that for all $\theta \in \mathbb{R}$,*

$$\left(\int_{\mathcal{I}} \left(\{\xi_\theta(x) - \xi_0(x)\}\,\upsilon(x,\theta) + \theta\dot{V}(x)\xi_0(x)/2\right)^2 \mathrm{d}x\right)^{1/2} \leq C|\theta|^\beta \,.$$

Proof. The proof follows as Lemma 7.2 and is omitted. □

Lemma 7.22. *Assume that* ***G****4 holds. Let X be a random variable distributed according to π and Z be a standard Gaussian random variable independent of X. Define*

$$\mathcal{D}_{\mathcal{I}} = \{X + \mathrm{r}\ell d^{-1/2} Z \in \mathcal{I}\} \cap \{X + (1-\mathrm{r})\ell d^{-1/2} Z \in \mathcal{I}\} \,.$$

Then,

(i) $\lim_{d\to+\infty} d \left\| \mathbb{1}_{\mathcal{D}_{\mathcal{I}}} \zeta^d(X,Z) + \ell Z \dot{V}(X)/(2\sqrt{d}) \right\|_2^2 = 0.$

(ii) Let p be given by ***G****4(i). Then,*

$$\lim_{d\to+\infty} \sqrt{d} \left\| \mathbb{1}_{\mathcal{D}_{\mathcal{I}}} \left\{V(X) - V(X + \ell Z/\sqrt{d})\right\} + \ell Z \dot{V}(X)/\sqrt{d} \right\|_p = 0 \,.$$

(iii) $\lim_{d\to\infty} d \left\| \mathbb{1}_{\mathcal{D}_{\mathcal{I}}} \left(\log(1 + \zeta_d(X,Z)) - \zeta^d(X,Z) + [\zeta^d]^2(X,Z)/2\right) \right\|_1 = 0,$

where ζ^d is given by (7.14).

Proof. Note by definition of ζ^d and ξ_θ (7.11), for $x \in \mathcal{I}$ and $x + \mathrm{r}\ell d^{-1/2} z \in \mathcal{I}$,

$$\zeta^d(x,z) = \xi_{\ell z d^{-1/2}}(x)/\xi_0(x) - 1 \,. \tag{7.52}$$

Using Lemma 7.21,

$$\begin{aligned}
&\left\| \mathbb{1}_{\mathcal{D}_{\mathcal{I}}} \zeta^d(X,Z) + \ell Z \dot{V}(X)/(2\sqrt{d}) \right\|_2^2 \\
&\quad = \mathbb{E}\left[\int_{\mathcal{I}} \left(\upsilon(x, \ell Z d^{-1/2})\,\{\xi_{\ell Z d^{-1/2}}(x) - \xi_0(x)\} + \ell Z \dot{V}(x)\xi_0(x)/(2\sqrt{d})\right)^2 \mathrm{d}x\right] \\
&\quad \leq C\ell^{2\beta} d^{-\beta} \mathbb{E}\left[|Z|^{2\beta}\right] .
\end{aligned}$$

The proof of (i) is completed using $\beta > 1$. For (ii), write for all $x \in \mathcal{I}$ and $x+\ell z d^{-1/2} z \in \mathcal{I}$, $\Delta V(x,z) = V(x) - V(x+\ell z d^{-1/2})$. By **G**4(i)

$$\left\| \mathbb{1}_{\mathcal{D}_{\mathcal{I}}} \Delta V(X,Z) + \ell Z \dot{V}(X)/\sqrt{d} \right\|_p^p = \mathbb{E}\left[\int_{\mathcal{I}} \left(\upsilon(x, \ell Z d^{-1/2}) \Delta V(X,Z) + \ell Z \dot{V}(x)/\sqrt{d} \right)^p \pi(x) \mathrm{d}x \right]$$
$$\leq C \ell^{\beta p} d^{-\beta p/2} \mathbb{E}\left[|Z|^{\beta p} \right]$$

and the proof of (ii) follows from $\beta > 1$. For (iii), note that for all $x > 0$, $u \in [0,x]$, $|(x-u)(1+u)^{-1}| \leq |x|$, and the same inequality holds for $x \in (-1,0]$ and $u \in [x,0]$. Then by (7.25) and (7.26), for all $x > -1$,

$$\left| \log(1+x) - x + x^2/2 \right| = |R(x)| \leq x^2 \left| \log(1+x) \right| .$$

Then by (7.52), for $x \in \mathcal{I}$ and $x + \ell d^{-1/2} z \in \mathcal{I}$,

$$\begin{aligned}
&\left| \log(1+\zeta_d(x,z)) - \zeta^d(x,z) + [\zeta^d]^2(x,z)/2 \right| \\
&\qquad \leq (\xi_{\ell z d^{-1/2}}(x)/\xi_0(x) - 1)^2 \left| \log(\xi_{\ell z d^{-1/2}}(x)/\xi_0(x)) \right| , \\
&\qquad \leq (\xi_{\ell z d^{-1/2}}(x)/\xi_0(x) - 1)^2 \left| V(x+\ell z d^{-1/2}) - V(x) \right| /2 .
\end{aligned}$$

Since for all $x \in \mathbb{R}$, $|\exp(x) - 1| \leq |x|(\exp(x)+1)$, this yields,

$$\begin{aligned}
&\left| \log(1+\zeta_d(x,z)) - \zeta^d(x,z) + [\zeta^d]^2(x,z)/2 \right| \\
&\qquad \leq \left| V(x+\ell z d^{-1/2}) - V(x) \right|^3 \left(\exp\left(V(x) - V(x+\ell z d^{-1/2}) \right) + 1 \right) /4 .
\end{aligned}$$

Therefore,

$$\int_{\mathcal{I}} \upsilon(x, \ell z d^{-1/2}) \left| \log(1+\zeta_d(x,z)) - \zeta^d(x,z) + [\zeta^d]^2(x,z)/2 \right| \pi(x) \mathrm{d}x \leq (I_1 + I_2)/4 ,$$

where

$$\begin{aligned}
I_1 &= \int_{\mathcal{I}} \upsilon(x, \ell z d^{-1/2}) \left| V(x+\ell z d^{-1/2}) - V(x) \right|^3 \pi(x) \mathrm{d}x \\
I_2 &= \int_{\mathcal{I}} \upsilon(x, \ell z d^{-1/2}) \left| V(x+\ell z d^{-1/2}) - V(x) \right|^3 \pi(x+\ell z d^{-1/2}) \mathrm{d}x .
\end{aligned}$$

By Hölder's inequality, a change of variable and using **G**4(i),

$$I_1 + I_2 \leq C \left(\left| \ell z d^{-1/2} \right|^3 \left(\int_{\mathcal{I}} \left| \dot{V}(x) \right|^4 \pi(x) \mathrm{d}x \right)^{3/4} + \left| \ell z d^{-1/2} \right|^{3\beta} \right) .$$

The proof follows from **G**4(ii) and $\beta > 1$.

□

For ease of notation, write for all $d \geq 1$ and $i, j \in \{1, \dots, d\}$,

$$\mathcal{D}^d_{\mathcal{I},j} = \left\{X^d_j + \mathrm{r}\ell d^{-1/2} Z^d_j \in \mathcal{I}\right\} \cap \left\{X^d_j + (1-\mathrm{r})\ell d^{-1/2} Z^d_j \in \mathcal{I}\right\} ,$$
$$\mathcal{D}^d_{\mathcal{I},i:j} = \bigcap_{k=i}^{j} \mathcal{D}^d_{\mathcal{I},k} . \tag{7.53}$$

Lemma 7.23. *Assume that* **G**4 *holds. For all $d \geq 1$, let X^d be distributed according to π^d, and Z^d be d-dimensional Gaussian random variable independent of X^d. Then,* $\lim_{d\to+\infty} \mathrm{J}^d_{\mathcal{I}} = 0$ *where*

$$\mathrm{J}^d_{\mathcal{I}} = \left\| \mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \sum_{i=2}^{d} \left\{ \left(\Delta V^d_i + \frac{\ell Z^d_i}{\sqrt{d}} \dot{V}(X^d_i) \right) - 2\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},i}} \zeta^d(X^d_i, Z^d_i) \right] + \frac{\ell^2}{4d} \dot{V}^2(X^d_i) \right\} \right\|_1 .$$

Proof. The proof follows the same lines as the proof of Lemma 7.14 and is omitted. □

Define for all $d \geq 1$,

$$\mathrm{E}^d_{\mathcal{I}} = \mathbb{E}\left[\left(Z^d_1\right)^2 \left| \mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1:d}} 1 \wedge \exp\left\{ \sum_{i=1}^{d} \Delta V^d_i \right\} \right.\right.$$
$$\left.\left. -1 \wedge \exp\left\{ -\ell d^{-1/2} Z^d_1 \dot{V}(X^d_1) + \sum_{i=2}^{d} b^d_{\mathcal{I}}(X^d_i, Z^d_i) \right\} \right| \right] ,$$

where ΔV^d_i is given by (7.5), for all $x \in \mathcal{I}$, $z \in \mathbb{R}$,

$$b^d_{\mathcal{I}}(x, z) = -\frac{\ell z}{\sqrt{d}} \dot{V}(x) + 2\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1}} \zeta^d(X^d_1, Z^d_1) \right] - \frac{\ell^2}{4d} \dot{V}^2(x) , \tag{7.54}$$

and ζ^d is given by (7.14).

Proposition 7.24. *Assume* **G**4 *holds. Let X^d be a random variable distributed according to π^d and Z^d be a zero-mean standard Gaussian random variable, independent of X. Then* $\lim_{d\to+\infty} \mathrm{E}^d_{\mathcal{I}} = 0$.

Proof. Let $\Lambda^d = -\ell d^{-1/2} Z^d_1 \dot{V}(X^d_1) + \sum_{i=2}^{d} \Delta V^d_i$. By the triangle inequality, $\mathrm{E}^d \leq \mathrm{E}^d_1 + \mathrm{E}^d_2 + \mathrm{E}^d_3$ where

$$\mathrm{E}^d_{1,\mathcal{I}} = \mathbb{E}\left[\left(Z^d_1\right)^2 \mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1:d}} \left| 1 \wedge \exp\left\{ \sum_{i=1}^{d} \Delta V^d_i \right\} - 1 \wedge \exp\left\{ \Lambda^d \right\} \right| \right] ,$$
$$\mathrm{E}^d_{2,\mathcal{I}} = \mathbb{E}\left[\left(Z^d_1\right)^2 \mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| 1 \wedge \exp\left\{ \Lambda^d \right\} - 1 \wedge \exp\left\{ -\ell d^{-1/2} Z^d_1 \dot{V}(X^d_1) + \sum_{i=2}^{d} b^d(X^d_i, Z^d_i) \right\} \right| \right] ,$$
$$\mathrm{E}^d_{3,\mathcal{I}} = \mathbb{E}\left[\left(Z^d_1\right)^2 \mathbb{1}_{\left(\mathcal{D}^d_{\mathcal{I},2:d}\right)^c} 1 \wedge \exp\left\{ -\ell d^{-1/2} Z^d_1 \dot{V}(X^d_1) + \sum_{i=2}^{d} b^d(X^d_i, Z^d_i) \right\} \right] ,$$

Since $t \mapsto 1 \wedge \mathrm{e}^t$ is 1-Lipschitz, by the Cauchy-Schwarz inequality we get

$$\mathrm{E}^d_{1,\mathcal{I}} \leq \mathbb{E}\left[\left(Z^d_1\right)^2 \mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1}} \left|\Delta V^d_1 + \ell d^{-1/2} Z^d_1 \dot{V}(X^d_1)\right|\right] \leq \|Z^d_1\|_4^2 \left\|\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1}} \Delta V^d_1 + \ell d^{-1/2} Z^d_1 \dot{V}(X^d_1)\right\|_2 .$$

By Lemma 7.13(ii), $\mathrm{E}^d_{1,\mathcal{I}}$ goes to 0 as d goes to $+\infty$. Using again that $t \mapsto 1 \wedge \mathrm{e}^t$ is 1-Lipschitz and Lemma 7.23, $\mathrm{E}^d_{2,\mathcal{I}}$ goes to 0 as well. Note that, as Z^d_1 and $\mathbb{1}_{(\mathcal{D}^d_{\mathcal{I},2:d})^c}$ are independent, by (7.18),

$$\mathrm{E}^d_{3,\mathcal{I}} \leq d\mathbb{P}\left(\left\{\mathcal{D}^d_{\mathcal{I},1}\right\}^c\right) \leq C d^{1-\gamma/2} .$$

Therefore, $\mathrm{E}^d_{3,\mathcal{I}}$ goes to 0 as d goes to $+\infty$ by **G**4(iii). □

Lemma 7.25. *Assume* ***G****4 holds. For all $d \in \mathbb{N}^*$, let X^d be a random variable distributed according to π^d and Z^d be a standard Gaussian random variable in $\mathbb{R}^d$, independent of X. Then,*

$$\lim_{d \to +\infty} 2d\,\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1}} \zeta^d(X^d_1, Z^d_1)\right] = -\frac{\ell^2}{4} I ,$$

where I is defined in (7.6) and ζ^d in (7.14).

Proof. Noting that for all $\theta \in \mathbb{R}$,

$$\int_{\mathcal{I}} \mathbb{1}_{\mathcal{I}}(x + \mathrm{r}\theta)\mathbb{1}_{\mathcal{I}}(x + (1-\mathrm{r})\theta)\pi(x+\theta)\mathrm{d}x = \int_{\mathcal{I}} \mathbb{1}_{\mathcal{I}}(x + (\mathrm{r}-1)\theta)\mathbb{1}_{\mathcal{I}}(x - \mathrm{r}\theta)\pi(x)\mathrm{d}x .$$

the proof follows the same steps as the the proof of Lemma 7.15 and is omitted. □

Proof of Theorem 7.9. The proof follows the same lines as the proof of Theorem 7.4 and is therefore omitted. □

7.5.2 Proof of Proposition 7.10

As for the proof of Proposition 7.5, the proof follows from Lemma 7.26.

Lemma 7.26. *Assume* ***G****4. Then, there exists $C > 0$ such that, for all $0 \leq k_1 < k_2$,*

$$\mathbb{E}\left[\left(X^d_{k_2,1} - X^d_{k_1,1}\right)^4\right] \leq C \sum_{p=2}^{4} \frac{(k_2 - k_1)^p}{d^p} .$$

Proof. We use the same decomposition of $\mathbb{E}[(X^d_{k_2,1} - X^d_{k_1,1})^4]$ as in the proof of Lemma 7.16 so that we only need to upper bound the following term:

$$d^{-2}\mathbb{E}\left[\left(\sum_{k=k_1+1}^{k_2} Z^d_{k,1}\mathbb{1}_{(\mathcal{A}^d_k)^c}\right)^4\right] = d^{-2}\sum \mathbb{E}\left[\prod_{i=1}^{4} Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right] ,$$

where the sum is over all the quadruplets $(m_p)_{p=1}^4$ satisfying $m_p \in \{k_1+1,\dots,k_2\}$, $p=1,\dots,4$. Let $(m_1,m_2,m_3,m_4)\in\{k_1+1,\dots,k_2\}^4$ and $(\tilde{X}_k^d)_{k\geq 0}$ be defined as:

$$\tilde{X}_0^d = X_0^d \quad\text{and}\quad \tilde{X}_{k+1}^d = \tilde{X}_k^d + \mathbb{1}_{k\notin\{m_1-1,m_2-1,m_3-1,m_4-1\}}\ell d^{-1/2} Z_{k+1}^d \mathbb{1}_{\tilde{\mathcal{A}}_{k+1}^d}\,,$$

where for all $k\geq 0$ and all $1\leq i\leq d$,

$$\begin{aligned}
\tilde{\mathcal{A}}_{k+1}^d &= \left\{U_{k+1}\leq \exp\left(\sum_{i=1}^d \Delta\tilde{V}_{k,i}^d\right)\right\}\\
\Delta\tilde{V}_{k,i}^d &= V\left(\tilde{X}_{k,i}^d\right) - V\left(\tilde{X}_{k,i}^d + \ell d^{-1/2} Z_{k+1,i}^d\right).
\end{aligned}$$

Define, for all $k_1+1\leq k\leq k_2$, $1\leq i,j\leq d$,

$$\begin{aligned}
\tilde{\mathcal{D}}_{\mathcal{I},j}^{d,k} &= \left\{\tilde{X}_{k,j}^d + \mathrm{r}\ell d^{-1/2} Z_{k+1,j}^d \in \mathcal{I}\right\}\cap\left\{\tilde{X}_{k,j}^d + (1-\mathrm{r})\ell d^{-1/2} Z_{k+1,j}^d \in \mathcal{I}\right\},\\
\tilde{\mathcal{D}}_{\mathcal{I},i:j}^{d,k} &= \bigcap_{\ell=i}^j \tilde{\mathcal{D}}_{\mathcal{I},\ell}^{d,k}.
\end{aligned}$$

Note that by convention $V(x)=-\infty$ for all $x\notin\mathcal{I}$, $\tilde{\mathcal{A}}_{k+1}^d\subset\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,k}$ so that $\left(\tilde{\mathcal{A}}_{k+1}^d\right)^c$ may be written $\left(\tilde{\mathcal{A}}_{k+1}^d\right)^c = \left(\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,k}\right)^c \cup\left(\left(\tilde{\mathcal{A}}_{k+1}^d\right)^c\cap\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,k}\right)$. Let $\mathcal{F}$ be the σ-field generated by $\left(\tilde{X}_k^d\right)_{k\geq 0}$. Consider the case $\#\{m_1,\dots,m_4\}=4$. The case $\#\{m_1,\dots,m_4\}=3$ is dealt with similarly and the two other cases follow the same lines as the proof of Lemma 7.26. As $\left\{\left(U_{m_j}, Z_{m_j,1}^d,\cdots,Z_{m_j,d}^d\right)\right\}_{1\leq j\leq 4}$ are independent conditionally to $\mathcal{F}$,

$$\mathbb{E}\left[\prod_{j=1}^4 Z_{m_j,1}^d \mathbb{1}_{\left(\mathcal{A}_{m_j}^d\right)^c}\middle|\mathcal{F}\right] = \prod_{j=1}^4\left\{\mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}\right)^c} Z_{m_j,1}^d\middle|\mathcal{F}\right] + \mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}}\mathbb{1}_{\left(\tilde{\mathcal{A}}_{m_j}^d\right)^c} Z_{m_j,1}^d\middle|\mathcal{F}\right]\right\}.$$

As U_{m_j} is independent of $(Z_{m_j,1}^d,\cdots,Z_{m_j,d}^d)$ conditionally to $\mathcal{F}$, the second term may be written:

$$\mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}}\mathbb{1}_{\left(\tilde{\mathcal{A}}_{m_j}^d\right)^c} Z_{m_j,1}^d\middle|\mathcal{F}\right] = \mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}} Z_{m_j,1}^d \left(1-\exp\left\{\sum_{i=1}^d \Delta\tilde{V}_{m_j-1,i}^d\right\}\right)_+\middle|\mathcal{F}\right].$$

Since the function $x\mapsto(1-\mathrm{e}^x)_+$ is 1-Lipschitz, on $\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}$

$$\left|\left(1-\exp\left\{\sum_{i=1}^d\Delta\tilde{V}_{m_j-1,i}^d\right\}\right)_+ - \Theta_{m_j}\right| \leq \left|\Delta\tilde{V}_{m_j-1,1}^d + \ell d^{-1/2}\dot{V}(\tilde{X}_{m_j-1,1}^d) Z_{m_j,1}^d\right|,$$

where $\Theta_{m_j} = (1-\exp\{-\ell d^{-1/2}\dot{V}(\tilde{X}_{m_j-1,1}^d)Z_{m_j,1}^d + \sum_{i=2}^d \Delta\tilde{V}_{m_j-1,i}^d\})_+$. Then,

$$\left|\mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}_{\mathcal{I},1:d}^{d,m_j-1}} Z_{m_j,1}^d\left(1-\exp\left\{\sum_{i=1}^d\Delta\tilde{V}_{m_j-1,i}^d\right\}\right)_+\middle|\mathcal{F}\right]\right| \leq A_{m_j}^d + B_{m_j}^d,$$

where

$$A^d_{m_j} = \mathbb{E}\left[\left|Z^d_{m_j,1}\right|\left|\mathbb{1}_{\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},1}}\Delta\tilde{V}^d_{m_j-1,1} + \ell d^{-1/2}\dot{V}(\tilde{X}^d_{m_j-1,1})Z^d_{m_j,1}\right|\middle|\mathcal{F}\right],$$

$$B^d_{m_j} = \left|\mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},2:d}}Z^d_{m_j,1}\Theta_{m_j}\middle|\mathcal{F}\right]\right|.$$

By Jensen inequality,

$$\left|\mathbb{E}\left[\prod_{j=1}^4 Z^d_{m_j,1}\mathbb{1}_{\left(\mathcal{A}^d_{m_j}\right)^c}\right]\right| \le \mathbb{E}\left[\prod_{j=1}^4\left\{\mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},1:d}\right)^c}|Z^d_{m_j,1}|\middle|\mathcal{F}\right] + A^d_{m_j} + B^d_{m_j}\right\}\right],$$
$$\le C\mathbb{E}\left[\sum_{j=1}^4\mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},1:d}\right)^c}|Z^d_{m_j,1}|^4\middle|\mathcal{F}\right] + \left(A^d_{m_j}\right)^4 + \left(B^d_{m_j}\right)^4\right],$$

By **G**4(iii) and Holder's inequality applied with $\alpha = 1/(1-2/\gamma) > 1$, for all $1 \le j \le 4$,

$$\mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},1:d}\right)^c}|Z^d_{m_j,1}|^4\right] \le \mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},1}\right)^c}|Z^d_{m_j,1}|^4\right] + \sum_{i=2}^d\mathbb{E}\left[\mathbb{1}_{\left(\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},i}\right)^c}\right],$$
$$\le \mathbb{E}\left[|Z^d_{m_j,1}|^{4\alpha/(\alpha-1)}\right]^{(\alpha-1)/\alpha}d^{-\gamma/(2\alpha)} + d^{1-\gamma/2},$$
$$\le Cd^{1-\gamma/2}.$$

By Lemma 7.22(ii) and the Holder's inequality, there exists $C > 0$ such that $\mathbb{E}\left[\left(A^d_{m_j}\right)^4\right] \le Cd^{-2}$. On the other hand, by [JLM15, Lemma 6] since $Z^d_{m_j,1}$ is independent of $\mathcal{F}$,

$$B^d_{m_j} = \left|\mathbb{E}\left[\mathbb{1}_{\tilde{\mathcal{D}}^{d,m_j-1}_{\mathcal{I},2:d}}\ell d^{-1/2}\dot{V}(\tilde{X}^d_{m_j-1,1})\mathcal{G}\left(\ell^2 d^{-1}\dot{V}(\tilde{X}^d_{m_j-1,1})^2, -2\sum_{i=2}^d\Delta\tilde{V}^d_{m_j-1,i}\right)\middle|\mathcal{F}\right]\right|,$$

where the function $\mathcal{G}$ is defined in (7.28). By **G**4(ii) and since $\mathcal{G}$ is bounded, $\mathbb{E}[(B^d_{m_j})^4] \le Cd^{-2}$. Since $\gamma \ge 6$ in **G**4(iii), $|\mathbb{E}[\prod_{j=1}^4 Z^d_{m_j,1}\mathbb{1}_{(\mathcal{A}^d_{m_j})^c}]| \le Cd^{-2}$, showing that

$$\sum_{(m_1,m_2,m_3,m_4)\in\mathcal{I}_4}\left|\mathbb{E}\left[\prod_{i=1}^4 Z^d_{m_i,1}\mathbb{1}_{(\mathcal{A}^d_{m_i})^c}\right]\right| \le Cd^{-2}\binom{k_2-k_1}{4}.$$

□

7.5.3 Proof of Proposition 7.11

Lemma 7.27. *Assume that* ***G****4 holds. Let μ be a limit point of the sequence of laws $(\mu_d)_{d\ge1}$ of $\left\{(Y^d_{t,1})_{t\ge0},\ d\in\mathbb{N}^*\right\}$. Then for all $t \ge 0$, the pushforward measure of μ by W_t is π.*

Proof. The proof is the same as in Lemma 7.17 and is omitted. □

We preface the proof by a lemma which provides a condition to verify that any limit point μ of $(\mu_d)_{d\geq 1}$ is a solution to the local martingale problem associated with (7.9).

Lemma 7.28. *Assume **G**4. Let μ be a limit point of the sequence $(\mu_d)_{d\geq 1}$. If for all $\phi \in C_c^\infty(\mathcal{I},\mathbb{R})$, the process $(\phi(W_t)-\phi(W_0)-\int_0^t \mathrm{L}\phi(W_u)\mathrm{d}u)_{t\geq 0}$ is a martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$, then μ solves the local martingale problem associated with (7.9).*

Proof. As for all $t \geq 0$ and $d \geq 1$, $Y_{t,1}^d \in \mathcal{I}$, for all $d \geq 1$ $\mu^d(C(\mathbb{R}_+,\overline{\mathcal{I}})) = 1$. Since $C(\mathbb{R}_+,\overline{\mathcal{I}})$ is closed in $\mathbf{W}$, we have by the Portmanteau theorem, $\mu(C(\mathbb{R}_+,\overline{\mathcal{I}})) = 1$. Therefore, we only need to prove that for all $\psi \in C^\infty(\overline{\mathcal{I}},\mathbb{R})$, the process $(\psi(W_t) - \psi(W_0) - \int_0^t \mathrm{L}\psi(W_u)\mathrm{d}u)_{t\geq 0}$ is a local martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$. Let $\psi \in C^\infty(\overline{\mathcal{I}},\mathbb{R})$.

Suppose first that for all $\varpi \in C_c^\infty(\overline{\mathcal{I}},\mathbb{R})$, $(\varpi(W_t) - \varpi(W_t) - \int_0^t \mathrm{L}\varpi(W_u)\mathrm{d}u)_{t\geq 0}$ is a martingale. Then, consider the sequence of stopping time defined for $k \geq 1$ by $\tau_k = \inf\{t \geq 0 \mid |W_t| \geq k\}$ and a sequence $(\varpi_k)_{k\geq 0}$ in $C_c^\infty(\overline{\mathcal{I}},\mathbb{R})$ satisfying:

1. for all $k \geq 1$ and all $x \in \overline{\mathcal{I}} \cap [-k,k]$, $\varpi_k(x) = \psi(x)$,
2. $\lim_{k\to+\infty} \varpi_k = \psi$ in $C^\infty(\overline{\mathcal{I}},\mathbb{R})$.

Since for all $k \geq 1$,

$$\left(\psi(W_{t\wedge\tau_k}) - \psi(W_0) - \int_0^{t\wedge\tau_k} \mathrm{L}\psi(W_u)\mathrm{d}u\right)_{t\geq 0} = \left(\varpi_k(W_{t\wedge\tau_k}) - \varpi_k(W_0) - \int_0^{t\wedge\tau_k} \mathrm{L}\varpi_k(W_u)\mathrm{d}u\right)_{t\geq 0}$$

and the sequence $(\tau_k)_{k\geq 1}$ goes to $+\infty$ as k goes to $+\infty$ almost surely, it follows that $(\psi(W_t) - \psi(W_0) - \int_0^t \mathrm{L}\psi(W_u)\mathrm{d}u)_{t\geq 0}$ is a local martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$. It remains to show that for all $\varpi \in C_c^\infty(\overline{\mathcal{I}},\mathbb{R})$, $(\varpi(W_t) - \varpi(W_0) - \int_0^t \mathrm{L}\varpi(W_u)\mathrm{d}u)_{t\geq 0}$ is a martingale under the assumption of the proposition. We only need to prove that for all $\varpi \in C_c^\infty(\overline{\mathcal{I}},\mathbb{R})$, $0 \leq s \leq t$, $m \in \mathbb{N}^*$, $g : \mathbb{R}^m \to \mathbb{R}$ bounded and continuous, and $0 \leq t_1 \leq \cdots \leq t_m \leq s \leq t$:

$$\mathbb{E}^\mu\left[\left(\varpi(W_t) - \varpi(W_s) - \int_s^t \mathrm{L}\varpi(W_u)\,\mathrm{d}u\right) g(W_{t_1},\ldots,W_{t_m})\right] = 0\,. \tag{7.55}$$

Let $(\phi_k)_{k\geq 0}$ be a sequence of functions in $C_c^\infty(\mathcal{I},\mathbb{R})$ and converging to ϖ in $C_c^\infty(\overline{\mathcal{I}},\mathbb{R})$. First note that for all $u \in [s,t]$, μ-almost everywhere,

$$\lim_{k\to+\infty} \phi_k(W_u) = \varpi(W_u)\,. \tag{7.56}$$

By Lemma 7.27, for all $u \in [s,t]$ the pushforward measure of μ by W_u has density π with respect to the Lebesgue measure and μ-almost everywhere, $\lim_{k\to+\infty} \mathrm{L}\phi_k(W_u) =$

$\mathrm{L}\varpi(W_u)$. On the other hand, there exists $C \geq 0$ such that for all $k \geq 0$, $|\mathrm{L}\phi_k(W_u)| \leq C(1 + |\dot{V}(W_u)|)$. Then,

$$\begin{aligned}\mathbb{E}^{\mu}\left[\int_s^t \left(1 + |\dot{V}(W_u)|\right) \mathrm{d}u\right] &\leq (t-s) + \int_s^t \mathbb{E}^{\mu}\left[|\dot{V}(W_u)|\right] \mathrm{d}u \\ &\leq (t-s)\left(1 + \int_{\mathcal{I}} |\dot{V}(x)|\pi(x)\mathrm{d}x\right) .\end{aligned}$$

Therefore, μ-almost everywhere by **G** 4(ii) and the Lebesgue dominated convergence theorem, we get

$$\lim_{k\to+\infty} \int_s^t \mathrm{L}\phi_k(W_u)\mathrm{d}u = \int_s^t \mathrm{L}\varpi(W_u)\mathrm{d}u\,. \tag{7.57}$$

Therefore, (7.55) follows from (7.56) and (7.57), using again the Lebesgue dominated convergence theorem and **G**4(ii). ☐

Proof of Proposition 7.11. Let μ be a limit point of $(\mu_d)_{d\geq 1}$. By Lemma 7.28, we only need to prove that for all $\phi \in C_c^\infty(\mathcal{I}, \mathbb{R})$, the process $(\phi(W_t) - \phi(W_0) - \int_0^t \mathrm{L}\phi(W_u)\mathrm{d}u)_{t\geq 0}$ is a martingale with respect to μ and the filtration $(\mathscr{B}_t)_{t\geq 0}$. Then, the proof follows the same line as the proof of Proposition 7.6 and is omitted. ☐

7.5.4 Proof of Theorem 7.12

Lemma 7.29. *Assume* **G**4 *holds. Let X^d be distributed according to π^d and Z^d be a d-dimensional standard Gaussian random variable, independent of X^d. Then, $\lim_{d\to+\infty} \mathrm{E}^d = 0$, where*

$$\mathrm{E}^d = \mathbb{E}\left[\left|\dot{V}(X_1^d)\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}}\left\{\mathcal{G}\left(\ell^2\dot{V}(X_1^d)^2/d, 2\bar{Y}_d\right) - \mathcal{G}\left(\ell^2\dot{V}(X_1^d)^2/d, 2\bar{X}_d\right)\right\}\right|\right] ,$$

where $\bar{Y}_d = \sum_{i=2}^d \Delta V_i^d$, ΔV_i^d and $\mathcal{D}^d_{\mathcal{I},2:d}$ are given by (7.5) and (7.53) and $\bar{X}_d = \sum_{i=2}^d b^d_{\mathcal{I},i}$, $b^d_{\mathcal{I},i} = b^d_{\mathcal{I}}(X_i^d, Z_i^d)$ with $b^d_{\mathcal{I}}$ given by (7.54).

Proof. Set for all $d \geq 1$, $\bar{Y}_d = \sum_{i=2}^d \Delta V_i^d$ and $\bar{X}_d = \sum_{i=2}^d b^d_{\mathcal{I},i}$. By definition of $b^d_{\mathcal{I}}$ (7.54), $\bar{X}_d$ may be expressed as $\bar{X}_d = \sigma_d \bar{S}_d + \mu_d$, where

$$\begin{aligned}\mu_d &= 2(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},1}}\zeta^d(X_1^d, Z_1^d)\right] - \frac{\ell^2(d-1)}{4d}\mathbb{E}\left[\dot{V}(X_1^d)^2\right] ,\\ \sigma_d^2 &= \ell^2\mathbb{E}\left[\dot{V}(X_1^d)^2\right] + \frac{\ell^4}{16d}\mathbb{E}\left[\left(\dot{V}(X_1^d)^2 - \mathbb{E}\left[\dot{V}(X_1^d)^2\right]\right)^2\right] ,\\ \bar{S}_d &= (\sqrt{d}\sigma_d)^{-1}\sum_{i=2}^d \beta_i^d ,\\ \beta_i^d &= -\ell Z_i^d \dot{V}(X_i^d) - \frac{\ell^2}{4\sqrt{d}}\left(\dot{V}(X_i^d)^2 - \mathbb{E}\left[\dot{V}(X_i^d)^2\right]\right) .\end{aligned}$$

By **G**4(ii) the Berry-Essen Theorem [Pet95, Theorem 5.7] can be applied to $\bar{S}_d$. Then, there exists a universal constant C such that for all $d > 0$,

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\sqrt{\frac{d}{d-1}} \bar{S}_d \leq x \right) - \Phi(x) \right| \leq C/\sqrt{d}\,.$$

It follows, with $\tilde{\sigma}_d^2 = (d-1)\sigma_d^2/d$, that

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\bar{X}_d \leq x \right) - \Phi((x-\mu_d)/\tilde{\sigma}_d) \right| \leq C/\sqrt{d}\,.$$

By this result and (7.39), Lemma 7.18 can be applied to obtain a constant $C \geq 0$, independent of d, such that:

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| \mathcal{G}\left(\ell^2 \dot{V}(X_1^d)^2/d, 2\bar{Y}_d\right) - \mathcal{G}\left(\ell^2 \dot{V}(X_1^d)^2/d, 2\bar{X}_d\right) \right| X_1^d \right] \\
&\qquad \leq C \left(\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| \bar{X}_d - \bar{Y}_d \right| \right] + d^{-1/2} + \sqrt{2\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| \bar{X}_d - \bar{Y}_d \right| \right] (2\pi\tilde{\sigma}_d^2)^{-1/2}} \right. \\
&\qquad\qquad \left. + \sqrt{\ell |\dot{V}(X_1^d)| / (2\pi d^{1/2} \tilde{\sigma}_d^2)} \right) .
\end{aligned}$$

Using this result, we have

$$\begin{aligned}
\mathrm{E}^d \leq C \Big\{ & \ell^{1/2} \mathbb{E}\left[|\dot{V}(X_1^d)|^{3/2} \right] (2\pi d^{1/2} \tilde{\sigma}_d^2)^{-1/2} + \mathbb{E}\left[|\dot{V}(X_1^d)| \right] \\
& \times \left(\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| \bar{X}_d - \bar{Y}_d \right| \right] + d^{-1/2} + \sqrt{2\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} \left| \bar{X}_d - \bar{Y}_d \right| \right] (2\pi\tilde{\sigma}_d^2)^{-1/2}} \right) \Big\} .
\end{aligned} \tag{7.58}$$

By Lemma 7.23, $\mathbb{E}[\mathbb{1}_{\mathcal{D}^d_{\mathcal{I},2:d}} |\bar{X}_d - \bar{Y}_d|]$ goes to 0 as d goes to infinity, and by **G** 4(ii) $\lim_{d\to+\infty} \tilde{\sigma}_d^2 = \ell^2 \mathbb{E}\left[\dot{V}(X)^2 \right]$. Combining these results with (7.58), it follows that E^d goes to 0 when d goes to infinity. □

For all $n \geq 0$, define $\mathcal{F}_n^d = \sigma(\{X_k^d, k \leq n\})$ and for all $\phi \in C_c^\infty(\mathbb{R}, \mathbb{R})$,

$$\begin{aligned}
M_n^d(\phi) = \frac{\ell}{\sqrt{d}} \sum_{k=0}^{n-1} \phi'(X_{k,1}^d) \left\{ Z_{k+1,1}^d \mathbb{1}_{\mathcal{A}_{k+1}^d} - \mathbb{E}^{\mathcal{F}_k^d}\left[Z_{k+1,1}^d \mathbb{1}_{\mathcal{A}_{k+1}^d} \right] \right\} \\
+ \frac{\ell^2}{2d} \sum_{k=0}^{n-1} \phi''(X_{k,1}^d) \left\{ (Z_{k+1,1}^d)^2 \mathbb{1}_{\mathcal{A}_{k+1}^d} - \mathbb{E}^{\mathcal{F}_k^d}\left[(Z_{k+1,1}^d)^2 \mathbb{1}_{\mathcal{A}_{k+1}^d} \right] \right\} .
\end{aligned} \tag{7.59}$$

Proposition 7.30. *Assume* ***G****4 and* ***G****5 hold. Then, for all $s \leq t$ and all $\phi \in C_c^\infty(\mathbb{R}, \mathbb{R})$,*

$$\lim_{d\to+\infty} \mathbb{E}\left[\left| \phi(Y_{t,1}^d) - \phi(Y_{s,1}^d) - \int_s^t \mathrm{L}\phi(Y_{r,1}^d) \mathrm{d}r - \left(M^d_{\lceil dt \rceil}(\phi) - M^d_{\lceil ds \rceil}(\phi) \right) \right| \right] = 0\,.$$

Proof. Using the same decomposition as in the proof of Proposition 7.20, we only need to prove that for all $1 \leq i \leq 5$, $\lim_{d\to+\infty} \mathbb{E}[|T_i^d|] = 0$, where

$$
\begin{aligned}
T_1^d &= \int_s^t \phi'(X^d_{\lfloor dr\rfloor,1}) \left(\ell\sqrt{d}\, \mathbb{E}^{\mathcal{F}^d_{\lfloor dr\rfloor}}\left[Z^d_{\lceil dr\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\right] + \frac{h(\ell)}{2}\dot{V}(X^d_{\lfloor dr\rfloor,1})\right)\mathrm{d}r\,,\\
T_2^d &= \int_s^t \phi''(X^d_{\lfloor dr\rfloor,1}) \left(\frac{\ell^2}{2}\,\mathbb{E}\left[(Z^d_{\lceil dr\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dr\rceil}}\,\middle|\mathcal{F}^d_{\lfloor dr\rfloor}\right] - \frac{h(\ell)}{2}\right)\mathrm{d}r\,,\\
T_3^d &= \int_s^t \left(\mathrm{L}\phi(Y^d_{\lfloor dr\rfloor/d,1}) - \mathrm{L}\phi(Y^d_{r,1})\right)\mathrm{d}r\,,\\
T_4^d &= \frac{\ell(\lceil dt\rceil - dt)}{\sqrt{d}}\phi'(X^d_{\lfloor dt\rfloor,1})\left(Z^d_{\lceil dt\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}} - \mathbb{E}\left[Z^d_{\lceil dt\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}\,\middle|\mathcal{F}^d_{\lfloor dt\rfloor}\right]\right)\\
&\quad + \frac{\ell^2(\lceil dt\rceil - dt)}{2d}\phi''(X^d_{\lfloor dt\rfloor,1})\left((Z^d_{\lceil dt\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}} - \mathbb{E}\left[(Z^d_{\lceil dt\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil dt\rceil}}\,\middle|\mathcal{F}^d_{\lfloor dt\rfloor}\right]\right)\,,\\
T_5^d &= \frac{\ell(\lceil ds\rceil - ds)}{\sqrt{d}}\phi'(X^d_{\lfloor ds\rfloor,1})\left(Z^d_{\lceil ds\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}} - \mathbb{E}\left[Z^d_{\lceil ds\rceil,1}\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}\,\middle|\mathcal{F}^d_{\lfloor ds\rfloor}\right]\right)\\
&\quad + \frac{\ell^2(\lceil ds\rceil - ds)}{2d}\phi''(X^d_{\lfloor ds\rfloor,1})\left((Z^d_{\lceil ds\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}} - \mathbb{E}\left[(Z^d_{\lceil ds\rceil,1})^2\mathbb{1}_{\mathcal{A}^d_{\lceil ds\rceil}}\,\middle|\mathcal{F}^d_{\lfloor ds\rfloor}\right]\right)\,.
\end{aligned}
$$

First, as ϕ' and ϕ'' are bounded, $\mathbb{E}\left[\left|T_4^d\right| + \left|T_5^d\right|\right] \leq Cd^{-1/2}$. Denote for all $r \in [s,t]$ and $d \geq 1$,

$$
\begin{aligned}
\Delta V^d_{r,i} &= V\left(X^d_{\lfloor dr\rfloor,i}\right) - V\left(X^d_{\lfloor dr\rfloor,i} + \ell d^{-1/2} Z^d_{\lceil dr\rceil,i}\right)\\
\Xi^d_r &= 1 \wedge \exp\left\{-\ell Z^d_{\lceil dr\rceil,1}\dot{V}(X^d_{\lfloor dr\rfloor,1})/\sqrt{d} + \sum_{i=2}^d b^{d,\lfloor dr\rfloor}_{\mathcal{I},i}\right\}\,,
\end{aligned}
$$

where for all $k, i \geq 0$, $b^{d,k}_{\mathcal{I},i} = b^d_{\mathcal{I}}(X^d_{k,i}, Z^d_{k+1,i})$, and for all $x, z \in \mathbb{R}$, $b^d_{\mathcal{I}}(x,y)$ is given by (7.54). For all $k \geq 0$, $1 \leq i,j \leq d$, define

$$
\begin{aligned}
\mathcal{D}^{d,k}_{\mathcal{I},j} &= \left\{X^d_{k,j} + \mathrm{r}\ell d^{-1/2} Z^d_{k+1,j} \in \mathcal{I}\right\} \cap \left\{X^d_{k,j} + (1-\mathrm{r})\ell d^{-1/2} Z^d_{k+1,j} \in \mathcal{I}\right\}\\
\mathcal{D}^{d,k}_{\mathcal{I},i:j} &= \bigcap_{\ell=i}^{j} \mathcal{D}^{d,k}_{\mathcal{I},\ell}\,.
\end{aligned}
$$

By the triangle inequality,

$$
|T_1| \leq \int_s^t \left|\phi'(X^d_{\lfloor dr\rfloor,1})\right|(A_{1,r} + A_{2,r} + A_{3,r} + A_{4,r})\mathrm{d}r\,, \tag{7.60}
$$

where

$$
\begin{aligned}
\Pi_r^d &= 1 \wedge \exp\left\{-\ell d^{-1/2} Z^d_{\lceil dr \rceil,1} \dot{V}(X^d_{\lfloor dr \rfloor,1}) + \sum_{i=2}^{d} \Delta V^d_{r,i}\right\} , \\
A_{1,r} &= \left|\ell\sqrt{d}\mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\left(\mathbb{1}_{\mathcal{A}^d_{\lceil dr \rceil}} - \mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1:d}}\Pi_r^d\right)\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right]\right| , \\
A_{2,r} &= \left|\ell\sqrt{d}\ \mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1:d}}\left(\Pi_r^d - \Xi_r^d\right)\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right]\right| , \\
A_{3,r} &= \left|\ell\sqrt{d}\ \mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\mathbb{1}_{\left(\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1:d}\right)^c}\Xi_r^d\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right]\right| , \\
A_{4,r} &= \left|\ell\sqrt{d}\ \mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\Xi_r^d\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right] + \dot{V}(X^d_{\lfloor dr \rfloor,1})h(\ell)/2\right| .
\end{aligned}
$$

Since $t \mapsto 1 \wedge \exp(t)$ is 1-Lipschitz,

$$
\begin{aligned}
\mathbb{E}\left[\left|A^d_{1,r}\right|\right] &\leq \ell\sqrt{d}\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1:d}}\left|Z^d_{\lceil dr \rceil,1}\right|\left|\Delta V^d_{r,1} - \ell d^{-1/2} Z^d_{\lceil dr \rceil,1}\dot{V}(X^d_{\lfloor dr \rfloor,1})\right|\right] , \\
&\leq \ell\sqrt{d}\mathbb{E}\left[\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1}}\left|Z^d_{\lceil dr \rceil,1}\right|\left|\Delta V^d_{r,1} - \ell d^{-1/2} Z^d_{\lceil dr \rceil,1}\dot{V}(X^d_{\lfloor dr \rfloor,1})\right|\right] , \\
&\leq \ell\sqrt{d}\mathbb{E}\left[\left|Z^d_{\lceil dr \rceil,1}\right|\left|\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1}}\Delta V^d_{r,1} - \ell d^{-1/2} Z^d_{\lceil dr \rceil,1}\dot{V}(X^d_{\lfloor dr \rfloor,1})\right|\right]
\end{aligned}
$$

and $\mathbb{E}[\left|A^d_{1,r}\right|]$ goes to 0 as $d \to +\infty$ for almost all r by Lemma 7.22(ii). So by the Fubini theorem, the first term in (7.60) goes to 0 as $d \to +\infty$. For $A^d_{2,r}$, note that

$$
A_{2,r} \leq \left|\ell\sqrt{d}\ \mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},2:d}}\left(\Pi_r^d - \Xi_r^d\right)\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right]\right| .
$$

Then, by [JLM15, Lemma 6],

$$
\begin{aligned}
\mathbb{E}\left[\left|A^d_{2,r}\right|\right] \leq \mathbb{E}\Bigg[\Bigg|\ell^2\dot{V}(X^d_{\lfloor dr \rfloor,1})\mathbb{1}_{\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},2:d}}\Bigg\{&\mathcal{G}\left(\frac{\ell^2\dot{V}(X^d_{\lfloor dr \rfloor,1})^2}{d}, 2\sum_{i=2}^{d}\Delta V^d_{r,i}\right) \\
&-\mathcal{G}\left(\frac{\ell^2\dot{V}(X^d_{\lfloor dr \rfloor,1})^2}{d}, 2\sum_{i=2}^{d} b^{d,\lfloor dr \rfloor}_{\mathcal{I},i}\right)\Bigg\}\Bigg|\Bigg] ,
\end{aligned}
$$

where $\mathcal{G}$ is defined in (7.28). By Lemma 7.29, this expectation goes to zero when d goes to infinity. Then by the Fubini theorem and the Lebesgue dominated convergence theorem, the second term of (7.60) goes 0 as $d \to +\infty$. On the other hand, by **G**4(iii) and Holder's inequality applied with $\alpha = 1/(1-2/\gamma) > 1$, for all $1 \leq j \leq 4$,

$$
\begin{aligned}
\mathbb{E}\left[\left|A^d_{3,r}\right|\right] &\leq \ell\sqrt{d}\left(\mathbb{E}\left[\left|Z^d_{\lceil dr \rceil,1}\right|\mathbb{1}_{\left(\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},1}\right)^c}\right] + \sum_{i=2}^{d}\mathbb{E}\left[\mathbb{1}_{\left(\mathcal{D}^{d,\lfloor dr \rfloor}_{\mathcal{I},i}\right)^c}\right]\right) , \\
&\leq \ell\sqrt{d}\left(\mathbb{E}\left[|Z^d_{m_j,1}|^{\alpha/(\alpha-1)}\right]^{(\alpha-1)/\alpha} d^{-\gamma/(2\alpha)} + d^{1-\gamma/2}\right) \leq C d^{3/2-\gamma/2}
\end{aligned}
$$

and $\mathbb{E}[\left|A^d_{3,r}\right|]$ goes to 0 as $d \to +\infty$ for almost all r. Define

$$\bar{V}_{d,1} = \sum_{i=1}^{d} \dot{V}(X^d_{\lfloor dr \rfloor,i})^2 \quad \text{and} \quad \bar{V}_{d,2} = \bar{V}_{d,1} - \dot{V}(X^d_{\lfloor dr \rfloor,1})^2 .$$

For the last term, by [JLM15, Lemma 6]:

$$\ell\sqrt{d}\ \mathbb{E}\left[Z^d_{\lceil dr \rceil,1}\Xi^d_r \middle| \mathcal{F}^d_{\lfloor dr \rfloor}\right] = -\ell^2\dot{V}(X^d_{\lfloor dr \rfloor,1}) \times \mathcal{G}\left(\frac{\ell^2}{d}\bar{V}_{d,1}, \left\{\frac{\ell^2}{2d}\bar{V}_{d,2} - 4(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)\right]\right\}\right) , \quad (7.61)$$

where $\mathcal{D}_\mathcal{I} = \left\{X + \ell d^{-1/2}Z \in \mathcal{I}\right\}$, X is distributed according to π and Z is a standard Gaussian random variable independent of X. As $\mathcal{G}$ is continuous on $\mathbb{R}_+ \times \mathbb{R} \setminus \{0,0\}$ (see [JLM15, Lemma 2]), by **G**4(ii), Lemma 7.25 and the law of large numbers, almost surely,

$$\lim_{d\to+\infty} \ell^2\mathcal{G}\left(\ell^2\bar{V}_{d,1}/d, \ell^2\bar{V}_{d,2}/(2d) - 4(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)\right]\right) = \ell^2\mathcal{G}\left(\ell^2\mathbb{E}[\dot{V}(X)^2], \ell^2\mathbb{E}[\dot{V}(X)^2]\right) = h(\ell)/2 , \quad (7.62)$$

where $h(\ell)$ is defined in (7.10). Therefore by Fubini's Theorem, (7.61) and Lebesgue's dominated convergence theorem, the last term of (7.60) goes to 0 as d goes to infinity. The proof for T^d_2 follows the same lines. By the triangle inequality,

$$\left|T^d_2\right| \le \left|\int_s^t \phi''(X^d_{\lfloor dr \rfloor,1})(\ell^2/2)\ \mathbb{E}\left[(Z^d_{\lceil dr \rceil,1})^2\left(\mathbb{1}_{\mathcal{A}^d_{\lceil dr \rceil}} - \Xi^d_r\right)\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right]\mathrm{d}r\right| + \left|\int_s^t \phi''(X^d_{\lfloor dr \rfloor,1})\left((\ell^2/2)\ \mathbb{E}\left[(Z^d_{\lceil dr \rceil,1})^2\Xi^d_r\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right] - h(\ell)/2\right)\mathrm{d}r\right| . \quad (7.63)$$

By Fubini's Theorem, Lebesgue's dominated convergence theorem and Proposition 7.24, the expectation of the first term goes to zero when d goes to infinity. For the second term, by [JLM15, Lemma 6 (A.5)],

$$(\ell^2/2)\mathbb{E}\left[(Z^d_{\lceil dr \rceil,1})^2 1 \wedge \exp\left\{-\frac{\ell Z^d_{\lceil dr \rceil,1}}{\sqrt{d}}\dot{V}(X^d_{\lfloor dr \rfloor,1}) + \sum_{i=2}^{d} b^{d,\lfloor dr \rfloor}_{\mathcal{I},i}\right\}\middle|\mathcal{F}^d_{\lfloor dr \rfloor}\right] = (B_1 + B_2 - B_3)/2 , \quad (7.64)$$

where

$$
\begin{aligned}
B_1 &= \ell^2\Gamma\left(\ell^2\bar{V}_{d,1}/d, \ell^2\bar{V}_{d,2}/(2d) - 4(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)\right]\right), \\
B_2 &= \frac{\ell^4\dot{V}(X^d_{\lfloor dr\rfloor,1})^2}{d}\mathcal{G}\left(\ell^2\bar{V}_{d,1}/d, \ell^2\bar{V}_{d,2}/(2d) - 4(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)\right]\right), \\
B_3 &= \frac{\ell^4\dot{V}(X^d_{\lfloor dr\rfloor,1})^2}{d}\left(2\pi\ell^2\bar{V}_{d,1}/d\right)^{-1/2} \\
&\qquad\qquad \times \exp\left\{-\frac{\left[-2(d-1)\mathbb{E}[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)] + (\ell^2/(4d))\bar{V}_{d,2}\right]^2}{2\ell^2\bar{V}_{d,1}/d}\right\},
\end{aligned}
$$

where Γ is defined in (7.29). As Γ is continuous on $\mathbb{R}_+\times\mathbb{R}\setminus\{0,0\}$ (see [JLM15, Lemma 2]), by **G**4(ii), Lemma 7.25 and the law of large numbers, almost surely,

$$
\begin{aligned}
\lim_{d\to+\infty} \ell^2\Gamma\left(\ell^2\bar{V}_{d,1}/d, \left\{\ell^2\bar{V}_{d,2}/(2d) - 4(d-1)\mathbb{E}\left[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)\right]\right\}\right) \\
= \ell^2\Gamma\left(\ell^2\mathbb{E}[\dot{V}(X)^2], \ell^2\mathbb{E}[\dot{V}(X)^2]\right) = h(\ell). \quad (7.65)
\end{aligned}
$$

By Lemma 7.25, by **G**4(ii) and the law of large numbers, almost surely,

$$
\lim_{d\to+\infty} \exp\left\{-\frac{\left[-2(d-1)\mathbb{E}[\mathbb{1}_{\mathcal{D}_\mathcal{I}}\zeta^d(X,Z)] + (\ell^2/(4d))\bar{V}_{d,2}\right]^2}{2\ell^2\bar{V}_{d,1}/d}\right\} = \exp\left\{-\frac{\ell^2}{8}\mathbb{E}[\dot{V}(X)^2]\right\}.
$$

Then, as $\mathcal{G}$ is bounded on $\mathbb{R}_+\times\mathbb{R}$,

$$
\lim_{d\to+\infty}\mathbb{E}\left[\left|\int_s^t \phi''(X^d_{\lfloor dr\rfloor,1})\,(B_2 - B_3)\,\mathrm{d}r\right|\right] = 0. \quad (7.66)
$$

Therefore, by Fubini's Theorem, (7.64), (7.65), (7.66) and Lebesgue's dominated convergence theorem, the second term of (7.63) goes to 0 as d goes to infinity. The proof for T_3^d follows exactly the same lines as the proof of Proposition 7.20. □

Proof of Theorem 7.12. Using Proposition 7.10, Proposition 7.11 and Proposition 7.30, the proof follows the same lines as the proof of Theorem 7.7. □

Chapter 8

Fast Langevin based algorithm for MCMC in high dimensions

Alain Durmus[1], Gareth O. Roberts [2], Gilles Vilmart[3] and Konstantinos C. Zygalakis[4]

Abstract

We introduce new Gaussian proposals to improve the efficiency of the standard Hastings-Metropolis algorithm in Markov chain Monte Carlo (MCMC) methods, used for the sampling from a target distribution in large dimension d. The improved complexity is $\mathcal{O}(d^{1/5})$ compared to the complexity $\mathcal{O}(d^{1/3})$ of the standard approach. We prove an asymptotic diffusion limit theorem and show that the relative efficiency of the algorithm can be characterised by its overall acceptance rate (with asymptotical value 0.704), independently of the target distribution. Numerical experiments confirm our theoretical findings.

8.1 Introduction

Consider a probability measure π on $\mathbb{R}^d$ with density again denoted by π with respect to the Lebesgue measure. The Langevin diffusion $\{x_t,\ t \geq 0\}$ associated with π is the solution of the following stochastic differential equation:

$$\mathrm{d}x_t = \frac{1}{2}\Sigma\nabla \log \pi(x_t)\mathrm{d}t + \Sigma^{1/2}\mathrm{d}W_t\ , \tag{8.1}$$

[1]LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2]Dept of Statistics University of Warwick, Coventry, CV4 7AL, UK. Gareth.O.Roberts@warwick.ac.uk

[3]Université de Genève, Section de mathématiques, 2-4 rue du Lièvre, CP 64, 1211 Genève 4, Switzerland. Gilles.Vilmart@unige.ch

[4]School of Mathematics and Maxwell Institute of Mathematical Sciences, University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. K.Zygalakis@ed.ac.uk

where $\{W_t,\ t \geq 0\}$ is a standard d-dimensional Brownian motion, and Σ is a given positive definite self-adjoint matrix. Under appropriate assumptions [Kha80] on π, it can be shown that the dynamic generated by (8.1) is ergodic with unique invariant distribution π. This is a key property of (8.1) and taking advantage of it permits to sample from the invariant distribution π. In particular, if one could solve (8.1) analytically and then take time t to infinity then it would be possible to generate samples from π. However, there exists a limited number of cases [KP92] where such an analytical formula exists. A standard approach is to discretise (8.1) using a one step integrator. The drawback of this approach is that it introduces a bias, because in general π is not invariant with respect to the Markov chain defined by the discretization, [TT90; MST10; AVZ14]. In addition, the discretization might fail to be ergodic [RT96a], even though (8.1) is geometrically ergodic.

An alternative way of sampling from π, which does not face the bias issue introduced by discretizing (8.1), is by using the Metropolis-Hastings algorithm [Has70]. The idea is to construct a Markov chain $\{x_j,\ j \in \mathbb{N}\}$, where at each step $j \in \mathbb{N}$, given x_j, a new candidate y_{j+1} is generated from a proposal density $q(x_j, \cdot)$. This candidate is then accepted ($x_{j+1} = y_{j+1}$) with probability $\alpha(x_j, y_{j+1})$ given by

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) , \tag{8.2}$$

and rejected ($x_{j+1} = x_j$) otherwise. The resulting Markov chain $(x_j)_{j\in\mathbb{N}}$ is reversible with respect to π and under mild assumptions is ergodic [Liu08; RC10].

The simplest proposals are random walks for which q is the transition kernel associated with the proposal

$$y = x + \sqrt{h}\Sigma^{1/2}\xi , \tag{8.3}$$

where ξ is a standard Gaussian random variable in $\mathbb{R}^d$, and leads to the well known Random Walk Metropolis Algorithm (RMW). This proposal is very simple to implement, but it suffers from (relatively) high rejection rate, due to the fact that it does not use information about π to construct appropriate candidate moves.

Another family of proposals commonly used, is based on the Euler-Maruyama discretization of (8.1), for which q is the transition kernel associated with the proposal

$$y = x + (h/2)\Sigma\nabla \log \pi(x) + \sqrt{h}\Sigma^{1/2}\xi , \tag{8.4}$$

where ξ is again a standard Gaussian random variable in $\mathbb{R}^d$. This algorithm is also known as the Metropolis Adjusted Langevin Algorithm (MALA), and it is well-established that it has better convergence properties than the RWM algorithm in general. This method directs the proposed moves towards areas of high probability for the distribution π, using the gradient of $\log \pi$. There is now a growing literature on gradient-based MCMC algorithms, as exemplified through the two papers [GC11; Cot+13] and the references therein. We also mention here function space MCMC methods [Cot+13]. Assuming that the target measure has a density w.r.t. a Gaussian measure on a Hilbert space, these algorithms are defined in infinite dimension and avoid completely the dependence on the dimension d faced by standard MCMC algorithms.

A natural question is if one can improve on the behaviour of MALA by incorporating more information about the properties of π in their proposal. A first attempt would be to use as proposal a one-step integrator with high weak order for (8.1), as suggested in the discussion of [GC11]. Although this turns out to not be sufficient, we shall show that, by slightly modifying this approach and not focusing on the weak order itself, we are able to construct a new proposal with better convergence properties than MALA. We mention that an analogous proposal is presented independently in [FS15] in a different context to improve the strong order of convergence of MALA.

Thus our main contribution in this paper is the introduction and theoretical analysis of the fMALA algorithm (*fast* MALA), and its cousins which will be introduced in Section 8.3. These algorithms provide for the first time, implementable gradient-based MCMC algorithms which can achieve convergence in $\mathcal{O}(d^{1/5})$ iterations, thus improving on the $\mathcal{O}(d^{1/3})$ of MALA and many related methods. These results are demonstrated as a result of high-dimensional diffusion approximation results. As well as giving these order of magnitude results for high-dimensional problems, we shall also give stochastic stability results, specifically results about the geometric ergodicity of the algorithms we introduce under appropriate regularity conditions.

Whilst the algorithms we describe have clear practical relevance for MCMC use, it is important to recognise the limitations of this initial study of these methodologies, and we shall note and comment on two which are particularly important. In order to obtain the diffusion limit results we give, it is necessary to make strong assumptions about the structure of the sequence of target distributions as d increases. In our analysis we assume that the target distribution consists of d i.i.d. components as in the initial studies of both high-dimensional RWM and MALA algorithms [RGG97; RR98]. Those analyses were subsequently extended (see for example [RR01a]) and supported by considerable empirical evidence from applied MCMC use. We also expect that in the context of this paper, our conclusions should provide practical guidance for MCMC practitioners well beyond the cases where rigorous results can be demonstrated, and we provide an example to illustrate this in Section 8.5.

Secondly, our diffusion limit results depend on the initial distribution of the Markov chain being the target distribution π, clearly impractical in real MCMC contexts. The works [CRR05; JLM14] study the case of MCMC algorithms (specifically RWM and MALA algorithms) started away from stationarity. On the one hand, it turns out that MALA algorithms are less robust than RWM when starting at under-dispersed values in that scaling strategies. Indeed, optimising mixing in stationarity can be highly suboptimal in the transient phase, often with initial moves having exponentially small acceptance probabilities (in d). On the other hand, a slightly more conservative strategy for MALA still achieves $\mathcal{O}(d^{1/2})$ compared to $\mathcal{O}(d)$ for RWM. It is natural to expect the story for fMALA to be at least as involved as that for MALA, and we give some empirical evidence to support this in the simulations study of Section 8.5. Future work will underpin these investigations with theoretical results analogous to those of [CRR05; JLM14]. From a practical MCMC perspective however, it should be noted that strategies which mix MALA-transient optimal scaling with fMALA-stationary optimal scaling will perform in a robust manner, both in the transient and stationary phases. Two of these effective strategies are illustrated in Section 8.5.

The paper is organised as follows. In Section 8.2 we provide a heuristic for the choice of the parameter h used in the proposal as a function of the dimension d of the target and present three different proposals that have better complexity scaling properties than RWM and MALA. In Section 8.3, we present fMALA and its variants, and prove our main results for the introduced methods. Section 8.4 investigates the ergodic properties of the different proposals for a wide variety of target densities π. Finally, in Section 8.5 we present numerical results that illustrate our theoretical findings.

8.2 Preliminaries

In this section we discuss some key issues regarding the convergence of MCMC algorithms. In particular, in Section 8.2.1 we discuss some issues related to the computational complexity of MCMC methods in high dimensions, while in Section 8.2.2 we present a useful heuristic for understanding the optimal scaling of a given MCMC proposal, and based on this heuristic formally derive a new proposal with desirable scaling properties.

8.2.1 Computational Complexity

Here we discuss a heuristic approach for selecting the parameter h in all proposals mentioned above as the dimension of the space d goes to infinity. In particular, we choose h proportional to an inverse power of the dimension d such that

$$h \propto d^{-\gamma} . \tag{8.5}$$

This implies that the proposal y is now a function of: (i) the current state x; (ii) the parameter γ through the scaling above; and (iii) the random variable ξ which appears in all the considered proposals. Thus $y = y(x, \xi; \gamma)$. Ideally γ should be as small as possible so the chain makes large steps and samples are correlated as little as possible. At the same time, the acceptance probability should not degenerate to 0 as $d \to \infty$, also to prevent high correlation amongst samples. This naturally leads to the definition of a critical exponent γ_0 given by

$$\gamma_0 = \inf_{\gamma_c \geq 0} \left\{ \gamma_c : \liminf_{d\to\infty} \mathbb{E}\left[\alpha(x,y)\right] > 0 \ , \quad \forall \gamma \in [\gamma_c, \infty) \right\} . \tag{8.6}$$

The expectation here is with respect to x distributed according to π and y chosen from the proposal distribution. In other words, we take the largest possible value for h, as function of d, constrained by asking that the average acceptance probability is bounded away from zero, uniformly in d. The time-step restriction (8.5) can be interpreted as a kind of Courant-Friedrichs-Lewy restriction arising in the numerical time-integration of PDEs.

If h is of the form (8.5), with $\gamma \geq \gamma_0$, the acceptance probability does not degenerate, and the Markov chain arising from the Metropolis-Hastings method can be thought of as an approximation of the Langevin SDE (8.1). This Markov chain travels with time-steps h on the paths of this SDE, and therefore requires a minimal number of steps to reach timescales of $\mathcal{O}(1)$ given by

$$M(d) = d^{\gamma_0} \ . \tag{8.7}$$

If it takes $\mathcal{O}(1)$ for the limiting SDE to reach stationarity, then we obtain that $M(d)$ gives the computational complexity of the algorithm.[1]

If we now consider the case of a product measure where

$$\pi(x) = \pi_d(x) = Z_d \prod_{i=1}^{d} e^{g(x_i)} \ , \tag{8.8}$$

and Z_d is the normalizing constant, then it is well known [RGG97] that for the RWM it holds $\gamma_0 = 1$, while for MALA it holds $\gamma_0 = 1/3$ [RR98]. In the next subsection, we recall the main ideas that allows one to obtain these scalings (valid also for some non-product cases), and derive a new proposal which we will call the fast Metropolis Adjusted Langevin algorithm (fMALA) and which satisfies $\gamma_0 = 1/5$ in the product case, i.e. it has a better convergence scaling.

8.2.2 Formal derivation

Here we explain the main idea that is used for proving the scaling of a Gaussian[2] proposal in high dimensions. In particular, the proposal y is now of the form

$$y = \mu(x,h) + S(x,h)\xi \ , \tag{8.9}$$

where $\xi \sim \mathcal{N}(0, \mathrm{I}_d)$ is a standard d dimensional Gaussian random variable. Note that in the case of the RWM,

$$\mu(x,h) = x, \quad S(x,h) = \sqrt{h}\Sigma^{1/2} \ ,$$

while in the case of MALA

$$\mu(x,h) = x + (h/2)\Sigma\nabla \log \pi(x), \quad S(x,h) = \sqrt{h}\Sigma^{1/2} \ .$$

The acceptance probability can be written in the form

$$\alpha(x,y) = \min\{1, \exp(R_d(x,y))\} \ ,$$

for some function $R_d(x,y)$ which depends on the Gaussian proposal (8.9). Now using the fact that y is related to x according to (8.9), $R_d(x,x) = 0$, together with appropriate smoothness properties on the function $g(x)$, one can expand R_d in powers of $\sqrt{h}$ using a Taylor expansion

$$R_d(x,y) = \sum_{i=1}^{k}\sum_{j=1}^{d} h^{i/2} C_{ij}(x,\xi) + h^{(k+1)/2} L_{k+1}(x,h^*,\xi) \ . \tag{8.10}$$

[1] In this definition of the cost one does not take into account the cost of generating a proposal. This is discussed in Remark 8.2.

[2] We point out that Gaussianity here is not necessary but it greatly simplifies the calculations.

It turns out [BS09] that the scaling associated with each proposal relates directly with how many of the C_{ij} terms are zero in (8.10). This simplifies if we further assume that $\Sigma = \mathrm{I}_d$ in (8.1) and that π satisfies (8.8), because we get for all $i \in \{1, \cdots, k\}$, $j \in \{1, \cdots, j\}$, $C_{ij}(x, \xi) = C_i(x_j, \xi_j)$ and (8.10) can be written as

$$R_d(x, y) = \sum_{i=1}^{k} \sum_{j=1}^{d} \frac{\sqrt{h^i d}}{\sqrt{d}} C_i(x_j, \xi_j) + h^{(k+1)/2} L_{k+1}(x, h^*, \xi) \; . \tag{8.11}$$

We then see that if $C_i = 0$, for $i = 1, \cdots, m$, then this implies that $\gamma_0 = 1/(m+1)$. Indeed, this value of γ_0 yields $h^{m+1} d = 1$ and the leading order term in (8.10) becomes

$$\frac{1}{\sqrt{d}} \sum_{j=1}^{d} C_{m+1}(x_j, \xi_j) \; .$$

To understand the behaviour for large d, we typically assume conditions to ensure that the above term has an appropriate (weak) limit. It turns out that $m + 1$ is generally an odd integer for known proposals, and the above expression is frequently approximated by a central limit theorem. The second dominant term in (8.10) turns out to be $C_{2(m+1)}$, although to turn this into a rigorous proof one also needs to be able to control the appropriate number of higher order terms, from $m + 1$ to $2(m + 1)$, as well as the remainder term in the above Taylor expansion.

8.2.3 Classes of proposals with $\gamma_0 = 1/5$

We introduce new Gaussian proposals for which $\gamma_0 = 1/5$ in (8.7). We start by presenting the simplest method, and then give two variations of it, motivated by the desire to obtain robust and stable ergodic properties (geometric ergodicity). The underlying calculations that show $C_i = 0, i = 1, \ldots, m$ with $m = 4$ and $\gamma_0 = 1/5$ for these methods are contained in the supplementary materials in the form of a Mathematica file. Recall that $f(x) = \Sigma \nabla \log \pi(x)$. In the sequel, we denote by Df and $D^2 f$ the Jacobian ($d \times d$-matrix) and the Hessian ($d \times d^2$-matrix) of f respectively. Thus $(Df(x))_{i,j} = \frac{\partial f_i(x)}{\partial x_j}$ and

$$D^2 f(x) = [\mathbf{H}_1(x) \quad \cdots \quad \mathbf{H}_d(x)] \quad , \quad \text{where } \{\mathbf{H}_i(x)\}_{j,k} = \frac{\partial f_i(x)}{\partial x_k \partial x_j}.$$

Finally for all $x \in \mathbb{R}^d$, $\{\Sigma : D^2 f(x)\} \in \mathbb{R}^d$ is defined by for $i = 1, \ldots, d$:

$$\left\{\Sigma : D^2 f(x)\right\}_i = \mathrm{trace}\left(\Sigma^T \mathbf{H}_i(x)\right) \; .$$

Notice that for $\Sigma = \mathrm{I}_d$, the above quantity reduces to the Laplacian and we have $\{\Sigma : D^2 f(x)\}_i = \Delta f_i$.

Fast Metropolis-Adjusted Langevin Algorithm (fMALA)

We first give a natural proposal for which $\gamma_0 = 1/5$ based on the discussion of Section 8.2.2. We restrict the class of proposal defined by (8.9) by setting for all $x \in \mathbb{R}^d$ and $h > 0$,

$$\mu(x,h) = x + h\,\mu_1(x) + h^2\mu_2(x)\ , \qquad S(x,h) = h^{1/2}S_1(x) + h^{3/2}S_2(x)\ .$$

By a formal calculation (see the supplementary materials), explicit expressions for the functions μ_1, μ_2, S_1, S_2 have to be imposed for the four first term $C_i(x,\xi)$, $i \in \{1,2,3,4\}$, in (8.11) to be zero. This result implies the following definition for μ and S:

$$\boxed{\begin{aligned}&\mu^{\mathrm{fM}}(x,h) = x + \frac{h}{2}f(x) - \frac{h^2}{24}\Big(Df(x)f(x) + \{\Sigma : D^2f(x)\}\Big)\,, \\ &S^{\mathrm{fM}}(x,h) = \Big(h^{1/2}\,\mathrm{I}_d + (h^{3/2}/12)Df(x)\Big)\,\Sigma^{1/2}\ .\end{aligned}} \qquad \begin{aligned}(8.12\text{a})\\(8.12\text{b})\end{aligned}$$

We will refer to (8.9) when μ, S are given by (8.12) as the fast Unadjusted Langevin Algorithm (fULA) when viewed as a numerical method for (8.1) and as the fast Metropolis-Adjusted Langevin Algorithm (fMALA) when used as a proposal in the Metropolis-Hastings framework.

Remark 8.1. *It is interesting to note that compared with Unadjusted Langevin Algorithm (ULA), fULA has the same order of weak convergence one, if applied as a one-step integrator for (8.1). One could obtain a second order weak method by changing the constants in front of the higher order coefficients, but in fact the corresponding method would not have better scaling properties than MALA when used in the Metropolis-Hastings framework. This observation answers negatively in part one of the questions in the discussion of [GC11] about the potential use of higher order integrators for the Langevin equation within the Metropolis-Hastings framework.*

Remark 8.2. *The proposal given by equation (8.12) contains higher order derivatives of the vector field $f(x)$, resulting in higher computational cost than the standard MALA proposal. This additional cost might offset the benefits of the improved scaling, since the corresponding Jacobian and Hessian can be full matrices in general. However, there exist cases of interest[3] where due to the structure of the Jacobian and Hessian the computational cost of the fMALA proposal is of the same order with respect to the dimension d as for the MALA proposal. Furthermore, we note that one possible way to avoid derivatives is by using finite differences or Runge-Kutta type approximations of the proposal (8.12). This, however, is out of the scope of the present paper.*

Modified Ozaki-Metropolis algorithm (mOMA)

One of the problems related to the MALA proposal is that it fails to be geometrically ergodic for a wide range of targets π [RT96a]. This issue was addressed in [RS02] where

[3] We study one of those in Section 8.5.

a modification of MALA based on the Ozaki discretization [Oza92] of (8.1) was proposed and studied. In the same spirit as in [RS02] we propose here a modification of fMALA, defined by

$$\boxed{\begin{aligned} \mu^{\mathrm{mO}}(x,h) &= x + \mathscr{T}_1(Df(x),h,1)f(x) - (h^2/6)Df(x)f(x) \\ &\qquad - (h^2/24)\{\Sigma : D^2 f(x)\} \\ S^{\mathrm{mO}}(x,h) &= \Big(\mathscr{T}_1(Df(x),2h,1) - (h^2/3)Df(x)\Big)^{1/2}\,\Sigma^{1/2}\,. \end{aligned}} \tag{8.13a, 8.13b}$$

where

$$\mathscr{T}_1(\mathrm{M},h,a) = (a\mathrm{M})^{-1}(e^{(ah/2)\mathrm{M}} - \mathrm{I}_d) \tag{8.14}$$

for all[4] $\mathrm{M} \in \mathbb{R}^{d\times d}$, $h > 0$, $a \in \mathbb{R}$.

The Markov chain defined by (8.13) will be referred to as the modified unadjusted Ozaki algorithm (mUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the modified Ozaki Metropolis algorithm (mOMA). Note that $t \mapsto (e^{ht}-1)/t - (1/3)h^2 t$ is positive on $\mathbb{R}$ for all $h > 0$, therefore $\mathscr{T}_1(Df(x),2h,1) - (h^2/3)Df(x)$ is a semi-positive matrix and $S^{\mathrm{mO}}(x,h)$ is well defined for all $x \in \mathbb{R}^d$ and $h > 0$.

Remark 8.3. *In regions where $\|\nabla \log \pi(x)\|$ is much greater than $\|x\|$, we need in practice to take h very small (of order $\|x\| / \|\nabla \log \pi(x)\|$) for MALA to exit these regions. However such a choice of h depends on x and cannot be used directly. Such a value of h can therefore be hard to find theoretically as well as computationally. This issue can be tackled by multiplying $f = \nabla \log \pi(x)$ by $\mathscr{T}_1(Df(x),h,a)$ in (8.13a). Indeed under some mild conditions, in that case, we can obtain an algorithm with good mixing properties for all $h > 0$; see [RS02, Theorem 4.1]. mOMA faces similar problems due to the term $Df(x)f(x)$.*

Generalised Boosted Ozaki-Metropolis Algorithm (gbOMA)

Having discussed the possible limitations of mOMA in Remark 8.3 we generalise here the approach in [RS02] to deal with the complexities arising to the presence of the $Df(x)f(x)$ term. In particular we now define

$$\boxed{\begin{aligned} \mu^{\mathrm{gbO}}(x,h) &= x + \mathscr{T}_1(Df(x),h,a_1)f(x) \\ &\qquad - (1/3)\mathscr{T}_3(Df(x),h,a_3)\{\Sigma : D^2 f(x)\} \\ &\qquad + ((a_1/2) + (1/6))\,\mathscr{T}_2(Df(x),h,a_2)f(x)\,, \\ S^{\mathrm{gbO}}(x,h) &= (\mathscr{T}_1(Df(x),2h,a_4) \\ &\qquad + ((a_4/2) - (1/6))\,\mathscr{T}_2(Df(x),2h,a_5))^{1/2}\,\Sigma^{1/2}\,. \end{aligned}} \tag{8.15a, 8.15b}$$

[4]Notice that the matrix functionals in (8.14),(8.16),(8.17) remain valid if matrix aM is not invertible, using the appropriate power series for the matrix exponentials.

where a_i, $i = 1, \cdots, 5$ are positive parameters, $\mathscr{T}_1$ is given by (8.14) and

$$\mathscr{T}_2(\mathrm{M}, h, a) = (a\mathrm{M})^{-1}(e^{-(ah^2/4)\mathrm{M}^2} - \mathrm{I}_d) \tag{8.16}$$

$$\mathscr{T}_3(\mathrm{M}, h, a) = (a\mathrm{M})^{-2}(e^{(ah/2)\mathrm{M}} - \mathrm{I}_d - (ah/2)\mathrm{M}) \tag{8.17}$$

with $\mathrm{M} \in \mathbb{R}^{d\times d}$, $h > 0$, $a \in \mathbb{R}$ and I_d is the identity matrix. The Markov chain defined by (8.15) will be referred to as the generalised boosted unadjusted Ozaki algorithm (gbUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the generalised boosted Ozaki Metropolis algorithm (gbOMA).

Note that S^{gbO} in (8.15b) is not always well defined in general and so the following condition on the constants a_4, a_5 is imposed to guarantee it.

H13. *The function* $t \mapsto (e^{a_4 t} - 1)/(a_4 t) + (a_4/2 - (1/6))(e^{-a_5 t^2} - 1)/(a_5 t)$ *is positive on* $\mathbb{R}$.

However for $a_4 = a_5 = 1$, this assumption is satisfied, and choosing $a_i = 1$ for all $i = 1, \ldots, 5$, (8.15) leads to a well defined proposal, which will be referred to as the boosted Unadjusted Ozaki Algorithm (bUOA), whereas when it is used in a Hastings-Metropolis algorithm, it will be referred to as the boosted Ozaki Metropolis Algorithm (bOMA). We will see in Section 8.4 that bOMA has nicer ergodic properties than fMALA.

8.3 Main scaling results

In this section, we present the optimal scaling results for fMALA and gbOMA introduced in Section 8.2. We recall from the discussion in Section 8.2 that the parameter h depends on the dimension and is given as $h_d = \ell^2 d^{-1/5}$, with $\ell > 0$. Finally, we prove our results for the case of target distributions of the product form given by (8.8), we take $\Sigma = \mathrm{I}_d$, and make the following assumptions on g.

H14. *We assume*

(i) $g \in C^{10}(\mathbb{R})$ *and* g'' *is bounded on* $\mathbb{R}$.

(ii) The derivatives of g *up to order* 10 *have at most a polynomial growth, i.e. there exists constants* C, κ *such that*

$$|g^{(i)}(t)| \leq C(1 + |t|^\kappa), \qquad t \in \mathbb{R}, i = 1, \ldots, 10.$$

(iii) for all $k \in \mathbb{N}$,

$$\int_{\mathbb{R}} t^k e^{g(t)} \mathrm{d}t < +\infty\,.$$

8.3.1 Optimal scaling of fMALA

The Markov chain produced by fMALA, with target density π_d and started at stationarity, will be denoted by $\{X_k^{d,\mathrm{fM}}\ ,\ k \in \mathbb{N}\}$. Let q_d^{fM} be the transition density associated with the proposal of fMALA relatively to π_d. In a similar manner, we denote by α_d^{fM} the acceptance probability. Now we introduce the jump process based on $\{X_k^{d,\mathrm{fM}},\ k \in \mathbb{N}\}$, which allows us to compare this Markov chain to a continuous-time process. Let $\{J_t,\ t \in \mathbb{R}_+\}$ be a Poisson process with rate $d^{1/5}$, and let $\Gamma^{d,\mathrm{fM}} = \{\Gamma_t^{d,\mathrm{fM}},\ t \in \mathbb{R}_+\}$ be the d-dimensional jump process defined by $\Gamma_t^{d,\mathrm{fM}} = X_{J_t}^{d,\mathrm{fM}}$. We denote by

$$a_d^{\mathrm{fM}}(\ell) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \pi_d(x) q_d^{\mathrm{fM}}(x,y)\alpha_d^{\mathrm{fM}}(x,y)\mathrm{d}x\mathrm{d}y$$

the mean under π_d of the acceptance rate.

Theorem 8.4. *Assume Assumption 14. Then*

$$\lim_{d\to+\infty} a_d^{\mathrm{fM}}(\ell) = a^{\mathrm{fM}}(\ell)\ ,$$

where $a^{\mathrm{fM}}(\ell) = 2\Phi(-K^{\mathrm{fM}}\ell^5/2)$ *with* $\Phi(t) = (1/(2\pi))\int_{-\infty}^{t} e^{-s^2/2}\mathrm{d}s$ *and the expression of* K^{fM} *is given in Appendix C.*

Theorem 8.5. *Assume Assumption 14. Let* $\{Y_t^{d,\mathrm{fM}} = \Gamma_{t,1}^{d,\mathrm{fM}},\ t \in \mathbb{R}_+\}$ *be the process corresponding to the first component of* $\Gamma^{d,\mathrm{fM}}$*. Then,* $\{Y^{d,\mathrm{fM}},\ d \in \mathbb{N}^*\}$ *converges weakly (in the Skorokhod topology), as* $d \to \infty$*, to the solution* $\{Y_t^{\mathrm{fM}},\ t \in \mathbb{R}_+\}$ *of the Langevin equation defined by:*

$$\mathrm{d}Y_t^{\mathrm{fM}} = (\mathrm{h}^{\mathrm{fM}}(\ell))^{(1/2)}\mathrm{d}B_t + (1/2)\mathrm{h}^{\mathrm{fM}}(\ell)\nabla\log\pi_1(Y_t^{\mathrm{fM}})\mathrm{d}t\ , \tag{8.18}$$

where $\mathrm{h}^{\mathrm{fM}}(\ell) = 2\ell^2\Phi(-K^{\mathrm{fM}}\ell^5/2)$ *is the speed of the limiting diffusion. Furthermore,* $\mathrm{h}^{\mathrm{fM}}(\ell)$ *is maximised at the unique value of* ℓ *for which* $a^{\mathrm{fM}}(\ell) = 0.704343$.

Proof. The proof of these two theorems are in Appendix 8.6. □

Remark 8.6. *The above analysis shows that for fMALA, the optimal exponent defined in (8.6) is given by* $\gamma_0 = 1/5$ *as discussed in Section 8.2.2. Indeed, if* h_d *has the form* $\ell^2 d^{-1/5+\epsilon}$*, then an adaptation of the proof of Theorem 8.4 implies that for all* $\ell > 0$*, if* $\epsilon \in (0, 1/5)$*,* $\lim_{d\to+\infty} a^{\mathrm{fM}}(\ell) = 0$*. In contrast, if* $\epsilon < 0$ *then* $\lim_{d\to+\infty} a^{\mathrm{fM}}(\ell) = 1$.

8.3.2 Scaling results for gbOMA

As in the case of fMALA, we assume π_d is of the form (8.8) and we take $\Sigma = \mathrm{I}_d$, $h_d = \ell^2 d^{-1/5}$. The Metropolis-adjusted Markov chain based on gbOMA, with target density π_d and started at stationarity, is denoted by $\{X_k^{d,\mathrm{gbO}},\ k \in \mathbb{N}\}$. We will denote by q_d^{gbO} the transition density associated with the proposals defined by gbOMA with respect to π_d. In a similar manner, the acceptance probability relatively to π_d and gbOMA will be denoted by α_d^{gbO}. Let $\{J_t,\ t \in \mathbb{R}_+\}$ be a Poisson process with rate $d^{1/5}$, and let $\Gamma^{d,\text{gbO}} = \{\Gamma_t^{d,\text{gbO}},\ t \in \mathbb{R}_+\}$ be the d-dimensional jump process defined by $\Gamma_t^{d,\text{gbO}} = X_{J_t}^{d,\text{gbO}}$. Denote also by

$$a_d^{\text{gbO}}(\ell) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \pi_d(x) q_d^{\text{gbO}}(x,y)\alpha_d^{\text{gbO}}(x,y)\mathrm{d}x\mathrm{d}y$$

the mean under π_d of the acceptance rate of the algorithm.

Theorem 8.7. *Assume Assumptions 13 and 14. Then*

$$\lim_{d\to+\infty} a_d^{\text{gbO}}(\ell) = a^{\text{gbO}}(\ell)\ ,$$

where $a^{\text{gbO}}(\ell) = 2\Phi(-K^{\text{gbO}}\ell^5/2)$ *with* $\Phi(t) = (1/(2\pi))\int_{-\infty}^t e^{-s^2/2}\mathrm{d}s$ *and* K^{gbO} *are given in Appendix C.*

Theorem 8.8. *Assume Assumptions 13 and 14. Let* $\{G_t^{d,\text{gbO}} = \Gamma_{t,1}^{d,\text{gbO}},\ t \in \mathbb{R}_+\}$ *be the process corresponding to the first component of* $\Gamma^{d,\text{gbO}}$. *Then,* $\left\{G^{d,\text{gbO}},\ d \in \mathbb{N}^*\right\}$ *converges weakly (in the Skorokhod topology) to the solution* $\{G_t^{\text{gbO}},\ t \in \mathbb{R}_+\}$ *of the Langevin equation defined by:*

$$\mathrm{d}G_t^{\text{gbO}} = (\mathrm{h}^{\text{gbO}}(\ell))^{(1/2)}\mathrm{d}B_t + (1/2)\mathrm{h}^{\text{gbO}}(\ell)\nabla\log\pi_c(G_t^{\text{gbO}})\mathrm{d}t\ ,$$

where $\mathrm{h}^{\text{gbO}}(\ell) = 2\ell^2\Phi(-K^{\text{gbO}}\ell^5/2)$ *is the speed of the limiting diffusion. Furthermore,* $\mathrm{h}^{\text{gbO}}(\ell)$ *is maximised at the unique value of* ℓ *for which* $a^{\text{gbO}}(\ell) = 0.704343$.

Proof. Note that under Assumption 14-(i), at fixed $a > 0$, using the regularity properties of $(x,h) \mapsto \mathscr{T}_i(x,h,a)$ on $\mathbb{R}^2$ for $i = 1,\ldots,3$, there exists an open interval I, which contains 0, and $M_0 \geq 0$ such that for all $x \in \mathbb{R}$, $k = 1,\cdots,11$, and $i = 1,\cdots,3$

$$\left|\frac{\partial^k\left(\mathscr{T}_i(g''(x),h,a)\right)}{\partial h^k}\right| \leq M_0 \quad \forall h \in I\ .$$

Using in addition Assumption 13 there exists $m_0 > 0$ such that for all $h \in I$ and for all $x \in \mathbb{R}$,

$$\mathscr{T}_1(g''(x),2h,a_4) + ((a_4/2) - (1/6))\,\mathscr{T}_2(g''(x),2h,a_5) \geq m_0\ .$$

Using these two results, the proof of both theorems follows the same lines as Theorems 8.4 and 8.5, which can be found in Appendix 8.6. □

8.4 Geometric ergodicity results for high order Langevin schemes

Having established the scaling behaviour of the different proposals in the previous section, we now proceed with establishing geometric ergodicity results for our new Metropolis algorithms. Furthermore, for completeness, we study the behaviour of the corresponding unadjusted proposal. For simplicity, we will take in the following $\Sigma = I_d$ and we limit our study of gbOMA to the one of bOMA, which is given by:

$$\begin{aligned} y^{\mathrm{bO}} &= \mu^{\mathrm{bO}}(x,h) + S^{\mathrm{bO}}(x,h)\ \xi\ , \\ \mu^{\mathrm{bO}}(x,h) &= x + \mathscr{T}_1(Df(x),h,1)f(x) + (2/3)\mathscr{T}_2(Df(x),h,1)f(x) \\ &\quad - (1/3)\mathscr{T}_3(Df(x),h,1)\{\Sigma : D^2 f(x)\}\ , \\ S^{\mathrm{bO}}(x,h) &= \left(\mathscr{T}_1(Df(x),2h,1) + (1/3)\mathscr{T}_2(Df(x),2h,1)\right)^{1/2}\ , \end{aligned} \tag{8.19}$$

where $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{T}_3$ are respectively defined by (8.14), (8.16) and (8.17). First, let us begin with some definitions. For a signed measure ν on $\mathbb{R}^d$, we define the total variation norm of ν by

$$\|\nu\|_{\mathrm{TV}} = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A)|\ ,$$

where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra of $\mathbb{R}^d$. Let P be a Markov kernel with invariant measure π. For a given measurable function $V : \mathbb{R}^d \to [1,+\infty)$, we will say that P is V-geometrically ergodic if there exist $C \geq 0$ and $\rho \in [0,1)$ such that for all $x \in \mathbb{R}^d$ and $n \geq 0$

$$\|P^n(x,\cdot) - \pi\|_V \leq C\rho^n V(x)\ ,$$

where for ν a signed measure on $\mathbb{R}^d$, the V-norm $\|\cdot\|_V$ is defined by

$$\|\nu\| = \sup_{\{f\ ;\ |f| \leq V\}} \int_{\mathbb{R}^d} f(x)\nu(\mathrm{d}x)\ .$$

We refer the reader to [MT09] for the definitions of small sets, φ-irreducibility and transience. Let P be a Markov kernel on $\mathbb{R}^d$, Leb^d-irreducible, where Leb^d is the Lebesgue measure on $\mathbb{R}^d$, and aperiodic and $V : \mathbb{R}^d \to [1,+\infty)$ be a measurable function. In order to establish that P is V-geometric ergodicity, a sufficient and necessary condition is given by a geometrical drift (see [MT09, Theorem 15.0.1]), namely for some small set $\mathcal{C}$, there exist $\lambda < 1$ and $b < +\infty$ such that for all $x \in \mathbb{R}^d$:

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_{\mathcal{C}}(x)\ . \tag{8.20}$$

Note that the different considered proposals belong to the class of Gaussian Markov kernels. Namely, let Q be a Markov kernel on $\mathbb{R}^d$. We say that Q is a Gaussian Markov kernel if for all $x \in \mathbb{R}^d$, $Q(x,\cdot)$ is a Gaussian measure, with mean $\mu(x)$ and covariance matrix $S(x)S^T(x)$, where $x \mapsto \mu(x)$ and $x \mapsto S(x)$ are measurable functions from $\mathbb{R}^d$ to respectively $\mathbb{R}^d$ and $\mathcal{S}_+^*(\mathbb{R}^d)$, the set of symmetric positive definite matrices of dimension d. These two functions will be referred to as the mean value map and the the variance map respectively. The Markov kernel Q has transition density q given by:

$$q(x,y) = \frac{1}{(2\pi)^{d/2}\,|S(x)|} \exp\left(-(1/2)\left\langle S(x)^{-2}(y-\mu(x)), (y-\mu(x))\right\rangle\right)\ , \tag{8.21}$$

where for $\mathrm{M} \in \mathbb{R}^{d\times d}$, $|\mathrm{M}|$ denotes the determinant of M. Geometric ergodicity of Markov Chains with Gaussian Markov kernels and the corresponding Metropolis-Hastings algorithms was the subject of study of [RT96a; Han03]. But contrary to [Han03], we assume for simplicity the following assumption on the functions $\mu : \mathbb{R}^d \to \mathbb{R}^d$ and $S : \mathbb{R}^d \to \mathcal{S}_+^*(\mathbb{R}^d)$:

H15. *The functions $x \mapsto \mu(x)$ and $x \mapsto S(x)$ are continuous.*

Note that if π, a target probability measure on $\mathbb{R}^d$, is absolutely continuous with respect to the Lebesgue measure with density still denoted by π, the following assumption ensures that the various different proposals introduced in this paper satisfy Assumption 15:

H16. *The log-density g of π belongs to $C^3(\mathbb{R}^d)$.*

We proceed in Section 8.4.1 with presenting and extending where necessary the main results about geometric ergodicity of Metropolis-Hasting algorithms using Gaussian proposals. In Section 8.4.2, we then introduce two different potential classes on which we apply our result in Section 8.4.3. Finally in Section 8.4.4, for completeness, we make the same kind of study but for unadjusted Gaussian Markov kernels on $\mathbb{R}$.

8.4.1 Geometric ergodicity of Hastings-Metropolis algorithm based on Gaussian Markov kernel

We first present an extension of the result given in in [Han03] for geometric ergodicity of Metropolis-Hastings algorithms based on Gaussian proposal kernels. In particular, let Q be a Gaussian Markov kernel with mean value map and variance map satisfying Assumption 15. We use such proposal in a Metropolis algorithm with target density π satisfying Assumption 16. Then, the produced Markov kernel P is given by

$$P(x, \mathrm{d}y) = \alpha(x, y)q(x, y)\mathrm{d}y + \delta_x(\mathrm{d}y) \int_{\mathbb{R}^d} (1 - \alpha(x, y))q(x, y)\mathrm{d}y \; , \tag{8.22}$$

where q and α are resp. given by (8.21) and (8.2).

H17. *We assume* $\liminf_{\|x\| \to +\infty} \int_{\mathbb{R}^d} \alpha(x, y)q(x, y)\mathrm{d}y > 0$.

Note that this condition is necessary to obtain the geometric ergodicity of a Metropolis-Hastings algorithm by [RT96a, Theorem 5.1]. We shall follow a well-known technique in MCMC theory in demonstrating that Assumption 17 allows us to ensure that geometric ergodicity of the algorithm is inherited from that of the proposal Markov chain itself. Thus, in the following lemma we combine the conditions given by [Han03], which imply geometric ergodicity of Gaussian Markov kernels, with Assumption 17 to get geometric ergodicity of the resultant Metropolis-Hastings Markov kernels.

Lemma 8.9. *Assume Assumptions 15, 17, and there exists $\tau \in (0, 1)$ such that*

$$\limsup_{\|x\| \to +\infty} \|\mu(x)\| \,/\, \|x\| = \tau, \;\; \textit{and} \;\; \limsup_{\|x\| \to +\infty} \|S(x)\| \,/\, \|x\| = 0 \; . \tag{8.23}$$

Then, the Markov kernel P given by (8.22) are V-geometrically ergodic, where $V(x) = 1 + \|x\|^2$.

Proof. The proof is postponed to Appendix 8.7.1. □

We now provide some conditions which imply that P is not geometrically ergodic.

Theorem 8.10. *Assume Assumptions 15,16, that π is bounded and there exists $\epsilon > 0$ such that*

$$\liminf_{\|x\|\to+\infty} \left\| S(x)^{-1}\mu(x) \right\| \|x\|^{-1} > \epsilon^{-1} , \qquad \liminf_{\|x\|\to+\infty} \inf_{\|y\|=1} \|S(x)y\| \geq \epsilon , \tag{8.24}$$

and

$$\lim_{\|x\|\to+\infty} \log\left(|S(x)|\right) / \|x\|^2 = 0 . \tag{8.25}$$

Then, P is not geometrically ergodic.

Proof. The proof is postponed to Appendix 8.7.2. □

8.4.2 Exponential potentials

We illustrate our results on the following classes of density.

The one-dimensional class $\mathscr{E}(\beta, \gamma)$

Let π be a probability density on $\mathbb{R}$ with respect to the Lebesgue measure. We will say that $\pi \in \mathcal{E}(\beta, \gamma)$ if π is positive, belongs to $C^3(\mathbb{R})$ and there exist $R_\pi, \beta > 0$ such that for all $x \in \mathbb{R}$, $|x| \geq R_\pi$,

$$\pi(x) \propto e^{-\gamma|x|^\beta} .$$

Then for $|x| \geq R_\pi$, $\log(\pi(x))' = -\gamma\beta x |x|^{\beta-2}$, $\log(\pi(x))'' = -\gamma\beta(\beta-1)|x|^\beta / x^2$ and $\log(\pi(x))^{(3)} = -\gamma\beta(\beta-1)(\beta-2)|x|^\beta / x^3$.

The multidimensional exponential class $\mathscr{P}_m$

Let π be a probability density on $\mathbb{R}^d$ with respect to the Lebesgue measure. We will say that $\pi \in \mathscr{P}_m$ if it is positive, belongs to $C^3(\mathbb{R}^d)$ and there exists $R_\pi \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq R_\pi$,

$$\pi(x) \propto e^{-\mathrm{q}(x)} ,$$

where q is a function of the following form. There exists a homogeneous polynomial p of degree m and a three-times continuously differentiable function r on $\mathbb{R}^d$ satisfying

$$\left\| D^2(\nabla r)(x) \right\| \underset{\|x\|\to+\infty}{=} o(\|x\|^{m-3}) , \tag{8.26}$$

and for all $x \in \mathbb{R}^d$

$$\mathrm{q}(x) = \mathrm{p}(x) + \mathrm{r}(x) .$$

Recall that p is an homogeneous polynomial of degree m if for all $t \in \mathbb{R}$ and $x \in \mathbb{R}^d$, $\mathrm{p}(tx) = t^m \mathrm{p}(x)$. Finally we define $\mathscr{P}_m^+$, the set of density $\pi \in \mathscr{P}_m$ such that the Hessian

of p at x, $\nabla^2 \mathrm{p}(x)$ is positive definite for all $x \neq 0$.
When p is an homogeneous polynomial of degree m, it can be written as

$$\mathrm{p}(x) = \sum_{|\mathbf{k}|=m} a_{\mathbf{k}} x^{\mathbf{k}} ,$$

where $\mathbf{k} \in \mathbb{N}^d$, $|\mathbf{k}| = \sum_i k_i$ and $x^{\mathbf{k}} = x_1^{k_1} \cdots x_d^{k_d}$. Then denoting by $\vec{n}_x = x/\|x\|$, it is easy to see that the following relations holds for all $x \in \mathbb{R}^d$.

$$\mathrm{p}(x) = \|x\|^m \ \mathrm{p}(\vec{n}_x) \tag{8.27}$$

$$\nabla \mathrm{p}(x) = \|x\|^{m-1} \nabla \mathrm{p}(\vec{n}_x) \tag{8.28}$$

$$\nabla^2 \mathrm{p}(x) = \|x\|^{m-2} \nabla^2 \mathrm{p}(\vec{n}_x) \tag{8.29}$$

$$D^2(\nabla \mathrm{p})(x) = \|x\|^{m-3} D^2(\nabla \mathrm{p})(x) \tag{8.30}$$

$$\langle \nabla \mathrm{p}(x), x \rangle = m \ \mathrm{p}(x) \tag{8.31}$$

$$\nabla^2 \mathrm{p}(x) x = (m-1) \ \nabla \mathrm{p}(x) \tag{8.32}$$

$$\left\langle \nabla^2 \mathrm{p}(x) x, x \right\rangle = m(m-1) \ \mathrm{p}(x) \ . \tag{8.33}$$

From (8.29), it follows that $\nabla^2 \mathrm{p}(x)$ is definite positive for all $x \in \mathbb{R}^d \setminus 0$ if and only if $\nabla^2 \mathrm{p}(\vec{n})$ is positive definite for all $\vec{n}$, with $\|\vec{n}\| = 1$. Then, p belongs to $\mathscr{P}_m^+$ only if $m \geq 2$.

8.4.3 Geometric ergodicity of the proposals: the case of Metropolis-Hastings algorithms

In this section we study the behaviour of our proposals within the Metropolis-Hastings framework. We will split our investigations in two parts: in the first we study fMALA and mOMA; while in the second we have a more detailed look in the properties of bOMA not only for the class $\mathscr{E}(\beta, \gamma)$, but also for the polynomial class $\mathscr{P}_m^+$.

Geometric ergodicity of fMALA, mOMA for the class $\mathscr{E}(\beta, \gamma)$

In the case $\beta \in (0, 2)$, fMALA and mOMA have their mean map behaving like $x - \beta\gamma x |x|^{\beta-2}/2$ at infinity and their variance map bounded from above. This is exactly the behaviour that MALA [RT96a] has for the same values of β, thus one would expect them to behave in the same way. This is indeed the case and thus using the same reasoning as in the proof [RT96a, Theorem 4.3] we deduce that the two algorithms are not geometrically ergodic for $\beta \in (0, 1)$. Similarly, the proof in [RT96a, Theorem 4.1] can be used to show that the two algorithms are geometrically ergodic for $\beta \in [1, 2)$. Furthermore, for values of $\beta \geq 2$ we have the following cases

(a) For $\beta = 2$,

- fMALA is geometrically ergodic if $h\gamma(1 + h\gamma/6) \in (0, 2)$ by [RT96a, Theorem 4.1], and not geometrically ergodic if $h\gamma(1+h\gamma/6) > 2$ by Theorem 8.10, since μ^{fM} is equivalent at infinity to $(1 - h\gamma(1 + h\gamma/6))x$ and $S^{\mathrm{fM}}(x)$ is constant for $|x| \geq R_\pi$.

- Since μ^{mO} is equivalent at infinity to $(e^{-\gamma h} - 2(h\gamma)^2/3)x$, we observe that mOMA is geometrically ergodic if $h\gamma \in (0, 1.22)$ by [RT96a, Theorem 4.1], and not geometrically ergodic if $h\gamma > 1.23$ by Theorem [RT96a, Theorem 5.1].

(b) For $\beta > 2$, fMALA and mOMA are not geometrically ergodic by Theorem 8.10 since the mean value maps of their proposal kernels are equivalent at infinity to $-C_1 |x|^{2\beta-2} /x$, their variance map to $C_2 |x|^{\beta-2}$ for some constants $C_1, C_2 > 0$, and the variance maps are bounded from below.

Geometric ergodicity of bOMA

In this section, we give some conditions under which bOMA is geometrically ergodic and some examples of density which satisfy such conditions. For a matrix $\mathrm{M} \in \mathbb{R}^{d\times d}$, we denote $\lambda_{\min}(\mathrm{M}) = \min \mathrm{Sp}(\mathrm{M})$ and $\lambda_{\max}(\mathrm{M}) = \max \mathrm{Sp}(\mathrm{M})$, where $\mathrm{Sp}(\mathrm{M})$ is the spectrum of M. We can observe three different behaviours of the proposal given by (8.19) when x is large, which are implied by the behaviour of $\lambda_{\min}(Df(x))$ and $\lambda_{\max}(Df(x))$.

If $\liminf_{\|x\|\to+\infty} \lambda_{\min}(Df(x)) = 0$. Then, $g(x) = o(\|x\|^2)$ as $\|x\| \to \infty$, and y^{bO} tends to be as the MALA proposal at infinity, and we can show that bOMA is geometrically ergodic with the same conditions introduced in [RT96a] for this one.

Example 8.11. *By [RT96a, Theorem 4.1] bOMA is geometrically ergodic for $\pi \in \mathscr{E}(\gamma, \beta)$ with $\beta \in [1, 2)$.*

Now, we focus on the case where $\limsup_{\|x\|\to+\infty} \lambda_{\max}(Df(x)) < 0$. For instance, this condition holds for $\pi \in \mathscr{E}(\gamma, \beta)$ when $\beta \geq 2$. We give conditions similar to the one for geometric convergence of the Ozaki discretization, given in [Han03], to check conditions of Lemma 8.9. Although these conditions does not cover all the cases, they seem to apply to interesting ones. Here are our assumptions where we denote by $\mathbb{S}^d = \{x \in \mathbb{R}^d,\ \|x\| = 1\}$, the sphere in $\mathbb{R}^d$ and $\vec{n}_x = x/\|x\|$.

H18. *We assume:*

1. $\limsup_{\|x\|\to+\infty} \lambda_{\max}(Df(x)) < 0$;
2. $\lim_{\|x\|\to+\infty} Df(x)^{-2}\{\mathrm{I}_d : D^2 f(x)\} = 0$;
3. *$Df(x)^{-1}f(x)$ is asymptotically homogeneous to x when $\|x\| \to +\infty$, i.e. there exists a function $c : \mathbb{S}^d \to \mathbb{R}$ such that*

$$\lim_{\|x\|\to+\infty} \left\| \frac{Df(x)^{-1}f(x)}{\|x\|} - c(\vec{n}_x)\vec{n}_x \right\| = 0\ .$$

The condition 1 in Assumption 18 implies that for all $x \in \mathbb{R}^d$, $\lambda_{\max}(Df(x)) \leq M_f$, and garantees that $S^{\mathrm{bO}}(x, h)$ is bounded for all $x \in \mathbb{R}^d$.

Lemma 8.12. *Assume Assumptions 16 and 18. There exists $M_\Sigma \geq 0$ such that for all $x \in \mathbb{R}^d$ $\left\|S^{\mathrm{bO}}(x,h)\right\| \leq M_\Sigma$.*

Proof. Since $S^{\mathrm{bO}}(x,h)$ is symmetric for all $x \in \mathbb{R}^d$, and $t \mapsto (e^{ht}-1)/t + (1/3)(e^{-(ht)^2} - 1)/t$ is bounded on $(-\infty, M]$ for all $M \in \mathbb{R}$, we just need to show that there exists $M_f \geq 0$ such that for all x, $\lambda_{\max}(Df(x)) \leq M_f$. First, by Assumption 18-(1), there exists $R \geq 0$, such that for all $x, \|x\| \geq R$, $\mathrm{Sp}(Df(x)) \subset \mathbb{R}_-$. In addition by Assumption 16 $x \mapsto Df(x)$ is continuous, and there exists $M \geq 0$ such that for all $x, \|x\| \leq R$, $\|Df(x)\| \leq M$. ☐

Theorem 8.13. *Assume Assumptions 16, 17 and 18. If*

$$0 < \inf_{n \in \mathbb{S}^d} c(n) \leq \sup_{n \in \mathbb{S}^d} c(n) < 6/5 \;, \tag{8.34}$$

then bOMA is geometrically ergodic.

Proof. We check that the conditions of Lemma 8.9 hold. By Assumption 16 and (8.19), Assumption 15 holds, thus it remains to check (8.23). First, Lemma 8.12 implies that the second equality of (8.23) is satisfied, and we just need to prove the first equality. By [Han03, Lemma 3.4], it suffices to prove that

$$\limsup_{\|x\| \to +\infty} \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle < 0 \;, \tag{8.35}$$

where $\eta(x) = \mu^{\mathrm{bO}}(x,h) - x$. Since $\limsup_{\|x\| \to +\infty} \lambda_{\max}(Df(x)) < 0$ we can write $\mathscr{G}(x) = \mathscr{B}(x) Df(x)^{-1} f(x)$, where

$$\mathscr{B}(x) = (e^{(h/2)Df(x)} - \mathrm{I}_d) + (2/3)(e^{-(hDf(x)/2)^2} - \mathrm{I}_d) \;,$$

and $x \mapsto \mathscr{B}(x)$ is bounded on $\mathbb{R}^d$. Since $\mathscr{B}$ is bounded on $\mathbb{R}^d$, by Assumption 18-(2)-(3) and (8.34),

$$\lim_{\|x\| \to +\infty} \left| \left\langle \frac{\eta(x)}{\|x\|}, \frac{\eta(x)}{\|x\|} + 2\vec{n}_x \right\rangle - \|\mathscr{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2\left\langle \mathscr{B}(x)\vec{n}_x, \vec{n}_x \right\rangle c(\vec{n}_x) \right| = 0 \;. \tag{8.36}$$

In addition, if we denote the eigenvalues of $\mathscr{B}(x)$ by $\{\lambda_i(x),\ i = 1, \ldots, d\}$ and $\{e_i(x),\ i = 1, \ldots, d\}$ an orthonormal basis of eigenvectors, we have

$$\begin{aligned} &\|\mathscr{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2\left\langle \mathscr{B}(x)\vec{n}_x, \vec{n}_x \right\rangle c(\vec{n}_x) \\ &\qquad = \sum_{i=1}^d c(\vec{n}_x)\lambda_i(x) \left\langle e_i(x), \vec{n}_x \right\rangle^2 (c(\vec{n}_x)\lambda_i(x) + 2) \end{aligned} \tag{8.37}$$

Since $\limsup_{\|x\| \to +\infty} Df(x) < 0$, for all i and $\|x\|$ large enough, $\lambda_i(x) \in [-5/3, 0)$. Therefore using (8.34) we get from (8.37):

$$\|\mathscr{B}(x)\vec{n}_x\|^2 c(\vec{n}_x)^2 + 2\left\langle \mathscr{B}(x)\vec{n}_x, \vec{n}_x \right\rangle c(\vec{n}_x) < 0 \;.$$

The proof is concluded using this result in (8.36). ☐

Application to the convergence of bOMA for $\pi \in \mathscr{P}_m^+$

For the proof of the main result of this section, we need the following lemma.

Lemma 8.14 ([Han03, Proof of Theorem 4.10]). *Let $\pi \in \mathscr{P}_m^+$ for $m \geq 2$, then π satisfies Assumption 18-(3) with $c(\vec{n}) = 1/(m-1) \in (0, 6/5)$ for all $\vec{n} \in \mathbb{S}^d$.*

Proposition 8.15. *Let $\pi \in \mathscr{P}_m^+$ for $m \geq 2$, then bOMA is V-geometrically ergodic, with $V(x) = \|x\|^2 + 1$.*

Proof. Let us denote $\pi \propto \exp(-\mathrm{p}(x) - \mathrm{r}(x))$, with p and r satisfying the conditions from the definition in Section 8.4.2. We prove that if $\pi \in \mathscr{P}_m^+$, Theorem 8.13 can be applied. First, by definition of $\mathscr{P}_m^+$, Assumption 16 is satisfied. Furthermore, Assumption 18-(1)-(2) follows from (8.26), (8.29), (8.30) and the condition that $\nabla^2 \mathrm{p}(\vec{n})$ is positive definite for all $\vec{n} \in \mathbb{S}^d$. Also by Lemma 8.14, Assumption 18-(3) is satisfied.
Now we focus on Assumption 17. For ease of notation, in the following we denote μ^{bO} and S^{bO} by μ and S , and do not mention the dependence in the parameter h of μ and S when it does not play any role. Note that

$$\int_{\mathbb{R}^d} \alpha(x,y) q(x,y) \mathrm{d}y = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \{1 \wedge \exp \tilde{\alpha}(x,\xi)\} \exp(-\|\xi\|^2/2) \mathrm{d}\xi \;, \tag{8.38}$$

where

$$\begin{aligned} \tilde{\alpha}(x,\xi) = &-\mathrm{p}(\mu(x) + S(x)\xi) + \mathrm{p}(x) - \mathrm{r}(\mu(x) \\ &+ S(x)\xi) + \mathrm{r}(x) - \log(|S(\mu(x) + S(x)\xi)|) + \log(|S(x)|) + (1/2)\|\xi\|^2 \\ &- (1/2) \left\langle (\tilde{S}(x,\xi))^{-1} \{x - \mu(\mu(x) + S(x)\xi)\}, x - \mu\left(\mu(x) + S(x)\xi\right) \right\rangle \;, \end{aligned} \tag{8.39}$$

and $\tilde{S}(x,\xi) = S(\mu(x) + S(x)\xi) S(\mu(x) + S(x)\xi)^T$. First, we consider $m \geq 3$, then we have the following estimate of the terms in (8.39) by (8.26)-(8.30) and Lemma 8.14:

$$\mu(w) \underset{\|w\| \to +\infty}{=} \{1 - 5/(3(m-1))\} w + o(\|w\|) \tag{8.40}$$

$$(S(w)S(w)^T)^{-1} \underset{\|w\| \to +\infty}{=} \frac{3}{4} m(m-1) \|w\|^{m-2} \nabla^2 \mathrm{p}(\vec{n}_w) + o(\|w\|^{m-2}) \tag{8.41}$$

$$\log(|S(w)|) \underset{\|w\| \to +\infty}{=} o(\|w\|) \tag{8.42}$$

Then by (8.40)-(8.42), if we define $\Psi : [3, +\infty) \to \mathbb{R}$ by

$$\begin{aligned} m \mapsto 1 - &\left\{1 - \frac{5}{3(m-1)}\right\}^m \\ &- (3/8) m(m-1) \left\{1 - \left(1 - \frac{5}{3(m-1)}\right)^2\right\}^2 \left\{1 - \frac{5}{3(m-1)}\right\}^{m-2} \;, \end{aligned}$$

we get

$$\tilde{\alpha}(x,\xi) \underset{\|x\| \to +\infty}{=} \|x\|^m \mathrm{p}(\vec{n}_x) \Psi(m) + o(\|x\|^m) \;.$$

Since Ψ is positive on $[3, +\infty)$, for all $\xi \in \mathbb{R}^d$ $\lim_{\|x\| \to +\infty} \widetilde{\alpha}(x, \xi) = +\infty$. This result, (8.38) and Fatou's Lemma imply that Assumption 17 is satisfied.
For $m = 2$, we can assume $\mathrm{p}(x) = \langle \mathrm{A}x, x\rangle$ with $\mathrm{A} \in \mathcal{S}_+^*(\mathbb{R}^d)$. Let us denote for M an invertible matrix of dimension $p \geq 1$,

$$\varrho(\mathrm{M}) = (e^{-\mathrm{M}} - \mathrm{I}_p) + (2/3)(e^{-\mathrm{M}^2} - \mathrm{I}_p)$$
$$\varsigma(\mathrm{M}) = (e^{-2\mathrm{M}} - \mathrm{I}_p) + (1/3)(e^{-4\mathrm{M}^2} - \mathrm{I}_p) \; .$$

Then we have the following estimates:

$$\widetilde{\alpha}(x, \xi) \underset{\|x\| \to +\infty}{=} \left\langle \mathrm{A}(\varsigma(h\mathrm{A}))^{-1} \left\{ \left(2\varrho(h\mathrm{A}) + \varrho(h\mathrm{A})^2\right) x \right\}, \left(2\varrho(h\mathrm{A}) + \varrho(h\mathrm{A})^2\right) x \right\rangle$$
$$+ \langle \mathrm{A}x, x\rangle - \langle \mathrm{A}\left\{(\mathrm{I}_d + \varrho(h\mathrm{A}))x\right\}, (\mathrm{I}_d + \varrho(h\mathrm{A}))x\rangle + o(\|x\|^2) \tag{8.43}$$

If we denote the eigenvalues of A by $\{\lambda_i, i = 1 \dots d\}$ and $\{x_i, i = 1, \dots, d\}$ the coordinates of x in an orthonormal basis of eigenvectors for A, (8.43) becomes

$$\widetilde{\alpha}(x, \xi) \underset{\|x\| \to +\infty}{=} \sum_{i=1}^{d} \Xi(h, \lambda_i) x_i^2 + o(\|x\|^2) \; . \tag{8.44}$$

where for $h, \lambda > 0$,

$$\Xi(h, \lambda) = \lambda \left(1 - (\varrho(h\lambda) + 1)^2 + \varsigma(h\lambda)^{-1} \left(4\varrho(h\lambda)^2 + 4\varrho(h\lambda)^3 + \varrho(h\lambda)^4\right)\right) \; .$$

Using that for any $h, \lambda > 0$, $\Xi(h, \lambda) > 0$ and (8.44), we have for all $\xi \in \mathbb{R}^d$, $\lim_{\|x\| \to +\infty} \widetilde{\alpha}(x, \xi) = +\infty$, and as in the first case Assumption 17 is satisfied. □

Remark 8.16. *Using the same reasoning as in Proposition 8.15, one can show that bOMA is geometrically ergodic for $\pi \in \mathscr{E}(\beta, \gamma)$ with $\beta \geq 2$.*

We now summarise the behaviour for all the different algorithms for the one dimensional class $\mathscr{E}(\beta, \gamma)$ in Table 8.1

Method	$\beta \in [1, 2)$	$\beta = 2$	$\beta > 2$
fMALA (8.12)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
mOMA (8.13)	geometrically ergodic	geometrically ergodic or not	not geometrically ergodic
bOMA (8.19)	geometrically ergodic	geometrically ergodic	geometrically ergodic

Table 8.1: Summary of ergodicity results for the Metropolis-Hastings algorithms for the class $\mathscr{E}(\beta, \gamma)$

8.4.4 Convergence of Gaussian Markov kernel on $\mathbb{R}$

We now present precise results for the ergodicity of the unadjusted proposals, by extending the results of [RT96a] for the ULA to Gaussian Markov kernels on $\mathbb{R}$. Under Assumption 15, it is straightforward to see that Q is Leb^d-irreducible, where Leb^d is the Lebesgue measure, aperiodic and all compact set of $\mathbb{R}^d$ are small; see [Han03, Theorem 3.1]. We now state our main theorems, which essentially complete [RT96a, Theorem 3.1-3.2]. Since their proof are very similar, they are omitted.

Theorem 8.17. *Assume Assumption 15, and there exist $s_\wedge, u_+, u_- \in \mathbb{R}_+^*$ and $\chi \in \mathbb{R}$ such that:*

$$\limsup_{|x|\to+\infty} S(x) \leq s_\wedge \ ,$$

$$\lim_{x\to+\infty} \{\mu(x) - x\} \, x^{-\chi} = -u_+ \ , \ \text{and} \ \lim_{x\to-\infty} \{\mu(x) - x\} \, |x|^{-\chi} = u_- \ .$$

(i) If $\chi \in [0,1)$, then Q is geometrically ergodic.

(ii) If $\chi = 1$ and $(1-u_+)(1-u_-) < 1$, then Q is geometrically ergodic.

(iii) If $\chi \in (-1,0)$, then Q is ergodic but not geometrically ergodic.

Proof. See the proof of [RT96a, Theorem 3.1]. □

Theorem 8.18. *Assume Assumption 15, and there exist $s_\vee, u_+, u_- \in \mathbb{R}_+^*$ and $\chi \in \mathbb{R}$ such that:*

$$\liminf_{|x|\to+\infty} S(x) \geq s_\vee \ ,$$

$$\lim_{x\to+\infty} S(x)^{-1}\mu(x)x^{-\chi} = -u_+ \ , \ \text{and} \ \lim_{x\to-\infty} S(x)^{-1}\mu(x) \, |x|^{-\chi} = u_- \ .$$

(i) If $\chi > 1$, then Q is transient.

(ii) If $\chi = 1$ and $(u_+ \wedge u_-)s_\vee > 1$, then Q is transient.

Proof. See the proof of [RT96a, Theorem 3.2]. □

Ergodicity of the unadjusted proposals for the class $\mathscr{E}(\beta,\gamma)$

We now apply Theorems 8.17 and 8.18 in order to study the ergodicity of the different unadjusted proposals applied to $\pi \in \mathscr{E}(\beta,\gamma)$. In the case $\beta \in (0,2)$ all the three algorithms (fULA,mUOA,bUOA) have their mean map behaving like $x - \beta\gamma x \, |x|^{\beta-2}/2$ at infinity and their variance map bounded from above. This is exactly the behaviour that ULA [RT96a] has for the same values of β, thus it should not be a surprise that Theorem 8.17 implies that all the three algorithms behaved as the ULA does for the corresponding values, namely being ergodic for $\beta \in (0,1)$ and geometrically ergodic for $\beta \in [1,2)$. Furthermore, for values of $\beta \geq 2$ we have the following cases.

(a) For $\beta = 2$,

- fULA is geometrically ergodic if $h\gamma(1 + h\gamma/6) \in (0, 2)$ by Theorem 8.17-((ii)), and is transcient if $h\gamma(1 + h\gamma/6) > 2$ by Theorem 8.18-((ii)), since μ^{fM} is equivalent at infinity to $(1 - h\gamma(1 + h\gamma/6))x$ and $S^{\mathrm{fM}}(x)$ is constant for $|x| \geq R_\pi$.
- mUOA is geometrically ergodic if $1 + 2(h\gamma)^2/3 - e^{-\gamma h} \in (0, 2)$ by Theorem 8.17-((ii)), and is transcient if $1 + 2(h\gamma)^2/3 - e^{-\gamma h} > 2$ by Theorem 8.18-((ii)), since μ^{mO} is equivalent at infinity to $(e^{-\gamma h} - 2(h\gamma)^2/3)x$ and $S^{\mathrm{mO}}(x)$ is constant for $|x| \geq R_\pi$.
- bUOA is geometrically ergodic by Theorem 8.17-((ii)), since μ^{bO} is equivalent at infinity to $-2x/3$ and $S^{\mathrm{bO}}(x)$ is constant for $|x| \geq R_\pi$.

(b) For $\beta > 2$,

- fULA and mUOA are transcient by Theorem 8.18-((i)) since their mean value map is equivalent at infinity to $-C_1 |x|^{2\beta-2}/x$, and their variance map to $C_2 |x|^{\beta-2}$ for some constants $C_1, C_2 > 0$, and their variance map are bounded from below.
- bUOA is geometrically ergodic by Theorem 8.17-((i)) since its mean value map is equivalent at infinity to $\{1 - 5/(3(\beta - 1))\}\, x$ and its variance map is bounded from above.

The summary of our findings can be found in Table 8.2.

Method	$\beta \in (0,1)$	$\beta \in [1,2)$	$\beta = 2$	$\beta > 2$
fULA (8.12)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
mUOA (8.13)	ergodic	geometrically ergodic	geometrically ergodic/transient	transient
bUOA (8.19)	ergodic	geometrically ergodic	geometrically ergodic	geometrically ergodic

Table 8.2: Summary of ergodicity results for the unadjusted proposals for the class $\mathscr{E}(\beta, \gamma)$.

8.5 Numerical illustration of the improved efficiency

In this section, we illustrate our analysis (Section 8.3.1) of the asymptotic behaviour of fMALA as the dimension d tends to infinity, and we demonstrate its gain of efficiency

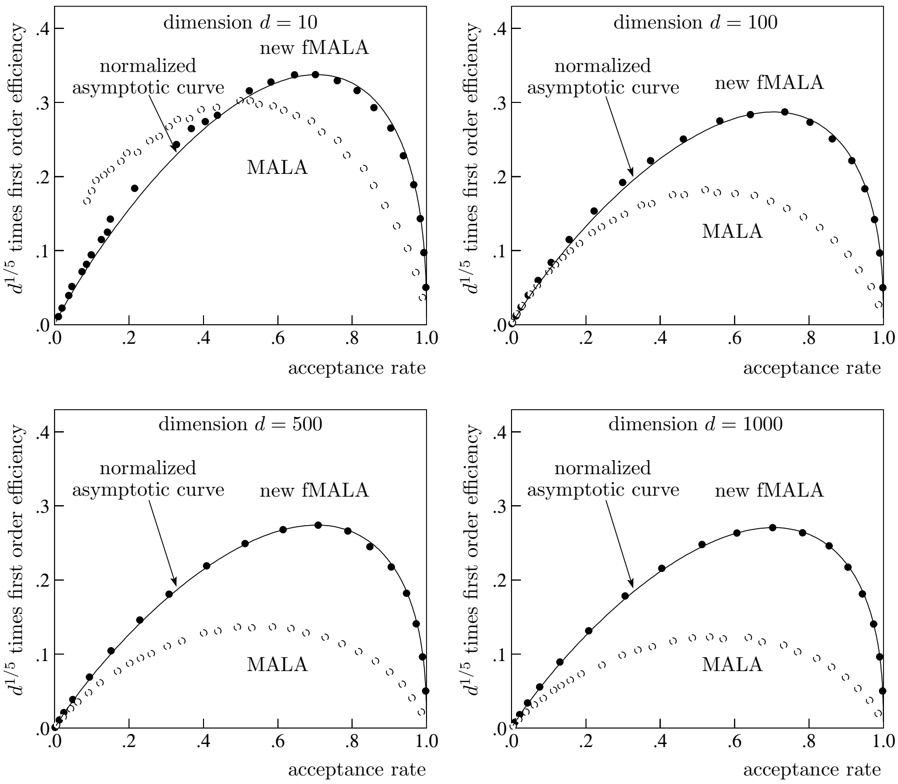


Figure 8.1: First-order efficiency of the new fMALA and the standard MALA for the double well potential $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$, as a function of the overall acceptance rates in dimensions $d = 10, 100, 500, 1000$. The solid line is the reference asymptotic curve of efficiency for the new fMALA, normalised to have the same maximum value as the finite dimensional fMALA.

as d increases compared to the standard MALA. Following [RR98], we define the first-order efficiency of a multidimensional Markov chain $\{X_k, k \in \mathbb{N}\}$ with first component denoted $X_k^{(1)}$ as $\mathbb{E}[(X_{k+1}^{(1)} - X_k^{(1)})^2]$. In Figure 8.1, we consider as a test problem the product case (8.8) using the double well potential with $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$ in dimensions $d = 10, 100, 500, 1000$, respectively. We consider many time stepsizes $h = \ell^2 d^{-1/5}$, plotting the first order efficiency (multiplied by $d^{1/5}$ because this is the scale which is asymptotically constant for fMALA as $d \to \infty$) as a function of the acceptance rate for the standard MALA (white bullets) and the acceptance rate $a_d^{\mathrm{fM}}(\ell)$ of the improved version fMALA (black bullets), respectively. For simplicity, each chain is started from the origin. The expectations are approximated as the average over 2×10^5 iterations of the algorithms and we use the same sets of generated random numbers for both methods. For comparison, we also include (as solid lines) the asymptotic efficiency curve of fMALA as d goes to infinity, normalised to have the same maximum as fMALA in finite dimension d. This corresponds to the (rescaled) limiting diffusion speed $\mathrm{h}^{\mathrm{fM}}(\ell)$ as a function of $a^{\mathrm{fM}}(\ell)$ (quantities given respectively in Theorems 8.4 and 8.5). We observe excellent agreement of the numerical first order efficiency compared to the asymptotic one, especially as d

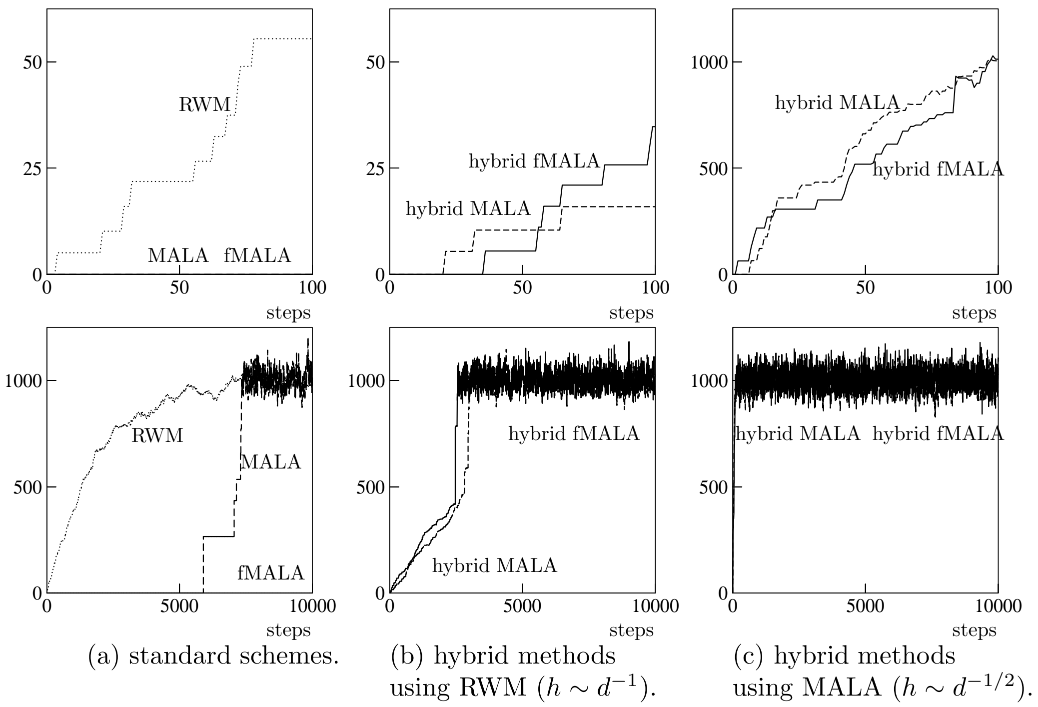
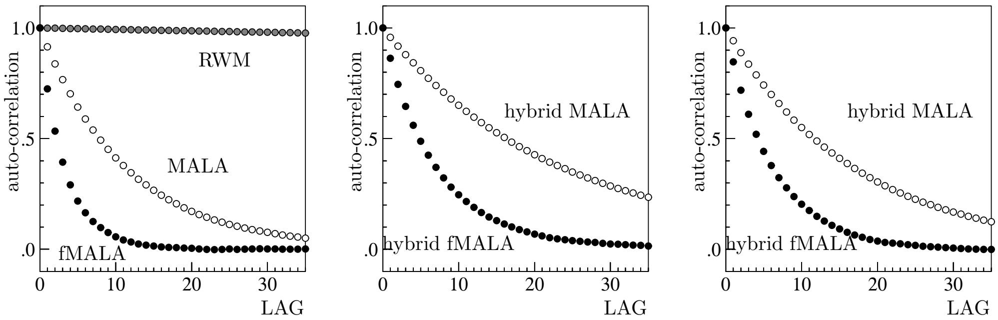


Figure 8.2: Trace plots of $\|X\|^2$ for the Gaussian target density in dimension $d = 1000$ when starting at the origin. Comparison of fMALA with $h \sim d^{-1/5}$ (solid lines), MALA with $h \sim d^{-1/3}$ (dashed lines), RWM with $h \sim d^{-1}$ (dotted lines).

increases, which corroborates the scaling results of fMALA. In addition, we observe for the considered dimensions d that the optimal acceptance rate maximizing the first-order efficiency remains very close to the limiting value of 0.704 predicted in Theorem 8.5. This numerical experiment shows that the efficiency improvement of fMALA compared to MALA is significant and indeed increases as the dimension d increases, which confirms the analysis of Section 8.3.1.

For our next experiments, we consider the d-dimensional zero-mean Gaussian distribution with covariance matrix I_d for $d = 1000$, as target distribution. We aim to numerically study the transient behaviour of fMALA and propose some solutions to overcome this issue. In Figure 8.2, we plot the squared norm of 10^4 samples generated by the RWM, MALA, fMALA and some hybrid strategies for MALA and fMALA, all started from the origin. We also include a zoom on the first 100 steps. In Figure 8.2a, we use standard implementations of the schemes. The time step h for each algorithm is chosen as the optimal parameter based on the optimal scaling results of all the algorithms at stationarity: for the RWM $h = 2.38^2 d^{-1}$, for MALA $h = 1.65^2 d^{-1/3}$ and for fMALA $h = 1.79^2 d^{-1/5}$. It can observed that MALA exhibits many rejected steps in contrast to RWM. This is a known issue of MALA in the transient phase [CRR05; JLM14] due to a tiny acceptance probability at first steps, and the same behaviour can be observed for fMALA, with zero accepted step in the present simulation. To circum-



(a) standard schemes. (b) hybrid methods using RWM ($h \sim d^{-1}$). (c) hybrid methods using MALA ($h \sim d^{-1/2}$).

Figure 8.3: Auto-correlation versus LAG for the Gaussian target density in dimension $d = 1000$. Comparison of fMALA with $h \sim d^{-1/5}$ (black), MALA with $h \sim d^{-1/3}$ (white), RWM with $h \sim d^{-1}$ (gray).

vent this issue, the following hybrid MALA scheme was presented in [CRR05]. The idea is to combine MALA with RWM at each step: with probability 1/2, we apply the MALA proposal (8.4) with step size $h = 1.65^2 d^{-1/3}$, the optimal parameter for MALA at stationarity. Otherwise, the RWM proposal (8.3) is used with step size $h = 2.38^2 d^{-1}$, the optimal parameter for the RWM at stationarity. Indeed, [CRR05] and [JLM14] have shown that the optimal scaling in the transient phase and at stationarity is the same and scales as d^{-1}. In Figure 8.2b, the plots for this hybrid MALA are presented, the same methodology is also applied for the hybrid fMALA scheme, showing a behaviour similar to hybrid MALA. In Figure 8.2c, the RWM proposal is replaced by the MALA proposal (8.4) with a different step size $h = 2d^{-1/2}$, which is the optimal parameter for MALA in the transient phase according to [CRR05]. Again, hybrid fMALA exhibits a behaviour similar to hybrid MALA.

In Figure 8.3, we consider again the same schemes and hybrid versions as in Figure 8.2, with the same step sizes, and we compare their autocorrelation function. We consider for each algorithms $2 \cdot 10^5$ iterations started at stationarity, where the first 10^3 iterations were discarded as burn-in. In Figure 8.3a, it can be observed that the autocorrelation associated with fMALA goes to 0 quicker than the RWM and MALA. In Figure 8.3b, and Figure 8.3c, we observe that by using hybrid strategies which are designed to robustify convergence from the transient phase, fMALA still comfortably outperforms MALA in terms of expected square efficiency (which is a stationary quantity).

Although our analysis applies only to product measure densities of the form (8.8), we next consider the following non-product density in $\mathbb{R}^d$, defined using a normalization constant Z_d and for $X_0 = 0$ as

$$\pi(X_1, \ldots, X_d) = Z_d \prod_{i=1}^{d} \frac{1}{1 + (X_i - \alpha(X_{i-1}))^2}, \tag{8.45}$$

where we consider the scalar functions $\alpha(x) = x/2$ and $\alpha(x) = \sin(x)$, respectively.

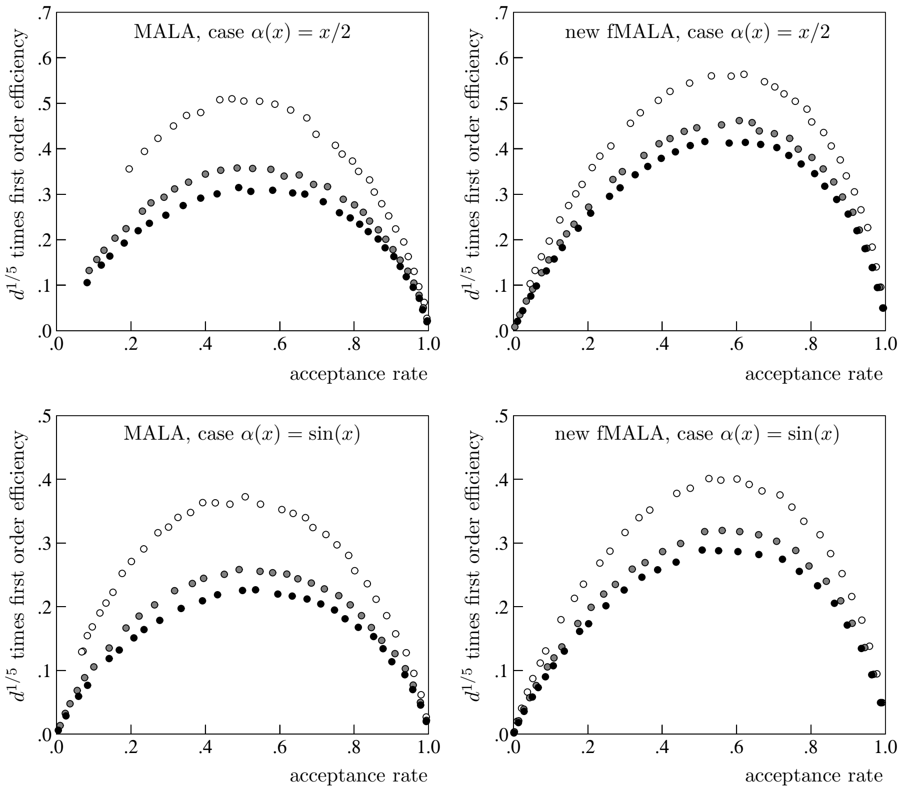


Figure 8.4: First-order efficiency of the new fMALA and the standard MALA as a function of the overall acceptance rates for the dimensions $d = 100$ (white points), $d = 500$ (gray points), $d = 1000$ (dark points), respectively, for the non product density (8.45) with $\alpha(x) = x/2$ (top pictures) and $\alpha(x) = \sin(x)$ (bottom pictures).

Notice that the density (8.45) is associated with the AR(1) process $X_i = \alpha(X_{i-1}) + Z_n$ with non Gaussian (Cauchy) increments Z_n. Furthermore, we observe that in this case the Jacobian in (8.12) is a symmetric tridiagonal matrix, which implies that the computational cost of the fMALA proposal is of the same order $\mathcal{O}(d)$ as the standard MALA proposal.

In Figure 8.4, we compare for many timesteps the standard MALA (left pictures) and the new fMALA (right pictures), and plot the (scaled) first order efficiency $\mathbb{E}[\|X_{k+1} - X_k\|^2/d]$ as a function of the overall acceptance rates, using the averages over 2×10^4 iterations of the algorithms. The initial condition for both algorithms is the same and is obtained after running 10^4 steps of the RWM algorithm to get close to the target probability measure. Analogously to the product case studied in Figure 8.1, we observe in both cases $\alpha(x) = x/2$ and $\alpha(x) = \sin(x)$ that the first-order efficiency of fMALA converges to a non-zero limiting curve with maximum close to the value 0.704. In contrast, the efficiency of the standard MALA drops to zero in this scaling where the first-order efficiency is multiplied with $d^{1/5}$. This numerical experiment suggests that our analysis in the product measure setting persists in the non product measure case.

8.6 Proof of Theorems 8.4 and 8.5

We provide here the proofs of Theorem 8.4 and Theorem 8.5 for the analysis of the optimal scaling properties of fMALA. We use tools analogous to that of [RGG97] and [RR98]. Consider the generator of the jump process $\Gamma^{d,\mathrm{fM}}$, defined for $\psi^d \in C_c^2(\mathbb{R}^d)$, and $x \in \mathbb{R}^d$ by

$$A_d^{\mathrm{fM}}\psi^d(x) = d^{1/5}\mathbb{E}\left[(\psi^d(y) - \psi^d(x))\alpha_d^{\mathrm{fM}}(x,y)\right] ,$$

where y follows the distribution defined by $q_d^{\mathrm{fM}}(x,\cdot)$. Also, consider the generator of the process $\{G_t, t \geq 0\}$, solution of (8.18), defined for $\psi \in C_c^2(\mathbb{R})$, and $x \in \mathbb{R}^d$ by

$$A^{\mathrm{fM}}\psi(x) = (\mathrm{h}(\ell)/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) .$$

We check that the assumptions of [EK86, Corollary 8.7, Chapter 4] are satisfied, which will imply Theorem 8.5. These assumptions consist in showing there exists a sequence of set $\{F_d \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$ such that for all $T \geq 0$:

$$\lim_{d\to+\infty} \mathbb{P}\left(\Gamma_s^{d,\mathrm{fM}} \in F_d \ , \ \forall s \in [0,T]\right) = 1$$

$$\lim_{d\to+\infty} \sup_{x\in F_d} \left|A_d^{\mathrm{fM}}\psi(x) - A^{\mathrm{fM}}\psi(x)\right| = 0 ,$$

for all functions ψ in a core of A^{fM}, which strongly separates points. Since A^{fM} is an operator on the set of function only depending on the first component, we restrict our study on this class of functions, which belong to $C_c^\infty(\mathbb{R})$, since by [EK86, Theorem 2.1, Chapter 8], this set of functions is a core for A^{fM} which strongly separates points. The following lemma is the proper result which was introduced in Section 8.2.2. For the sequel, let $\{\xi_i, i \in \mathbb{N}^*\}$ be a sequence of i.i.d. standard one-dimensional Gaussian random variables and X be a random variable distributed according to π_1. Also, for all $x \in \mathbb{R}^d$, denote by y^{fM} the proposal of fMALA, defined by (8.9), (8.12a) and (8.12b), started at $x \in \mathbb{R}^d$, with parameter h_d and associated with the d-dimensionnal Gaussian random variable $\{\xi_i, i = 1, \cdots, d\}$.

Lemma 8.19. *Assume Assumption 14. The following Taylor expansion in $\sqrt{h_d}$ holds: for all $x \in \mathbb{R}^d$ and $i \in \{1, \cdots, d\}$,*

$$\log\left(\frac{\pi(y_i^{\mathrm{fM}})q^{\mathrm{fM}}(y_i^{\mathrm{fM}}, x_i)}{\pi(x_i)q^{\mathrm{fM}}(x_i, y_i^{\mathrm{fM}})}\right) = C_5^{\mathrm{fM}}(x_i,\xi_i)d^{-1/2} + C_6^{\mathrm{fM}}(x_i,\xi_i)d^{-3/5} + C_7^{\mathrm{fM}}(x_i,\xi_i)d^{-7/10}$$
$$+ C_8^{\mathrm{fM}}(x_i,\xi_i)d^{-4/5} + C_9^{\mathrm{fM}}(x_i,\xi_i)d^{-9/10} + C_{10}^{\mathrm{fM}}(x_i,\xi_i)d^{-1} + C_{11}^{\mathrm{fM}}(x_i,\xi_i,h_d) , \quad (8.46)$$

where $C_5^{\mathrm{fM}}(x_1,\xi_1)$ is given in Appendix C. Furthermore, for $j = 6, \cdots, 10$, $C_j^{\mathrm{fM}}(x_i,\xi_i)$ are polynomials in ξ_i and derivatives of g at x_i and

$$\mathbb{E}\left[C_j^{\mathrm{fM}}(\mathrm{X},\xi_1)\right] = 0 \text{ for } j = 5, \cdots, 9 , \quad (8.47)$$

$$\mathbb{E}\left[\left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X},\xi_1)\,|\mathrm{X})\right]\right)^2\right] = \ell^{10}(K^{\mathrm{fM}})^2 = -2\mathbb{E}\left[C_{10}^{\mathrm{fM}}(\mathrm{X},\xi_1)\right] . \quad (8.48)$$

In addition, there exists a sequence of sets $\{F_d^1 \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$ *such that* $\lim_{d\to+\infty} d^{1/5}\pi_d((F_d^1)^c) = 0$ *and for* $j = 6, \cdots, 10$

$$\lim_{d\to+\infty} d^{-3/5} \sup_{x\in F_d^1} \mathbb{E}\left[\left|\sum_{i=2}^{d} C_j(x_i^d, \xi_i) - \mathbb{E}\left[C_j^{\mathrm{fM}}(\mathrm{X}, \xi_i)\right]\right|\right] = 0 \,, \tag{8.49}$$

and

$$\lim_{d\to+\infty} \sup_{x\in F_d^1} \mathbb{E}\left[\left|\sum_{i=2}^{d} C_{11}(x_i^d, \xi_i, h_d)\right|\right] = 0 \,. \tag{8.50}$$

Finally,

$$\lim_{d\to+\infty} \sup_{x\in F_d^1} \mathbb{E}\left[\left|\zeta^d\right|\right] = 0 \,, \tag{8.51}$$

with

$$\zeta^d = \sum_{i=2}^{d} \log\left(\frac{\pi(y_i^{\mathrm{fM}})q^{\mathrm{fM}}(y_i^{\mathrm{fM}}, x_i)}{\pi(x_i)q^{\mathrm{fM}}(x_i, y_i^{\mathrm{fM}})}\right) - \left(\left(d^{-1/2}\sum_{i=2}^{d} C_5(x_i^d, \xi_i)\right) - \ell^{10}(K^{\mathrm{fM}})^2/2\right) \,.$$

Proof. The Taylor expansion was computed using the computational software Mathematica [Wol14]. Then, since just odd powers of ξ_i occur in C_5, C_7 and C_9, we deduce (8.47) for $j = 5, 7, 9$. Furthermore by explicit calculation, the anti-derivative in x_1 of $e^{g(x_1)}\mathbb{E}\left[C_j^{\mathrm{fM}}(x_1, \xi_1)\right]$, for $j = 6, 8$, and $e^{g(x_1)}\mathbb{E}\left[C_5^{\mathrm{fM}}(x_1, \xi_1)^2 + 2C_{10}^{\mathrm{fM}}(x_1, \xi_1)\right]$ are on the form of some polynomials in the derivatives of g in x_1 times $e^{g(x_1)}$. Therefore, Assumption 14-(iii) implies (8.47) for $j = 6, 8$ and (8.48). We now build the sequence of sets F_d^1, which satisfies the claimed properties.
Denote for $j = 6, \cdots, 10$ and $x_i \in \mathbb{R}$, $\tilde{C}_j^{\mathrm{fM}}(x_i) = \mathbb{E}\left[C_j^{\mathrm{fM}}(x_i, \xi_i)\right]$ and $\mathrm{V}_j^{\mathrm{fM}}(x_i) = \mathrm{Var}\left[C_j^{\mathrm{fM}}(x_i, \xi_i)\right]$, which are bounded by a polynomial P_1 in x_i by Assumption 14-(ii) since $C_j^{\mathrm{fM}}(x_i, \xi_i)$ are polynomials in ξ_i and the derivatives of g at x_i. Therefore for all $k \in \mathbb{N}^*$,

$$\mathbb{E}\left[\left|\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right|^k\right] + \mathbb{E}\left[\left|\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right|^k\right] < \ddot{o} + \infty \,. \tag{8.52}$$

Consider for all $j = 6, \cdots, 10$, the sequence of sets $F_{d,j}^1 \in \mathbb{R}^d$ defined by $F_{d,j}^1 = F_{d,j,1}^1 \cap F_{d,j,2}^1$ where

$$F_{d,j,1}^1 = \left\{x \in \mathbb{R}^d \;;\; \left|\sum_{i=2}^{d} \tilde{C}_j^{\mathrm{fM}}(x_i) - \mathbb{E}\left[\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right]\right| \leq d^{23/40}\right\} \tag{8.53}$$

$$F_{d,j,2}^1 = \left\{x \in \mathbb{R}^d \;;\; \left|\sum_{i=2}^{d} \mathrm{V}_j^{\mathrm{fM}}(\mathrm{X}) - \mathbb{E}\left[\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right]\right| \leq d^{23/20}\right\} \,. \tag{8.54}$$

Note that $\lim_{d\to+\infty} d^{1/5}\pi_d((F_{d,j}^1)^c) = 0$ for all $j = 6 \cdots 10$, is implied by $\lim_{d\to+\infty} d^{1/5}\pi_d((F_{d,j,1}^1)^c) = 0$ and $\lim_{d\to+\infty} d^{1/5}\pi_d((F_{d,j,2}^1)^c) = 0$. Let $\{\mathrm{X}_i, i \geq 2\}$ be a sequence of i.i.d.random variables with distribution π_1. By definition of $F^1_{d,j,1}$, the Markov inequality and independence, we get

$$\begin{aligned} d^{1/5}\pi_d((F^1_{d,j,1})^c) &\leq d^{-21/10}\mathbb{E}\left[\left(\sum_{i=2}^{d}\tilde{C}_j^{\mathrm{fM}}(\mathrm{X}_i) - \mathbb{E}\left[\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^4\right] \\ &\leq \sum_{i_1,i_2=2}^{d} \mathbb{E}\left[\left(\tilde{C}_j^{\mathrm{fM}}(\mathrm{X}_{i_1}) - \mathbb{E}\left[\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^2 \left(\tilde{C}_j^{\mathrm{fM}}(\mathrm{X}_{i_2}) - \mathbb{E}\left[\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^2\right] \\ &\leq d^{-1/10}\ \mathbb{E}\left[\left(\tilde{C}_j^{\mathrm{fM}}(\mathrm{X}) - \mathbb{E}\left[\tilde{C}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^4\right] , \end{aligned} \tag{8.55}$$

where we have used the Young inequality for the last line. On another hand, using the Chebyshev and Hölder inequality, we get

$$\begin{aligned} d^{1/5}\pi_d((F^1_{d,j,2})^c) &\leq d^{-21/10}\mathbb{E}\left[\left(\sum_{i=2}^{d}\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X}_i) - \mathbb{E}\left[\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^2\right] \\ &\leq d^{-1/10}\ \mathbb{E}\left[\left(\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X}) - \mathbb{E}\left[\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right]\right)^2\right] . \end{aligned} \tag{8.56}$$

Therefore (8.52), (8.55) and (8.56) imply that $\lim_{d\to+\infty} d^{1/5}\pi_d((F^1_{d,j})^c) = 0$ for all $j = 6,\cdots,10$. In addition, for all $x \in F^1_{d,j}$, by the triangle inequality and the Cauchy-Schwarz inequality we have for all $j = 6,\cdots,10$

$$\begin{aligned} \mathbb{E}\left[\left|\sum_{i=2}^{d} C_j^{\mathrm{fM}}(x_i,\xi_i) - \mathbb{E}\left[C_j^{\mathrm{fM}}(\mathrm{X},\xi_i)\right]\right|\right] &\leq \left|\sum_{i=2}^{d}\mathrm{V}_j^{\mathrm{fM}}(x_i) - \mathbb{E}\left[\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right]\right|^{1/2} \\ &+ d^{1/2}\mathbb{E}\left[\mathrm{V}_j^{\mathrm{fM}}(\mathrm{X})\right]^{1/2} + \left|\sum_{i=2}^{d}\tilde{C}_j^{\mathrm{fM}}(x_i) - \mathbb{E}\left[C_j^{\mathrm{fM}}(\mathrm{X},\xi_i)\right]\right| . \end{aligned}$$

Therefore by this inequality, (8.53) and (8.54), there exists a constant M_1 such that

$$d^{3/5}\sup_{x\in F^1_{d,j}}\mathbb{E}\left[\left|\sum_{i=2}^{d} C_j^{\mathrm{fM}}(x_i,\xi_i) - \mathbb{E}\left[C_j^{\mathrm{fM}}(\mathrm{X},\xi_i)\right]\right|\right] \leq d^{-1/40}M_1 ,$$

and (8.49) follows. It remains to show (8.50). By definition, C_{11} is the remainder in the eleventh order expansion in $\sigma_d := \sqrt{h_d}$ given by (8.46) of the function Θ defined by $\Theta(x_i,\xi_i,\sigma_d) = \log(\pi_1(y_i^{\mathrm{fM}})q_1^{\mathrm{fM}}(y_i^{\mathrm{fM}},x_i)) - \log(\pi_1(x_i)q_1^{\mathrm{fM}}(x_i,y_i^{\mathrm{fM}}))$. Therefore, by the mean-value form of the remainder, there exists $u_d \in [0,\sigma_d]$ such that

$$C_{11}(x_i,\xi_i,h_d) = (\sigma_d^{11}/(11!))\frac{\partial^{11}\Theta}{\partial\sigma_d^{11}}(x_i,\xi_i,u_d) .$$

By Assumption 14-(i) which implies that g'' is bounded, and Assumption 14-(ii), for all $u_d \in [0,\sigma_d]$, the eleventh derivative of Θ with respect to σ_d, taken in (x_i,ξ_i,u_d), can be bounded by a positive polynomial in (x_i, ξ_i) on the form $\mathrm{P}_2(x_i)\mathrm{P}_3(\xi_i)$. Hence, there exists a constant M_2 such that

$$\mathbb{E}\left[|C_{11}(x_i, \xi_i, h_d)|\right] \leq M_2\ d^{-11/10}\ \mathrm{P}_2(x_i)\ . \tag{8.57}$$

And if we define

$$F^1_{d,11} = \left\{ x \in \mathbb{R}^d\ ;\ \left| \sum_{i=2}^{d} \mathrm{P}_2(x_i) - \mathbb{E}\left[\mathrm{P}_2(\mathrm{X})\right] \right| \leq d \right\}\ ,$$

then we have by the Chebychev inequality, this definition and (8.57)

$$d^{1/5}\pi_d((F^1_{d,11})^c) \leq \mathrm{Var}\left[\mathrm{P}_2(\mathrm{X})\right] d^{-4/5}$$

$$\sup_{x\in F^1_{d,11}} \sum_{i=2}^{d} \mathbb{E}\left[|C_{11}(x_i, \xi_i, h_d)|\right] \leq M_2(\mathbb{E}\left[\mathrm{P}_2(x)\right] + 1)d^{-1/10}\ .$$

These results, combined with Assumption 14-(iii), imply $\lim_{d\to+\infty} d^{1/5}\pi_d((F^1_{d,11})^c) = 0$ and (8.50). Finally, $F^1_d = \bigcap_{j=6}^{11} F^1_{d,j}$ satisfies the claimed properties of the Lemma, and (8.51) directly follows from all the previous results. □

To isolate the first component of the process $\Gamma^{d,\mathrm{fM}}$, we consider the modified generators defined for $\psi \in C^2_c(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ by

$$\tilde{A}^{\mathrm{fM}}_d \psi(x) = d^{1/5}\mathbb{E}\left[(\psi(y^{\mathrm{fM}}) - \psi(x))\alpha^{\mathrm{fM}}_{-1,d}(x, y^{\mathrm{fM}})\right]$$

where for all $x, y \in \mathbb{R}^d$,

$$\alpha^{\mathrm{fM}}_{-1,d}(x, y) = \prod_{i=2}^{d} \frac{\pi_1(y_i)q_{1,\mathrm{fM}}(y_i, x_i)}{\pi_1(x_i)q_{1,\mathrm{fM}}(x_i, y_i)}\ .$$

The next lemma shows that we can approximate A^{fM}_d by $\tilde{A}^{\mathrm{fM}}_d$, and thus, in essence, the first component becomes "asymptotically independent" from the others.

Theorem 8.20. *There exists a sequence of sets* $\{F^2_d \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$ *such that* $\lim_{d\to+\infty} d^{1/5}\pi_d((F^2_d)^c) = 0$ *and for all* $\psi \in C^\infty_c(\mathbb{R})$ *(seen as function of* $\mathbb{R}^d$ *for all* d *which only depends on the first component):*

$$\lim_{d\to+\infty} \sup_{x\in F^2_d} \left| A^{\mathrm{fM}}_d \psi(x) - \tilde{A}^{\mathrm{fM}}_d \psi(x) \right| = 0\ .$$

In addition,

$$\lim_{d\to+\infty} \sup_{x\in F^2_d} d^{1/5}\mathbb{E}\left[\left|\alpha^{\mathrm{fM}}_d(x, y^{\mathrm{fM}}) - \alpha^{\mathrm{fM}}_{-1,d}(x, y^{\mathrm{fM}})\right|\right] = 0\ . \tag{8.58}$$

Proof. Using that ψ is bounded and the Jensen inequality, there exists a constant M_1 such that

$$\left|A_d^{\mathrm{fM}}\psi(x) - \tilde{A}_d^{\mathrm{fM}}\psi(x)\right| \leq M_1 d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] .$$

Thus it suffices to show (8.58). Set $\sigma_d = \sqrt{h_d}$. Since $t \mapsto 1 \wedge \exp(t)$ is 1-Lipschtz on $\mathbb{R}$ and, by definition we have

$$d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] \leq d^{1/5}\mathbb{E}\left[|\Theta(x_1, \xi_1, \sigma_d)|\right] , \tag{8.59}$$

where $\Theta(x_1, \xi_1, \sigma_d) = \log(\pi_1(y_1^{\mathrm{fM}})q_1^{\mathrm{fM}}(y_1^{\mathrm{fM}}, x_1)) - \log(\pi_1(x_1)q_1^{\mathrm{fM}}(x_1, y_1^{\mathrm{fM}}))$. By a fifth order Taylor expansion of Θ in σ_d, and since by (8.46) $\partial^j\Theta(x_1, \xi_1, 0)/(\partial\sigma_d^j) = 0$ for $j = 0 \cdots 4$, we have

$$\Theta(x_1, \xi_1, \sigma_d) = \frac{\partial^5\Theta}{\partial\sigma_d^5}(x_1, \xi_1, u_d)(\sigma_d^5/5!) ,$$

for some $u_d \in [0, \sigma_d]$. Using Assumption 14-(i)-(ii), and an explicit expression of $\partial^j\Theta(x_1, \xi_1, u_d)/(\partial\sigma_d^j)$, there exists two positive polynomials P_1 and P_2 such that

$$|\Theta(x_1, \xi_1, \sigma_d)| \leq (\sigma_d^5/5!)\mathrm{P}_1(x_1)\mathrm{P}_2(\xi_1) .$$

Plugging this result in (8.59) and since $\sigma_d^5 = \ell^{5/2}d^{-1/2}$, we get

$$d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] \leq \ell^{5/2}d^{-3/10}\mathrm{P}_1(x_1) .$$

Setting $F_d^2 = \{x \in \mathbb{R}^d \ ; \ \mathrm{P}_1(x_1) \leq d^{1/10}\}$, we have

$$\sup_{x \in F_d^2} d^{1/5}\mathbb{E}\left[\left|\alpha_d^{\mathrm{fM}}(x, y^{\mathrm{fM}}) - \alpha_{-1,d}^{\mathrm{fM}}(x, y^{\mathrm{fM}})\right|\right] \leq \ell^{5/2}d^{-1/5} ,$$

and (8.58) follows. Finally, F_d^2 satisfied $\lim_{d\to+\infty} d^{1/5}\pi_d((F_d^2)^c) = 0$ since by the Markov inequality

$$d^{1/5}\pi_d((F_d^2)^c) \leq d^{-1/10}\mathbb{E}\left[\mathrm{P}_1(\mathrm{X})^3\right] ,$$

where $\mathbb{E}\left[\mathrm{P}_1(\mathrm{X})^3\right]$ is finite by Assumption 14-(iii). □

Lemma 8.21. *For all $\psi \in C_c^\infty(\mathbb{R})$,*

$$\lim_{d\to+\infty}\sup_{x_1\in\mathbb{R}}\left|d^{1/5}\mathbb{E}\left[\psi(y_1^{\mathrm{fM}}) - \psi(x_1)\right] - (\ell^2/2)(\psi'(x_1)f(x_1) + \psi''(x_1))\right| = 0 .$$

Proof. Consider $\sigma_d = \sqrt{h_d}$ and $W(x_1, \xi_1, \sigma_d) = \psi(y_1^{\mathrm{fM}})$. Note that $W(x_1, \xi_1, 0) = \psi(x_1)$. Then using that $\psi \in C_c^\infty(\mathbb{R})$, a third order Taylor expansion of this function in σ_d implies there exists $u_d \in [0, h_d]$ and $M_1 \geq 0$ such that

$$\mathbb{E}\left[W(x_1, \xi_1, \sigma_d) - \psi(x_1)\right] = (\ell^2 d^{-1/5}/2)(\psi'(x_1)f(x_1) + \psi''(x_1)) + M_1 d^{-3/10} + \frac{\partial^3 W}{\partial\sigma_d^3}(x_1, \xi_1, u_d)\sigma_d^3 .$$

Moreover since $\psi \in C_c^\infty(\mathbb{R})$, the third partial derivative of W in σ_d are bounded for all x_1, ξ_1 and σ_d. Therefore there exists $M_2 \geq 0$ such that for all $x_1 \in \mathbb{R}$,

$$\left| d^{1/5}\mathbb{E}\left[\psi(y_1^{\mathrm{fM}}) - \psi(x_1)\right] - (\ell^2/2)(\psi'(x_1)f(x_1) + \psi''(x_1))\right| \leq M_2 \ell^{3/2} d^{-1/10} ,$$

which concludes the proof. □

As proceed in [RR98], we prove a uniform central limit theorem for the sequence of random variables defined for $i \geq 2$ and $x_i \in \mathbb{R}$ by $C_5^{\mathrm{fM}}(x_i, \xi_i)$. Define now for $d \geq 2$ and $x \in \mathbb{R}^d$,

$$\bar{\mathsf{M}}_d(x) = n^{-1/2} \sum_{i=2}^{d} C_5^{\mathrm{fM}}(x_i, \xi_i) ,$$

and the characteristic function of $\bar{\mathsf{M}}_d$ for $t \in \mathbb{R}$ by

$$\varphi_d(x,t) = \mathbb{E}[e^{\mathrm{i}t\bar{\mathsf{M}}_d(x)}] .$$

Finally define the characteristic function of the zero-mean Gaussian distribution with standard deviation K^{fM}, given in Lemma 8.19, by: for $t \in \mathbb{R}$,

$$\varphi(t) = e^{-(K^{\mathrm{fM}})^2 t/2} .$$

Lemma 8.22. *There exists a sequence of set* $\{F_d^3 \subset \mathbb{R}^d, d \in \mathbb{N}^*\}$, *satisfying* $\lim_{d\to+\infty} d^{1/5}\pi_d((F_d^3)^c) = 0$ *and we have the following properties:*

(i) for all $t \in \mathbb{R}$, $\lim_{d\to+\infty} \sup_{x\in F_d^3} |\varphi_d(x,t) - \varphi(t)| = 0$,

(ii) for all bounded continuous function $\mathrm{b} : \mathbb{R} \to \mathbb{R}$,

$$\lim_{d\to+\infty} \sup_{x\in F_d^3} \left| \mathbb{E}\left[\mathrm{b}\left(\bar{\mathsf{M}}_d(x)\right)\right] - (2\pi\ell^{10}(K^{\mathrm{fM}})^2)^{-1/2} \int_{\mathbb{R}} \mathrm{b}(u) e^{-u^2/(2\ell^{10}(K^{\mathrm{fM}})^2)} \mathrm{d}u \right| = 0 .$$

In particular, we have

$$\lim_{d\to+\infty} \sup_{x\in F_d^3} \left| \mathbb{E}\left[1 \wedge e^{\bar{\mathsf{M}}_d(x) - \ell^{10}(K^{\mathrm{fM}})^2/2}\right] - 2\Phi(\ell^5 K^{\mathrm{fM}}/2) \right| = 0 .$$

Proof. We first define for all $d \geq 1$, $F_d^3 = F_{d,1}^3 \cap F_{d,2}^3$ where

$$F_{d,1}^3 = \bigcap_{j=2,4} \left\{ x \in \mathbb{R}^d \; ; \left| \; d^{-1} \sum_{i=2}^{d} \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^j\right] - \mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1,\xi_1)^j\right] \right| \leq d^{-1/4} \right\} , \tag{8.60}$$

$$F_{d,2}^3 = \left\{ x \in \mathbb{R}^d \; ; \; \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^2\right] \leq d^{3/4} \;\; \forall i \in \{2, \cdots, d\} \right\} . \tag{8.61}$$

It follows from (8.52), and the Chebychev and Markov inequalities that there exists a constant M such that $\pi_d((F_{d,1}^3)^c) + \pi_d((F_{d,2}^3)^c) \leq M d^{-1/2}$. Therefore $\lim_{d\to+\infty} d^{1/5}\pi_d((F_d^3)^c) =$

0.
(i). Let $t \in \mathbb{R}$ and $x \in F_d^3$ and denote

$$\mathrm{V}(x_i) = \mathrm{Var}[C_5^{\mathrm{fM}}(x_i, \xi_i)] = \mathbb{E}\left[C_5^{\mathrm{fM}}(x_i, \xi_i)^2\right] ,$$

where the second equality follows from Lemma 8.19. By the triangle inequality

$$\begin{aligned} |\varphi_d(x,t) - \varphi(t)| \leq \left|\varphi_d(x,t) - \prod_{i=2}^{d}\left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)\right| \\ + \left|\prod_{i=2}^{d}\left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right) - e^{-\ell^{10}(K^{\mathrm{fM}})^2 t^2/2}\right| . \end{aligned} \tag{8.62}$$

We bound the two terms of the right hand side separately. Note that by independence for all d, $\varphi_d(x,t) = \prod_{i=2}^{d} \varphi_1(x_i, t/\sqrt{d})$. Since $x \in F_d^3$, by (8.61), for d large enough $\ell^{10}\mathrm{V}(x_i)t^2/(2d) \leq 1$ for all $i \in \{2, \cdots, d\}$. Thus, by [Bil95, Eq. 26.5], we have for such large d, all $i \in \{2, \cdots, d\}$ and all $\delta > 0$:

$$\begin{aligned} \left|\varphi_1(x_i, t/\sqrt{d}) - \left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)\right| &\leq \mathbb{E}\left[\left(\frac{|t|^3\,\ell^{15}}{6d^{3/2}}\left|C_5^{\mathrm{fM}}(x_i,\xi_i)\right|^3\right) \wedge \left(\frac{t^2\ell^{10}}{d} C_5^{\mathrm{fM}}(x_i,\xi_i)^2\right)\right] \\ &\leq \mathbb{E}\left[\frac{|t|^3\,\ell^{15}}{6d^{3/2}}\left|C_5^{\mathrm{fM}}(x_i,\xi_i)\right|^3 \mathbb{1}_{\{|C_5^{\mathrm{fM}}(x_i,\xi_i)| \leq \delta d^{1/2}\}}\right] \\ &\quad + \mathbb{E}\left[\frac{t^2\ell^{10}}{d} C_5^{\mathrm{fM}}(x_i,\xi_i)^2 \mathbb{1}_{\{|C_5^{\mathrm{fM}}(x_i,\xi_i)| > \delta d^{1/2}\}}\right] \\ &\leq \frac{\delta\,|t|^3\,\ell^{15}}{6d}\mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^2\right] + \frac{\ell^{10}t^2}{\delta^2 d^2}\mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^4\right] , \end{aligned}$$

In addition, by [Bil95, Lemma 1, Section 27] and using this result we get:

$$\begin{aligned} \left|\varphi_d(x,t) - \prod_{i=2}^{d}\left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)\right| &\leq \sum_{i=2}^{d} \frac{\delta\,|t|^3\,\ell^{15}}{6d}\mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^2\right] \\ &\quad + \frac{\ell^{10}t^2}{\delta^2 d^2}\mathbb{E}\left[C_5^{\mathrm{fM}}(x_i,\xi_i)^4\right] \\ &\leq \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1,\xi_1)^2\right] + d^{-1/4}\right)\ell^{15}\delta\,|t|^3/6 \\ &\quad + \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1,\xi_1)^4\right] + d^{-1/4}\right)\ell^{10}t^2/(\delta^2 d) , \end{aligned}$$

where the last inequality follows from $x \in F_d^3$ and (8.60) Let now $\epsilon > 0$, and choose δ small enough such that the fist term is smaller than $\epsilon/2$. Then there exists $d_0 \in \mathbb{N}^*$ such that for all $d \geq d_0$, the second term is smaller than $\epsilon/2$ as well. Therefore, for $d \geq d_0$ we get

$$\sup_{x \in F_d^3}\left|\varphi_d(x,t) - \prod_{i=2}^{d}\left(1 - \frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)\right| \leq \epsilon .$$

Consider now the second term of (8.62), by the triangle inequality,

$$\left|\prod_{i=2}^{d}\left(1-\frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)-e^{-\ell^{10}(K^{\mathrm{fM}})^2t^2/2}\right| \leq \left|\prod_{i=2}^{d}\left(1-\frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)-\prod_{i=2}^{d}e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)}\right| \\ +\left|\prod_{i=2}^{d}e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)}-e^{-\ell^{10}(K^{\mathrm{fM}})^2t^2/2}\right| . \quad (8.63)$$

We deal with the two terms separatly. First since for all x_i, $\mathrm{V}(x_i) \geq 0$, we have

$$\left|1-\mathrm{V}(x_i)\ell^{10}t^2/(2d)-e^{-\mathrm{V}(x_i)\ell^{10}t^2/(2d)}\right| \leq \mathrm{V}(x_i)^2\ell^{20}t^4/(8d^2) .$$

Using this result, [Bil95, Lemma 1, Section 27] and the Cauchy-Schwarz inequality, it follows:

$$\begin{aligned}&\left|\prod_{i=2}^{d}\left(1-\frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)-\prod_{i=2}^{d}e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)}\right| \\ &\leq \sum_{i=2}^{d}\left|1-\mathrm{V}(x_i)\ell^{10}t^2/(2d)-e^{-\mathrm{V}(x_i)\ell^{10}t^2/(2d)}\right| \\ &\leq \sum_{i=2}^{d}\mathrm{V}(x_i)^2\ell^{20}t^4/(8d^2) \leq \left(\mathbb{E}\left[C_5^{\mathrm{fM}}(\mathrm{X}_1,\xi_1)^4\right]+d^{-1/4}\right)\ell^{20}t^4/(8d) , \end{aligned} \quad (8.64)$$

where the last inequality is implied by (8.60). Finally since on $\mathbb{R}_-$, $u \mapsto e^u$ is 1-Lipschitz and using (8.60), we get

$$\begin{aligned}\left|\prod_{i=2}^{d}e^{-\ell^{10}\mathrm{V}(x_i)t^2/(2d)}-e^{-\ell^{10}(K^{\mathrm{fM}})^2t^2/2}\right| &\leq (t^2\ell^{10}/2)\left|\sum_{i=2}^{d}d^{-1}\mathrm{V}(x_i)-(K^{\mathrm{fM}})^2\right| \\ &\leq t^2\ell^{10}d^{-1/4}/2 . \end{aligned} \quad (8.65)$$

Therefore, combining (8.64) and (8.65) in (8.63), we get:

$$\lim_{d\to+\infty}\sup_{x\in F_d^3}\left|\prod_{i=2}^{d}\left(1-\frac{\ell^{10}\mathrm{V}(x_i)t^2}{2d}\right)-e^{-\ell^{10}(K^{\mathrm{fM}})^2t^2/2}\right|=0 ,$$

which concludes the proof of (i).
(ii) follows now from (i) by the continuity theorem applied to an appropriate sequence $\{x^d\ ,\ d\in\mathbb{N}^*\}$. □

proof of Theorem 8.4. The theorem follows from Lemma 8.19, (8.58) in Theorem 8.20 and the last statement in Lemma 8.22. □

proof of Theorem 8.5. Consider $F_d = \bigcap_{j=1,2,3} F_d^j$, where the sets F_d^j are given resp. in Lemma 8.19 Theorem 8.20 and Lemma 8.22. We then obtain $\lim_{d\to+\infty} d^{-1/5}\pi_d((F_d)^c) = 0$ and by the union bound, for all $T \geq 0$,

$$\lim_{d\to+\infty} \mathbb{P}\left(\Gamma_s^{d,\mathrm{fM}} \in F_d \ , \ \forall s \in [0,T]\right) = 1 \ .$$

Furthermore, combining the former results with Lemma 8.21, we have for all $\psi \in C_c^\infty(\mathbb{R})$ (seen as a function of the first component):

$$\lim_{d\to+\infty} \sup_{x\in F_d} \left| A_d^{\mathrm{fM}}\psi(x) - A^{\mathrm{fM}}\psi(x) \right| = 0 \ .$$

Then, the weak convergence follows from [EK86, Corollary 8.7, Chapter 4]. □

8.7 Postponed proofs

8.7.1 Proof of Lemma 8.9

By Assumption 15-16, π and q are positive and continuous. It follows from [MT96, Lemma 1.2] that P is Leb^d-irreducible aperiodic, where Leb^d is the Lebesgue measure on $\mathbb{R}^d$. In addition, all compact set $\mathcal{C}$ such that $\mathrm{Leb}^d(\mathcal{C}) > 0$ are small for P. Now by [MT09, Theorem 15.0.1], we just need to check the drift condition (8.20). But by a simple calculation, using $\alpha(x,y) \leq 1$ for all $x, y \in \mathbb{R}^d$, and the Cauchy-Schwarz inequality, we get

$$\begin{aligned} PV(x) &\leq 1 + \|x\|^2 + (\|\mu(x)\|^2 - \|x\|^2) \int_{\mathbb{R}^d} \alpha(x,y) q(x,y) \mathrm{d}y \\ &\quad + (2\pi)^{-d/2} (2\|\mu(x)\|\,\|S(x)\| + \|S(x)\|^2) \int_{\mathbb{R}^d} \max(\|\xi\|^2, 1) e^{-\|\xi\|^2/2} \mathrm{d}\xi \ . \end{aligned}$$

By (8.23), $\limsup_{\|x\|\to+\infty} (2\|\mu(x)\|\,\|S(x)\| + \|S(x)\|^2)\,\|x\|^{-2} = 0$. Therefore, using again the first inequality of (8.23) and Assumption 17:

$$\limsup_{\|x\|\to+\infty} PV(x)/V(x) \leq 1 - (1-\tau^2) \liminf_{\|x\|\to+\infty} \int_{\mathbb{R}^d} \alpha(x,y) q(x,y) \mathrm{d}y < 1 \ .$$

This concludes the proof of Lemma 8.9. □

8.7.2 Proof of Theorem 8.10

We prove this result by contradiction. The strategy of the proof is the following: first, under our assumptions, most of the proposed moves by the algorithm has a norm which is greater than the current point. However, if P is geometrically ergodic, then it implies a upper bound on the rejection probability of the algorithm by some constant strictly smaller than 1. But combining these facts, we can exhibit a sequence of point $\{x_n, n \in \mathbb{N}\}$, such that $\lim_{n\to+\infty} \pi(x_n) = +\infty$. Since we assume that π is bounded, we have our contradiction.

If P is geometrically ergodic, then by [RT96a, Theorem 5.1], there exists $\eta > 0$ such that for almost every $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} \alpha(x,y)q(x,y)\mathrm{d}y \geq \eta \ , \tag{8.66}$$

and let $M \geq 0$ such that

$$\mathbb{P}\left[\|\xi\| \geq M\right] \leq \eta/2 \ , \tag{8.67}$$

where ξ is a standard d-dimensional Gaussian random variable. By (8.24), there exist $R_\epsilon, \delta > 0$ such that

$$\inf_{\{\|x\|\geq R_\epsilon\}} \left\|S(x)^{-1}\mu(x)\right\| \|x\|^{-1} \geq \epsilon^{-1} + \delta \tag{8.68}$$

$$\inf_{\{\|x\|\geq R_\epsilon\}} \inf_{\|z\|=1} \|S(x)z\| \geq \epsilon(1+\delta\epsilon/2)^{-1} \ . \tag{8.69}$$

Note that we can assume R_ϵ is large enough so that

$$\epsilon\delta R_\epsilon/2 \geq M \ . \tag{8.70}$$

Now define for $x \in \mathbb{R}^d$, $\|x\| \geq R_\epsilon$

$$B(x) = \left\{y \in \mathbb{R}^d \mid \left\|S(x)^{-1}(y-\mu(x))\right\| \leq M\right\} \ . \tag{8.71}$$

Note if $y \in B(x)$, we have by definition and the triangle inequality $\|S(x)^{-1}y\| \geq \|S(x)^{-1}\mu(x)\| - M$. Therefore by (8.68)-(8.69) and (8.70)

$$\begin{aligned}\|y\| = \left\|S(x)S(x)^{-1}y\right\| &\geq \epsilon(1+\delta\epsilon/2)^{-1}\left\|S(x)^{-1}y\right\| \\ &\geq \epsilon(1+\delta\epsilon/2)^{-1}\left\{(\epsilon^{-1}+\delta)\|x\| - M\right\} \geq \|x\| \ .\end{aligned} \tag{8.72}$$

We then show that this inequality implies

$$\liminf_{\|x\|\to+\infty} \inf_{y\in B(x)} \frac{q(y,x)}{q(x,y)} = 0 \ . \tag{8.73}$$

Let $x \in \mathbb{R}^d$, $\|x\| \geq R_\epsilon$, $y \in B(x)$. First, it is straightforward by (8.71), that $|S(x)|\, q(x,y)$ is uniformly bounded away from 0, and it suffices to consider $|S(x)|\, q(y,x)$. By (8.69)-(8.72), we have $\|y\| \geq R_\epsilon$ and for all $z \in \mathbb{R}^d$, $\|S(y)z\| \geq \epsilon(1+\delta\epsilon/2)^{-1}\|z\|$, which implies for all $z \in \mathbb{R}^d$, $\epsilon^{-1}(1+\delta\epsilon/2)\|z\| \geq \|S(y)^{-1}z\|$. By this inequality and (8.68), we have

$$\begin{aligned}\left|\left\|S(y)^{-1}\mu(y)\right\| - \left\|S(y)^{-1}x\right\|\right| &\geq \left\|S(y)^{-1}\mu(y)\right\| - \left\|S(y)^{-1}x\right\| \\ &\geq (\epsilon^{-1}+\delta)\|y\| - \epsilon^{-1}(1+\delta\epsilon/2)\|x\| \geq (\delta/2)\|y\| \ ,\end{aligned} \tag{8.74}$$

where the last inequality follows from (8.72). Using this result, the triangle inequality, (8.74)-(8.69) and (8.72), we get

$$\begin{aligned}
q(y,x) &= (2\pi)^{-d/2}\exp\left\{-(1/2)\left\|S(y)^{-1}(x-\mu(y))\right\|^2 - \log(|S(y)|)\right\} \\
&\leq (2\pi)^{-d/2}\exp\left\{-(1/2)\left(\left\|S(y)^{-1}\mu(y)\right\| - \left\|S(y)^{-1}x\right\|\right)^2 - \log(|S(y)|)\right\} \\
&\leq (2\pi)^{-d/2}\exp\left\{-(\delta^2/8)\left\|y\right\|^2 - \log(|S(y)|)\right\} \\
&\leq (2\pi)^{-d/2}\exp\left\{-(\delta^2/8)\left\|x\right\|^2 - d\log(\epsilon(1+\delta\epsilon/2)^{-1})\right\} .
\end{aligned}$$

Using this inequality and (8.25) imply $\lim_{\|x\|\to+\infty}\inf_{y\in B(x)}|S(x)|\,q(y,x) = 0$ and then (8.73). Therefore there exists $R_q \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq R_q$

$$\inf_{y\in B(x)}\frac{q(y,x)}{q(x,y)} \leq \eta/4 . \tag{8.75}$$

Now we are able to build the sequence $\{x_n, n \in \mathbb{N}\}$ such that for all $n \in \mathbb{N}$, $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$ and $\lim_{n\to+\infty}\pi(x_n) = +\infty$. Indeed let $x_0 \in \mathbb{R}^d$ such that $\|x_0\| \geq \max(R_\epsilon, R_q)$. Assume, we have built the sequence up to the nth term and such that for all $k = 0, \ldots, n-1$, $\|x_{k+1}\| \geq \max(R_\epsilon, R_q)$ and $\pi(x_{k+1}) \geq (3/2)\pi(x_k)$. Now we choose x_{n+1} depending on x_n, satisfying $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$ and $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$. Since $\|x_n\| \geq \max(R_\epsilon, R_q)$, by (8.66)-(8.67) and (8.75)

$$\begin{aligned}
\eta &\leq \int_{\mathbb{R}^d}\alpha(x_n,y)q(x_n,y)\mathrm{d}y \leq \eta/2 + \int_{B(x_n)}\min\left(1, \frac{\pi(y)q(y,x_n)}{\pi(x_n)q(x_n,y)}\right)q(x_n,y)\mathrm{d}y \\
&\leq \eta/2 + (\eta/4)\int_{B(x_n)}\frac{\pi(y)}{\pi(x_n)}q(x_n,y)\mathrm{d}y .
\end{aligned}$$

This inequality implies that $\int_{B(x_n)}\frac{\pi(y)}{\pi(x_n)}q(x_n,y)\mathrm{d}y \geq 2$ and therefore there exists $x_{n+1} \in B(x_n)$ such that $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$, and since $x_{n+1} \in B(x_n)$ by (8.72), $\|x_{n+1}\| \geq \max(R_\epsilon, R_q)$. Therefore, we have a sequence $\{x_n, n \in \mathbb{N}\}$ such that for all $n \in \mathbb{N}$, $\pi(x_{n+1}) \geq (3/2)\pi(x_n)$. Since by assumption $\pi(x_0) > 0$, we get $\lim_{n\to+\infty}\pi(x_n) = +\infty$, which contradicts the assumption that π is bounded. This concludes the proof of Theorem 8.10. □

Appendix A

Markov processes

In this section, we give some definitions and fundamental results on Markov processes and related objects. However, basic knowledge and definition is assumed to be known.

A.1 Markov chains

Let P be a Markov kernel on a state space $(\mathsf{E}, \mathcal{E})$. For all initial distribution μ_0 on $(\mathsf{E}, \mathcal{E})$, $\mathbb{P}_{\mu_0}$ and $\mathbb{E}_{\mu_0}$ denote the probability and the expectation respectively, associated with P and μ_0 on the canonical space $(\mathsf{E}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$, and $(Y_i)_{i \in \mathbb{N}}$ denotes the canonical process.

Tensor product of Markov kernels

Definition-Proposition A.1 ([DMS14, Proposition 5.8])**.** *Let $(\mathsf{E}, \mathcal{E})$, $(\mathsf{F}, \mathcal{F})$ and $(\mathsf{G}, \mathcal{G})$ be measurable spaces. Let P and Q be two Markov kernels on $\mathsf{E} \times \mathcal{F}$ and $\mathsf{F} \times \mathcal{G}$ respectively.*

(a) There exists a Markov kernel $P \otimes Q$ on $\mathsf{E} \times (\mathcal{F} \otimes \mathcal{G})$ such that for all $x \in \mathsf{E}$ and $\mathtt{A} \in \mathcal{F} \otimes \mathcal{G}$:

$$P \otimes Q(x, \mathtt{A}) = \int_{\mathsf{F}} P(x, \mathrm{d}y) \int_{\mathsf{G}} \mathbb{1}_{\mathtt{A}}(y, z) Q(y, \mathrm{d}z) \; .$$

In particular, if P is simply a measure μ on $(\mathsf{E}, \mathcal{E})$, then it yields that $\mu \otimes Q$ is a measure on $(\mathsf{F} \times \mathsf{G}, \mathcal{F} \otimes \mathcal{G})$.

(b) There exists a Markov kernel $P \otimes^{\mathrm{T}} Q$ on $\mathsf{E} \times (\mathcal{G} \otimes \mathcal{F})$ such that for all $x \in \mathsf{E}$ and $\mathtt{A} \in \mathcal{G} \otimes \mathcal{F}$:

$$P \otimes^{\mathrm{T}} Q(x, \mathtt{A}) = \int_{\mathsf{F}} P(x, \mathrm{d}y) \int_{\mathsf{G}} \mathbb{1}_{\mathtt{A}}(z, y) Q(y, \mathrm{d}z) \; .$$

In particular, if P is simply a measure μ on $(\mathsf{E}, \mathcal{E})$, then it yields that $\mu \otimes^{\mathrm{T}} Q$ is a measure on $(\mathsf{G} \times \mathsf{F}, \mathcal{G} \otimes \mathcal{F})$.

Irreducibility

Definition A.2. *1. Let μ be a non-trivial σ-finite measure on $(\mathsf{E}, \mathcal{E})$. P is said μ-irreducible if for all $\mathtt{A} \in \mathcal{E}$, $\mu(\mathtt{A}) > 0$, for all $x \in \mathsf{E}$, there exists $n \in \mathbb{N}^*$ such that $P^n(x, \mathtt{A}) > 0$.*

2. P is said to be irreducible is there exists μ, a non-trivial σ-finite measure on $(\mathsf{E}, \mathcal{E})$, such that P is μ-irreducible.

Definition A.3. *A set $\mathtt{A} \in \mathcal{E}$ is said accessible for P if for all $x \in \mathsf{E}$, there exists $n \in \mathbb{N}^*$ such that $P^n(x, A) > 0$.*

Definition A.4. *We say that a σ-finite measure $\mu \in \mathcal{P}(\mathsf{E})$ is a maximal irreducibility measure for P if P is μ-irreducible and $\mathtt{A}$ is accessible if and only if $\mu(\mathtt{A}) > 0$.*

Definition A.5. *Let $(a_k)_{k\in\mathbb{N}^*}$ be a sequence of nonnegative real numbers such that $\sum_{k=0}^{+\infty} a_k = 1$. Then the sampled kernel associated with $(a_k)_{k\in\mathbb{N}^*}$ is the Markov kernel defined for all $x \in \mathsf{E}$ and $\mathtt{A} \in \mathcal{E}$ by*

$$K_a(x, \mathtt{A}) = \sum_{k=0}^{+\infty} a_k P^k(x, \mathtt{A}) .$$

Let $\epsilon \in (0, 1)$, we denote by K_ϵ the sample chain associated with the geometric sequence $a_n = \epsilon^n/(1 - \epsilon)$ for all $n \geq 0$.

Theorem A.6 ([MT09, Proposition 4.2.2])**.** *Let P be an irreducible chain, with irreducibility measure μ. Then μK_ϵ is a maximal irreducibility measure for all $\epsilon \in (0, 1)$. In addition if μ_1 is a maximal irreducibility measure, for any irreducibility measure μ_2 satisfies $\mu_2 << \mu_1$.*

Petite and small sets

Definition A.7. *$\mathcal{C}$ est un $\mathtt{n}$-small set pour P si ils existent $\mathtt{n} \in \mathbb{N}^*$, $\epsilon > 0$ et une mesure de probabilité ν sur $(\mathsf{E}, \mathcal{E})$ tels que pour tout $x \in \mathcal{C}$, $P^{\mathtt{n}}(x, \dots) \geq \epsilon\nu(\cdot)$.*

Definition A.8. *A set $\mathcal{C} \in \mathcal{E}$ is said to be a petite set for P if there exist a sequence of nonnegative real numbers $(a_k)_{k\in\mathbb{N}^*}$ such that $\sum_{k=0}^{+\infty} a_k = 1$, a non trivial σ-finite positive measure μ on $(\mathsf{E}, \mathcal{E})$ such that for all $x \in \mathcal{C}$*

$$K_a(x, \mathtt{A}) \geq \mu(\mathtt{A}) , \qquad \textit{for all } \mathtt{A} \in \mathcal{E} ,$$

where K_a is the sample kernel associated with P and $(a_k)_{k\in\mathbb{N}^*}$. If such relation holds then $\mathcal{C}$ is said to be a (μ, a)-petite set.

Recurrence and transience

Define for all $\mathtt{A} \in \mathcal{E}$

$$N_A = \sum_{i=0}^{+\infty} \mathbb{1}_{\mathtt{A}}(Y_i) \ , U(x, \mathtt{A}) = \mathbb{E}_x \left[N_A \right] \ .$$

Definition A.9. *a) A set $\mathtt{A} \in \mathcal{E}$ is said to be recurrent if for all $x \in \mathtt{A}$, $U(x, \mathtt{A}) = +\infty$.*

b) P is recurrent if all accessible sets are recurrent.

Definition A.10. *a) A set $\mathtt{A} \in \mathcal{E}$ is said to be uniformly transient if for all $x \in \mathtt{A}$, $U(x, \mathtt{A}) < +\infty$.*

b) A set $\mathtt{A} \in \mathcal{E}$ is transient if there exists a countable sequence $(\mathtt{A}_n)_{n \in \mathbb{N}}$ of uniformly transient sets such that $\mathtt{A} = bigcup_{n=0}^{+\infty} \mathtt{A}_n$.

c) P is transient if E is transient.

Theorem A.11 ([MT09, Theorem 8.3.4])**.** *If P is irreducible. Then it is either transient or recurrent.*

Proposition A.12 ([MT09, Proposition 10.1.1])**.** *Assume that P is irreducible and admits a invariant distribution π. Then it is recurrent.*

Define for all $\mathtt{A} \in \mathcal{E}$, the following stopping time:

$$\boldsymbol{\sigma}_{\mathtt{A}} = \inf \left\{ k \geq 1 \,|\, Y_k \in \mathtt{A} \right\} \ .$$

Theorem A.13 ([MT09, Theorem 10.4.9, Theorem 10.4.10])**.** *Let P be an irreducible and recurrent Markov kernel on $(\mathsf{E}, \mathcal{E})$.*

1. *Then there exists a unique (up to multiplicative constant) invariant measure $\tilde{\pi}$ on $(\mathsf{E}, \mathcal{E})$ for P, which is a maximal irreducibility measure and satisfies for any accessible set $\mathtt{A} \in \mathcal{E}$ and any measurable set $\mathtt{B} \in \mathcal{E}$,*

$$\tilde{\pi}(\mathtt{B}) = \int_{\mathtt{A}} \mathbb{E}_y \left[\sum_{k=1}^{\boldsymbol{\sigma}_A} \mathbb{1}_B(Y_k) \right] \tilde{\pi}(\mathrm{d}y) \ .$$

2. *$\tilde{\pi}(\mathsf{E}) < +\infty$ if there exists an accessible petite set $\mathcal{C}$ such that*

$$\sup_{x \in \mathcal{C}} \mathbb{E}_x \left[\boldsymbol{\sigma}_{\mathcal{C}} \right] < +\infty \ .$$

In addition, P is Harris recurrent if for all $x \in \mathsf{E}$,

$$\mathbb{E}_x \left[\boldsymbol{\sigma}_{\mathcal{C}} \right] < +\infty \ .$$

Definition A.14. *a) A set $\mathtt{A} \in \mathcal{E}$ is said to be Harris recurrent if for all $x \in \mathtt{A}$, $\mathbb{P}_x \left[N_A = +\infty \right] = 1$.*

b) P is Harris recurrent if all accessible sets are Harris recurrent.

Theorem A.15 ([MT09, Theorem 9.1.5])**.** *Let P be irreducible with maximal irreducible measure μ, and recurrent on $(\mathsf{E}, \mathcal{E})$. Then there exists a partition of $\mathsf{E} = \mathsf{H} \cup \mathsf{N}$ such that*

(i) $\mathsf{H} \in \mathcal{E}$ is a non-empty absorbing set and P restricted to H is Harris recurrent.

(ii) $\mathsf{N} \in \mathcal{E}$ is transient and $\mu(\mathsf{N}) = 0$.

Periodicity and aperiodicity

Definition A.16. *a) P is said to be periodic with period $\mathsf{d} \in \mathbb{N}^*$, if there exists a partition of E, $\{\mathtt{A}_i \in \mathcal{E} \,|1 \leq i \leq N\}$, $N \in \mathbb{N}^*$, such that $\mathtt{A}_1 = \mathtt{A}_N$, and for all $i \in \{1, \cdots, N-1\}$, for all $x \in \mathtt{A}_i$, $P(x, \mathtt{A}_{i+1}) = 1$. P is said to be aperiodic if it is periodic with period 1.*

b) P is said to be strongly aperiodic if there exists a small set $(1, \mu)$-$\mathcal{C}$ for P such that $\mu(\mathcal{C}) > 0$.

Theorem A.17 ([MT09, Theorem 5.5.7])**.** *If P is irreducible and aperiodic, then all petite sets are small.*

Dynkin's formula and consequences

Theorem A.18 (Dynkin's formula, [MT09, Theorem 11.3.1])**.** *Let $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\in\mathbb{N}}, \mathbb{P})$ be a filtered probability space. Let $(X_k)_{k\in\mathbb{N}}$ be a bounded and $(\mathcal{F}_k)_{k\in\mathbb{N}}$-adapted sequence of random variables and let τ be a bounded stopping time. Then, the following identity holds*

$$\mathbb{E}\left[X_\tau\right] - \mathbb{E}\left[X_0\right] = \mathbb{E}\left[\sum_{k=1}^{\tau} \left\{\mathbb{E}\left[X_k \,|\mathcal{F}_{k-1}\right] - X_{k-1}\right\}\right] .$$

Corollary A.19. *Let P be a Markov kernel on $(\mathsf{E}, \mathcal{E})$ and $(Y_k)_{k\in\mathbb{N}}$ be the corresponding canonical chain. Then for all bounded measurable function $f : \mathsf{E} \to \mathbb{R}$ and bounded stopping time τ,*

$$\mathbb{E}\left[f(Y_\tau)\right] + \mathbb{E}\left[\sum_{k=1}^{\tau} f(Y_{k-1})\right] = \mathbb{E}\left[f(Y_0)\right] + \mathbb{E}\left[\sum_{k=1}^{\tau} Pf(Y_{k-1})\right] .$$

Corollary A.20. *Assume there exist a measurable function $V : \mathsf{E} \to \bar{\mathbb{R}}_+$, $b \in \mathbb{R}_+$ and $\mathcal{C} \in \mathcal{E}$ such that for all $x \in \mathsf{E}$,*

$$PV(x) \leq V(x) - 1 + b\mathbb{1}_{\mathcal{C}}(x) .$$

Then for all $x \in \mathsf{E}$,

$$\mathbb{E}_x\left[\boldsymbol{\sigma}_{\mathcal{C}}\right] \leq V(x) + b\mathbb{1}_{\mathcal{C}}(x) .$$

In particular in the case where $\mathcal{C}$ is petite and P is irreducible and recurrent, we have

(i) If $\sup_{x\in\mathcal{C}} V(x) < +\infty$, P has a unique invariant distribution.

(ii) If in addition for all $x \in \mathsf{E}$, $V(x) < +\infty$, P is Harris recurrent.

Feller chains

We assume here that E is a locally compact metric space and $\mathcal{E}$ is its Borel σ-field.

Definition A.21. *P is said to be Feller if for all $f \in C_b(\mathsf{E})$, then $Pf \in C_b(\mathsf{E})$ as well.*

Proposition A.22. *Let P be a Feller irreducible chain on $(\mathsf{E}, \mathcal{E})$. Let μ be a maximal irreducibility measure for P. If the support of μ has non empty interior then all compact sets of E are petite sets.*

A.2 Markov processes in continuous time

Let E be a locally compact Polish space, and $\mathcal{E} = \mathcal{B}(\mathsf{E})$ its Borel σ-algebra.

Definition A.23. *$(\mathbf{P}_t)_{t\geq 0}$ is a Markov semi-group (or transition function) on $(\mathsf{E}, \mathcal{E})$ if it is a family of Markov kernels such that for all $s, t \geq 0$, $\mathbf{P}_s\mathbf{P}_t = \mathbf{P}_{s+t}$.*

Definition A.24. *A function $f \in C(\mathsf{E})$ is said to vanish at infinity if for all $a \geq 0$, there exists a compact set $\mathcal{K} \subset \mathsf{E}$ such that for all $x \notin \mathcal{K}$, $|f(x)| \leq a$. The set of functions vanishing at infinity is denoted by $C_0(\mathsf{E})$.*

Definition A.25. *Let $(\mathbf{P}_t)_{t\geq 0}$ be a Markov semi-group on $(\mathsf{E}, \mathcal{E})$. $\mathbf{P}$ is said to be a Feller semi-group if for all $f \in C_0(\mathsf{E})$, for all $t \geq 0$ $\mathbf{P}_t f \in C_0(\mathsf{E})$, and for all $x \in \mathsf{E}$,*

$$\lim_{t\to 0} \mathbf{P}_t f(x) = f(x) \ .$$

Theorem A.26 ([RY99, Theorem 1.5, Theorem 2.7, Chapter III])**.** *Let $(\mathbf{P}_t)_{t\geq 0}$ be a Markov semi-group on $(\mathsf{E}, \mathcal{E})$ and $\nu \in \mathcal{P}(\mathsf{E})$, then there exists a unique probability measure $\mathbb{P}_\nu$ on the canonical space $(\mathsf{E}^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ such that the canonical process $(Y_t)_{t\geq 0}$ is a Markov process associated with $(\mathbf{P}_t)_{t\geq 0}$ and initial distribution ν, with respect to $(\mathcal{F}_t^Y)_{t\in 0}$ the corresponding filtration defined by for all $t \geq 0$ $\mathcal{F}_t^Y = \sigma(Y_s, s \in [0, t])$.*

In addition if $(\mathbf{P}_t)_{t\geq 0}$ is Feller, then $(Y_t)_{t\geq 0}$ admits a cadlag modification.

Definition A.27. *A measurable set $\mathtt{A} \in \mathcal{E}$ is said to be accessible for the Markov semi-group $(\mathbf{P}_t)_{t\geq 0}$ if for all $x \in \mathsf{E}$, there exists $t \geq 0$ such that $\mathbf{P}_t(x, \mathtt{A}) > 0$.*

Definition A.28. *Let $(\mathbf{P}_t)_{t\geq 0}$ be a Markov semi-group on $(\mathsf{E}, \mathcal{E})$.*

(a) A non-trivial σ-finite measure ν on $(\mathsf{E}, \mathcal{E})$ is said to be an ν-irreducibility measure for $(\mathbf{P}_t)_{t\geq 0}$ if for all $\mathtt{A} \in \mathcal{E}$ satisfying $\nu(\mathtt{A}) > 0$ is accessible for $(\mathbf{P}_t)_{t\geq 0}$.

(b) $(\mathbf{P}_t)_{t\geq 0}$ is irreducible is there exists a non-trivial σ-finite measure ν on $(\mathsf{E}, \mathcal{E})$ such that $(\mathbf{P}_t)_{t\geq 0}$ is ν irreducible.

Define for all $\mathtt{A} \in \mathcal{E}$,

$$N_A = \int_0^{+\infty} \mathbb{1}_A(Y_s)\mathrm{d}s \ , \qquad U_A(x) = \mathbb{E}_x\left[N_A\right] \ .$$

Definition A.29. *Let $(\mathbf{P}_t)_{t\geq 0}$ be a Markov semi-group on $(\mathsf{E}, \mathcal{E})$. A set $\mathtt{A} \in \mathcal{E}$ is said to be recurrent if for all $x \in \mathtt{A}$, $U_A(x) = +\infty$. $(\mathbf{P}_t)_{t\geq 0}$ is said to be recurrent if any accessible set is recurrent.*

Definition A.30. *Let $(\mathbf{P}_t)_{t\geq 0}$ be a Markov semi-group on $(\mathsf{E}, \mathcal{E})$. A set $\mathtt{A} \in \mathcal{E}$ is said to be Harris recurrent if for all $x \in \mathtt{A}$, $\mathbb{P}_x[N_A = +\infty] = 1$. $(\mathbf{P}_t)_{t\geq 0}$ is said to be Harris recurrent if any accessible set $\mathtt{A}$ is Harris recurrent.*

A.3 Results on diffusions

Let $b : \mathbb{R}^d \to \mathbb{R}$ and σ be a function from $\mathbb{R}^d$ to $\mathcal{S}_+^*(\mathbb{R}^d)$. We assume that these two functions are locally Lipschitz. The the associated SDE is defined by:

$$\mathrm{d}\mathbf{Y}_t = b(\mathbf{Y}_t)\mathrm{d}t + \sigma(\mathbf{Y}_t)\mathrm{d}B_t^d \,. \tag{A.1}$$

By [IW89, Theorem 2.3, Theorem 3.1, Chapter 4], for all initial condition $x \in \mathbb{R}^d$, this SDE admits a unique solution $(\mathbf{Y}_t)_{t\in 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_x)$ up to the time ξ which is the stopping time defined by

$$\xi = \inf \{ t \geq 0 \,|\, \mathbf{Y}_t = \infty \} \;.$$

We assume that almost surely $\xi = +\infty$. We denote by $(\mathbf{P}_t)_{t\in 0}$ the associated Markov semi-group.

Definition A.31. *(a) We say that a point $x \in \mathbb{R}^d$ is recurrent if for all $\epsilon > 0$, $\mathbb{P}_x[N_{\mathrm{B}(x,\epsilon)} = +\infty] = 1$.*

(b) We say that a point $x \in \mathbb{R}^d$ is transient if $\mathbb{P}_x[\lim_{t\to+\infty} \|\mathbf{Y}_t\| = +\infty] = 1$.

Proposition A.32 ([Bha78, Proposition 3.1])**.** *$(\mathbf{P}_t)_{t\in 0}$ is Harris recurrent if and only if any point $x \in \mathbb{R}^d$ is recurrent.*

Theorem A.33 ([Bha78, Theorem 3.2])**.** *(i) If there exists a recurrent point for $(\mathbf{P}_t)_{t\in 0}$, then $(\mathbf{P}_t)_{t\in 0}$ is Harris recurrent.*

(ii) If no recurrent point exists, then any point $x \in \mathbb{R}^d$ is transient. We say that the diffusion is transient.

Corollary A.34. *If $(\mathbf{P}_t)_{t\in 0}$ is recurrent[1] then it is Harris recurrent.*

[1] Definition A.29

Appendix B

Supplement to "Sampling from a strongly log-concave distribution with the Unadjusted Langevin Algorithm"

ALAIN DURMUS[1], ÉRIC MOULINES [2]

B.1 Discussion of Theorem 6.4

Note that

$$u_n^{(2)}(\gamma) \leq \sum_{i=1}^{n} \left\{ \mathsf{A}_0 \gamma_i^2 + \mathsf{A}_1 \gamma_i^3 \right\} \prod_{k=i+1}^{n} (1 - \kappa \gamma_k / 2) \;, \tag{B.1}$$

where κ is given by (6.4), and

$$\mathsf{A}_0 \stackrel{\text{def}}{=} 2L^2 \kappa^{-1} d \;, \tag{B.2}$$

$$\mathsf{A}_1 \stackrel{\text{def}}{=} 2L^2 \kappa^{-1} d (m+L)^{-1} + d L^4 (\kappa^{-1} + (m+L)^{-1})(m^{-1} + 6^{-1}(m+L)^{-1}) \;. \tag{B.3}$$

If $(\gamma_k)_{k \geq 1}$ is a constant step size ,$\gamma_k = \gamma$ for all $k \geq 1$, then a straightforward consequence of Theorem 6.4 and (B.1) if the following result, which gives the minimal number of iterations n and a step-size γ to get $W_2(\delta_{x^\star} Q_\gamma^n, \pi)$ smaller than $\epsilon > 0$.

Corollary B.1 (of Theorem 6.4). *Assume* ***L1*** *and* ***H11****. Let $x^\star$ be the unique minimizer*

[1] LTCI, Telecom ParisTech 46 rue Barrault, 75634 Paris Cedex 13, France. alain.durmus@telecom-paristech.fr

[2] Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France. eric.moulines@polytechnique.edu

of U. Let $x \in \mathbb{R}^d$ and $\epsilon > 0$. Set for all $k \in \mathbb{N}$, $\gamma_k = \gamma$ with

$$\gamma = \frac{-2\mathsf{A}_0\kappa^{-1} + (4\kappa^{-1}(\mathsf{A}_0^2\kappa^{-1} + \epsilon^2\mathsf{A}_1))^{1/2}}{4\mathsf{A}_1\kappa^{-1}} \wedge (m+L)^{-1} \; , \qquad \text{(B.4)}$$
$$n = \left\lceil 2(\kappa\gamma)^{-1}\left\{-\log(\epsilon^2/4) + \log(2d/m)\right\}\right\rceil \; .$$

Then $W_2(\delta_{x^\star}Q_n^\gamma, \pi) \leq \epsilon$.

Note that if γ is given by (B.4), and is different from $1/(m+L)$, then $\gamma \leq \epsilon(4\mathsf{A}_1\kappa^{-1})^{-1/2}$ and $2\kappa^{-1}(\mathsf{A}_0\gamma + \mathsf{A}_1\gamma^2) = \epsilon^2/2$. Therefore,

$$\gamma \geq (\epsilon^2\kappa/4)\left\{\mathsf{A}_0 + \epsilon(\mathsf{A}_1/(4\kappa))^{1/2}\right\}^{-1} \; .$$

It is shown in [Dal16, Corollary 1] that under **H**11, for constant step size for any $\epsilon > 0$, we can choose γ and $n \geq 1$ such that if for all $k \geq 1$, $\gamma_k = \gamma$, then $\|\nu^\star Q_n^\gamma - \pi\|_{\mathrm{TV}} \leq \epsilon$ where $\nu^\star$ is the Gaussian measure on $\mathbb{R}^d$ with mean $x^\star$ and covariance matrix $L^{-1}\,\mathrm{I}_d$. We stress that the results in [Dal16, corollary 1] hold only for a particular choice of the initial distribution $\nu^\star$, (which might seem a rather artificial assumption) whereas Theorem 6.4 hold for any initial distribution in $\mathcal{P}_2(\mathbb{R}^d)$.

We compare the optimal value of γ and n obtained from Corollary B.1 with those given in [Dal16, Corollary 1]. This comparison is summarized in Table B.1 and Table B.2; for simplicity, we provide only the dependencies of the optimal stepsize γ and minimal number of simulations n as a function of the dimension d, the precision ϵ and the constants m, L. It can be seen that the dependency on the dimension is significantly better than those in [Dal16, Corollary 1].

Parameter	d	ϵ	L	m
Theorem 6.4 and (6.20)	d^{-1}	ϵ^2	L^{-2}	m^2
[Dal16, Corollary 1]	d^{-2}	ϵ^2	L^{-2}	m

Table B.1: Dependencies of γ

Parameter	d	ϵ	L	m
Theorem 6.4 and (6.20)	$d\log(d)$	$\epsilon^{-2}\,\lvert\log(\epsilon)\rvert$	L^2	$\lvert\log(m)\rvert\, m^{-3}$
[Dal16, Corollary 1]	d^3	$\epsilon^{-2}\,\lvert\log(\epsilon)\rvert$	L^3	$\lvert\log(m)\rvert\, m^{-2}$

Table B.2: Dependencies of n

B.1.1 Explicit bounds for $\gamma_k = \gamma_1 k^\alpha$ with $\alpha \in (0,1]$

We give here a bound on the sequences $(u_n^{(1)}(\gamma))_{n\geq 0}$ and $(u_n^{(2)}(\gamma))_{n\geq 0}$ for $(\gamma_k)_{k\geq 1}$ defined by $\gamma_1 < 1/(m+L)$ and $\gamma_k = \gamma_1 k^{-\alpha}$ for $\alpha \in (0,1]$. Also for that purpose we introduce for $t \in \mathbb{R}_+^*$,

$$\boldsymbol{\psi}_\beta(t) = \begin{cases} (t^\beta - 1)/\beta & \text{for } \beta \neq 0 \\ \log(t) & \text{for } \beta = 0 \; . \end{cases} \qquad \text{(B.5)}$$

We easily get for $a \geq 0$ that for all $n, p \geq 1$, $n \leq p$

$$\boldsymbol{\psi}_{1-a}(p+1) - \boldsymbol{\psi}_{1-a}(n) \leq \sum_{k=n}^{p} k^{-a} \leq \boldsymbol{\psi}_{1-a}(p) - \boldsymbol{\psi}_{1-a}(n) + 1 \;, \tag{B.6}$$

and for $a \in \mathbb{R}$

$$\sum_{k=n}^{p} k^{-a} \leq \boldsymbol{\psi}_{1-a}(p+1) - \boldsymbol{\psi}_{1-a}(n) + 1 \;. \tag{B.7}$$

1. For $\alpha = 1$, using that for all $t \in \mathbb{R}$, $(1+t) \leq e^t$ and by (B.6) and (B.7), we have

$$u_n^{(1)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \;,\; u_n^{(2)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \sum_{j=0}^{1} \mathsf{A}_j(\boldsymbol{\psi}_{\kappa\gamma_1/2-1-j}(n+1)+1) \;.$$

2. For $\alpha \in (0,1)$, by (B.6) and Lemma 6.24 applied with $\ell = \lceil n/2 \rceil$, where $\lceil \cdot \rceil$ is the ceiling function, we have

$$\begin{aligned} u_n^{(1)}(\gamma) &\leq \exp\left(-\kappa\gamma_1\boldsymbol{\psi}_{1-\alpha}(n+1)/2\right) \\ u_n^{(2)}(\gamma) &\leq \sum_{j=0}^{1} \mathsf{A}_j \left(2\kappa^{-1}\gamma_1^{j+1}(n/2)^{-\alpha(j+1)} + \gamma_1^{j+2}\left(\boldsymbol{\psi}_{1-\alpha(j+2)}(\lceil n/2 \rceil) + 1\right) \right. \\ &\qquad \left. \times \exp\left\{-(\kappa\gamma_1/2)(\boldsymbol{\psi}_{1-\alpha}(n+1) - \boldsymbol{\psi}_{1-\alpha}(\lceil n/2 \rceil))\right\}\right) \;. \end{aligned} \tag{B.8}$$

B.1.2 Optimal strategy with a fixed number of iterations

Corollary B.2. *Let $n \in \mathbb{N}^*$ be a fixed number of iteration. Assume* ***L*** *1,* ***H*** *11, and $(\gamma_k)_{k\geq 1}$ is a constant sequence, $\gamma_k = \gamma$ for all $k \geq 1$. Set*

$$\begin{aligned} \gamma^+ &\overset{\text{def}}{=} 2(\kappa n)^{-1}(\log(\kappa n/2) + \log(2(\|x - x^\star\|^2 + d/m)) - \log(2\kappa^{-1}\mathsf{A}_0)) \\ \gamma_- &\overset{\text{def}}{=} 2(\kappa n)^{-1}(\log(\kappa n/2) + \log(2(\|x - x^\star\|^2 + d/m)) - \log(2\kappa^{-1}(\mathsf{A}_0 + 2\mathsf{A}_1(m+L)^{-1}))) \;. \end{aligned}$$

Assume $\gamma^+ \in (0, (m+L)^{-1})$. Then, the optimal choice of γ to minimize the bound on $W_2(\delta_x Q_n^\gamma, \pi)$ given by Theorem 6.4 belongs to $[\gamma_-, \gamma^+]$. Moreover if $\gamma = \gamma_+$, then the bound on $W_2^2(\delta_x Q_n^\gamma, \pi)$ is equivalent to $4\mathsf{A}_0(\kappa^2 n)^{-1}\log(\kappa n/2)$ as $n \to +\infty$.

Similarly, we have the following result.

Corollary B.3. *Assume* ***L*** *1 and* ***H*** *11. Let $(\gamma_k)_{k\geq 1}$ be the decreasing sequence, defined by $\gamma_k = \gamma_\alpha k^{-\alpha}$, with $\alpha \in (0,1)$. Let $n \geq 1$ and set*

$$\gamma_\alpha \overset{\text{def}}{=} 2(1-\alpha)\kappa^{-1}(2/n)^{1-\alpha}\log(\kappa n/(2(1-\alpha))) \;.$$

Assume $\gamma_\alpha \in (0, (m+L)^{-1})$. Then the bound on $W_2^2(\delta_x Q_n^\gamma, \pi)$ given by Theorem 6.4 is smaller for large n than $8(1-\alpha)\mathsf{A}_0(\kappa^2 n)^{-1}\log(\kappa n/(2(1-\alpha)))$.

Proof. Follows from (B.1), (B.8) and the choice of γ_α. □

B.2 Discussion of Theorem 6.7

Based on Theorem 6.7, we can follow the same discussion than for Theorem 6.4. Note that

$$u_n^{(3)}(\gamma) \le \sum_{i=1}^{n} \left\{ \mathsf{B}_0 \gamma_i^3 + \mathsf{B}_1 \gamma_i^4 \right\} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) \;, \tag{B.9}$$

where κ is given by (6.4), and

$$\mathsf{B}_0 \stackrel{\text{def}}{=} d\left(2L^2 + \kappa^{-1}(\tilde{L}^2/3 + 4L^4/(3m))\right) \;, \tag{B.10}$$

$$\mathsf{B}_1 \stackrel{\text{def}}{=} d\left(\kappa^{-1}L^4 + L^4(1/(6(m+L)) + m^{-1})\right) \;. \tag{B.11}$$

The following result gives the minimal number of iterations n and a step-size γ to get $W_2(\delta_{x^\star} Q_\gamma^n, \pi)$ smaller than $\epsilon > 0$, when $(\gamma_k)_{k\ge1}$ is a constant step size, $\gamma_k = \gamma$ for all $k \ge 1$.

Corollary B.4 (of Theorem 6.7)**.** *Assume* $\boldsymbol{L}$ *1,* $\boldsymbol{H}$ *11 and* $\boldsymbol{H}$ *12. Let* $x^\star$ *be the unique minimizer of* U*. Let* $x \in \mathbb{R}^d$ *and* $\epsilon > 0$*. Set for all* $k \in \mathbb{N}$*,* $\gamma_k = \gamma$ *with*

$$\gamma = (\epsilon/2)\kappa^{-1}(\mathsf{B}_0 + \mathsf{B}_1(m+L)^{-1})^{-1/2} \;, \tag{B.12}$$

$$n = \left\lceil 2(\kappa\gamma)^{-1} \left\{ -\log(\epsilon^2/4) + \log(2d/m) \right\} \right\rceil \;. \tag{B.13}$$

Then $W_2(\delta_{x^\star} Q_n^\gamma, \pi) \le \epsilon$*.*

We provide only the dependencies of the optimal stepsize γ and minimal number of simulations n as a function of the dimension d, the precision ϵ and the constants $m, L, \tilde{L}$ in Table B.3 and Table B.4.

Parameter	d	ϵ	L	m
Theorem 6.7 and (6.20)	$d^{-1/2}$	ϵ	$L^{-7/2}$	m^2

Table B.3: Dependencies of γ

Parameter	d	ϵ	L	m
Theorem 6.7 and (6.20)	$d^{1/2}\log(d)$	$\epsilon^{-1}\left\|\log(\epsilon)\right\|$	$L^{9/2}$	$\left\|\log(m)\right\| m^{-3}$

Table B.4: Dependencies of n

B.2.1 Explicit bounds for $\gamma_k = \gamma_1 k^\alpha$ with $\alpha \in (0,1]$

We give here a bound on the sequence $(u_n^{(3)}(\gamma))_{n\ge0}$ for $(\gamma_k)_{k\ge1}$ defined by $\gamma_1 < 1/(m+L)$ and $\gamma_k = \gamma_1 k^{-\alpha}$ for $\alpha \in (0,1]$. Bounds for $(u_n^{(1)}(\gamma))_{n\ge0}$ have already been given in Appendix B.1.1. Recall that the function $\boldsymbol{\psi}$ is defined by (B.5).

1. For $\alpha = 1$, using that for all $t \in \mathbb{R}$, $(1+t) \leq e^t$ and by (B.6) and (B.7), we have

$$u_n^{(3)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \sum_{j=1}^{2} \mathsf{B}_{j-1}(\boldsymbol{\psi}_{\kappa\gamma_1/2-1-j}(n+1)+1) \; .$$

2. For $\alpha \in (0,1)$, by (B.6) and Lemma 6.24 applied with $\ell = \lceil n/2 \rceil$, where $\lceil \cdot \rceil$ is the ceiling function, we have

$$u_n^{(3)}(\gamma) \leq \sum_{j=1}^{2} \mathsf{B}_{j-1} \left(2\kappa^{-1}\gamma_1^{j+1}(n/2)^{-\alpha(j+1)} + \gamma_1^{j+2} \left(\boldsymbol{\psi}_{1-\alpha(j+2)}(\lceil n/2 \rceil) + 1 \right) \right. \\ \left. \times \exp\{-(\kappa\gamma_1/2)(\boldsymbol{\psi}_{1-\alpha}(n+1) - \boldsymbol{\psi}_{1-\alpha}(\lceil n/2 \rceil))\} \right) \; . \tag{B.14}$$

B.2.2 Optimal strategy with a fixed number of iterations

Corollary B.5. *Let $n \in \mathbb{N}^*$ be a fixed number of iteration. Assume* ***L** 1*, ***H** 11*, ***H** 12* and $(\gamma_k)_{k\geq 1}$ is a constant sequence, $\gamma_k = \gamma^\star$ for all $k \geq 1$, with*

$$\gamma^\star = 4(\kappa n)^{-1} \left\{ \log(\kappa n/2) + \log(2(\|x - x^\star\|^2 + d/m)) \right\} \; .$$

Assume $\gamma^\star \in (0, (m+L)^{-1})$. Then the bound on $W_2^2(\delta_x Q_n^\gamma, \pi)$ given by Theorem 6.7 is of order $\mathcal{O}(\mathsf{B}_0(\kappa^2 n)^{-2}\log^2(n))$ as $n \to +\infty$.

Similarly, we have the following result.

Corollary B.6. *Assume* ***L** 1*, ***H** 11* and ***H** 12*. Let $(\gamma_k)_{k\geq 1}$ be the decreasing sequence, defined by $\gamma_k = \gamma_\alpha k^{-\alpha}$, with $\alpha \in (0,1)$. Let $n \geq 1$ and set*

$$\gamma_\alpha \stackrel{\text{def}}{=} 2(1-\alpha)\kappa^{-1}(2/n)^{1-\alpha}\log(\kappa n/(2(1-\alpha))) \; .$$

Assume $\gamma_\alpha \in (0, (m+L)^{-1})$. Then the bound on $W_2^2(\delta_x Q_n^\gamma, \pi)$ given by Theorem 6.4 is smaller for large n than $\mathcal{O}(\mathsf{B}_0(\kappa^2 n)^{-2}\log^2(n))$.

Proof. Follows from (B.9), (B.14) and the choice of γ_α. □

B.3 Generalization of Theorem 6.4

In this section, we weaken the assumption $\gamma_1 \leq 1/(m+L)$ of Theorem 6.4. We assume now:

G 6. *The sequence $(\gamma_k)_{k\geq 1}$ is non-increasing, and there exists $\rho > 0$ and n_1 such that $(1+\rho)\gamma_{n_1} \leq 2/(m+L)$.*

Under **G**6, we denote by

$$n_0 \stackrel{\text{def}}{=} \min\{k \in \mathbb{N} \mid \gamma_k \leq 2/(m+L)\} \tag{B.15}$$

We first give an extension of Theorem 6.2. Denote in the sequel $(\cdot)_+ = \max(\cdot, 0)$. Recall that under **H**11, $x^\star$ is the unique minimizer of U, and κ is defined in (6.6)

Theorem B.7. *Assume **L**1, **H**11 and **G**6. Then for all $n, p \in \mathbb{N}^*$, $n \leq p$*

$$\int_{\mathbb{R}^d} \|x - x^\star\|^2 \, \mu_0 Q_n^p(\mathrm{d}x) \leq E_{n,p}(\mu_0, \gamma) \; ,$$

where

$$E_{n,p}(\mu_0,\gamma) \stackrel{\text{def}}{=} \exp\left(-\sum_{k=n}^{p} \gamma_k \kappa + \sum_{k=n}^{n_0-1} L^2\gamma_k^2\right) \int_{\mathbb{R}^d} \|x - x^\star\|^2 \, \mu_0(\mathrm{d}x)$$
$$+ 2d\kappa^{-1} + 2d\left\{\prod_{k=n}^{n_0-1} (\gamma_{n_0-1}L^2)^{-1}\left(1 + L^2\gamma_k^2\right)\right\} \exp\left(-\sum_{k=n}^{p} \kappa\gamma_k + \sum_{k=n}^{n_0-1} \gamma_k^2 m L\right) \; . \tag{B.16}$$

Proof. For any $\gamma > 0$, we have for all $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} \|y - x^\star\|^2 \, R_\gamma(x, \mathrm{d}y) = \|x - \gamma\nabla U(x) - x^\star\|^2 + 2\gamma d \; .$$

Using that $\nabla U(x^\star) = 0$, (6.4) and **L**1, we get from the previous inequality:

$$\begin{aligned}
&\int_{\mathbb{R}^d} \|y - x^\star\|^2 R_\gamma(x, \mathrm{d}y) \\
&\qquad \leq (1 - \kappa\gamma) \|x - x^\star\|^2 + \gamma\left(\gamma - \frac{2}{m+L}\right) \|\nabla U(x) - \nabla U(x^\star)\|^2 + 2\gamma d \\
&\qquad \leq \eta(\gamma) \|x - x^\star\|^2 + 2\gamma d \; ,
\end{aligned}$$

where $\eta(\gamma) = (1 - \kappa\gamma + \gamma L(\gamma - 2/(m+L))_+)$. Denote for all $k \geq 1$, $\eta_k = \eta(\gamma_k)$. By a straightforward induction, we have by definition of Q_n^p for $p, n \in \mathbb{N}$, $p \leq n$,

$$\int_{\mathbb{R}^d} \|x - x^\star\|^2 \, \mu_0 Q_n^p(\mathrm{d}x) \leq \prod_{k=n}^{p} \eta_k \int_{\mathbb{R}^d} \|x - x^\star\| \, \mu_0(\mathrm{d}x) + (2d) \sum_{i=n}^{p} \prod_{k=i+1}^{p} \eta_k \gamma_i \; , \tag{B.17}$$

with the convention that for $n, p \in \mathbb{N}$, $n < p$, $\prod_p^n = 1$. For the first term of the right hand side, we simply use the bound, for all $x \in \mathbb{R}$, $(1 + x) \leq e^x$, and we get by **G**6

$$\prod_{k=n}^{p} \eta_k \leq \exp\left(-\sum_{k=n}^{p} \kappa\gamma_k + \sum_{k=n}^{n_0-1} L^2\gamma_k^2\right) \; , \tag{B.18}$$

where n_0 is defined in (B.15). Consider now the second term in the right hand side of (B.17).

$$\begin{aligned}\sum_{i=n}^{p}\prod_{k=i+1}^{p}\eta_k\gamma_i &\leq \sum_{i=n_0}^{p}\prod_{k=i+1}^{p}(1-\kappa\gamma_k)\,\gamma_i+\sum_{i=n}^{n_0-1}\prod_{k=i+1}^{p}\eta_k\gamma_i \\ &\leq \kappa^{-1}\sum_{i=n_0}^{p}\left\{\prod_{k=i+1}^{p}(1-\kappa\gamma_k)-\prod_{k=i}^{p}(1-\kappa\gamma_k)\right\} \\ &+\left\{\sum_{i=n}^{n_0-1}\prod_{k=i+1}^{n_0-1}\left(1+L^2\gamma_k^2\right)\gamma_i\right\}\prod_{k=n_0}^{p}(1-\kappa\gamma_k)\end{aligned} \tag{B.19}$$

Since $(\gamma_k)_{k\geq1}$ is nonincreasing, we have

$$\begin{aligned}\sum_{i=n}^{n_0-1}\prod_{k=i+1}^{n_0-1}\left(1+L^2\gamma_k^2\right)\gamma_i &= \sum_{i=n}^{n_0-1}(\gamma_i L^2)^{-1}\left\{\prod_{k=i}^{n_0-1}\left(1+L^2\gamma_k^2\right)-\prod_{k=i+1}^{n_0-1}\left(1+L^2\gamma_k^2\right)\right\} \\ &\leq \prod_{k=n}^{n_0-1}(\gamma_{n_0-1}L^2)^{-1}\left(1+L^2\gamma_k^2\right) .\end{aligned}$$

Furthermore for $k<n_0$ $\gamma_k>2/(m+L)$. This implies with the bound $(1+x)\leq e^x$ on $\mathbb{R}$:

$$\begin{aligned}\prod_{k=n_0}^{p}(1-\kappa\gamma_k) &\leq \exp\left(-\sum_{k=n}^{p}\kappa\gamma_k\right)\exp\left(\sum_{k=n}^{n_0-1}\kappa\gamma_k\right) \\ &\leq \exp\left(-\sum_{k=n}^{p}\kappa\gamma_k\right)\exp\left(\sum_{k=n}^{n_0-1}\gamma_k^2 mL\right) .\end{aligned}$$

Using the two previous inequalities in (B.19), we get

$$\sum_{i=n}^{p}\prod_{k=i+1}^{p}\eta_k\gamma_i\leq\kappa^{-1}+\left\{\prod_{k=n}^{n_0-1}(\gamma_{n_0-1}L^2)^{-1}\left(1+L^2\gamma_k^2\right)\right\}\exp\left(-\sum_{k=n}^{p}\kappa\gamma_k+\sum_{k=n}^{n_0-1}\gamma_k^2 mL\right) . \tag{B.20}$$

Combining (B.18) and (B.20) in (B.17) concluded the proof. □

We now deal with bounds on $W_2(\mu_0 Q_\gamma^n,\pi)$ under **G** 6. But before we preface our result by some techincal lemmas.

Lemma B.8. *Assume* ***L*** *1 and* ***H*** *11. Let* $\zeta_0\in\mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d)$, $(Y_t,\overline{Y}_t)_{t\geq0}$ *such that* $(Y_0,\overline{Y}_0)$ *is distributed according to* ζ_0 *and given by (5.12). Let* $(\mathcal{F}_t)_{t\geq0}$ *be the filtration associated with* $(B_t)_{t\geq0}$ *with* $\mathcal{F}_0$*, the* σ*-field generated by* $(Y_0,\overline{Y}_0)$*. Then for all* $n\geq0$*,* $\epsilon_1>0$ *and*

$\epsilon_2 > 0$,

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|Y_{\Gamma_{n+1}} - \overline{Y}_{\Gamma_{n+1}}\right\|^2\right]$$
$$\leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon_1) + \gamma_{n+1}L((1+\epsilon_2)\gamma_{n+1} - 2/(m+L))_+\}\left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2$$
$$+ \gamma_{n+1}^2(1/(2\epsilon_1) + (1+\epsilon_2^{-1})\gamma_{n+1})\left(dL^2 + (L^4\gamma_{n+1}/2)\left\|Y_{\Gamma_n} - x^\star\right\|^2 + dL^4\gamma_{n+1}^2/12\right) .$$

Proof. Let $n \geq 0$ and $\epsilon_1 > 0$, and set $\Delta_n = Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}$ by definition we have:

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Delta_{n+1}\|^2\right] = \|\Delta_n\|^2 + \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\int_{\Gamma_n}^{\Gamma_{n+1}}\left\{\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\right\}\mathrm{d}s\right\|^2\right]$$
$$- 2\gamma_{n+1}\left\langle \Delta_n, \nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n})\right\rangle - 2\int_{\Gamma_n}^{\Gamma_{n+1}}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\langle \Delta_n, \{\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\}\rangle \mathrm{d}s\right] .$$

Using the two inequalities $|\langle a, b\rangle| \leq \epsilon_1\|a\|^2 + (4\epsilon_1)^{-1}\|b\|^2$ and (6.4), we get

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Delta_{n+1}\|^2\right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon_1)\}\|\Delta_n\|^2 - 2\gamma_{n+1}/(m+L)\left\|\nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n})\right\|^2$$
$$+ \mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\int_{\Gamma_n}^{\Gamma_{n+1}}\{\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\}\mathrm{d}s\right\|^2\right] + \frac{1}{2\epsilon_1}\int_{\Gamma_n}^{\Gamma_{n+1}}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\right]\mathrm{d}s . \tag{B.21}$$

Using $\|a+b\|^2 \leq (1+\epsilon_2)\|a\|^2 + (1+\epsilon_2^{-1})\|b\|^2$ and the Jensen's inequality, we have

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\left\|\int_{\Gamma_n}^{\Gamma_{n+1}}\left\{\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\right\}\mathrm{d}s\right\|^2\right] \leq (1+\epsilon_2)\gamma_{n+1}^2\left\|\nabla U(Y_{\Gamma_n}) - \nabla U(\overline{Y}_{\Gamma_n})\right\|^2$$
$$\gamma_{n+1}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\int_{\Gamma_n}^{\Gamma_{n+1}}\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\mathrm{d}s\right] .$$

This result and **L**1 imply,

$$\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\Delta_{n+1}\|^2\right] \leq \{1 - \gamma_{n+1}(\kappa - 2\epsilon_1) + \gamma_{n+1}L((1+\epsilon_2)\gamma_{n+1} - 2/(m+L))_+\}\|\Delta_n\|^2$$
$$+ ((1+\epsilon_2^{-1})\gamma_{n+1} + (2\epsilon_1)^{-1})\int_{\Gamma_n}^{\Gamma_{n+1}}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\right]\mathrm{d}s . \tag{B.22}$$

By **L**1, the Markov property of $(Y_t)_{t\geq 0}$ and ((ii)), we have

$$\int_{\Gamma_n}^{\Gamma_{n+1}}\mathbb{E}^{\mathcal{F}_{\Gamma_n}}\left[\|\nabla U(Y_s) - \nabla U(Y_{\Gamma_n})\|^2\right]\mathrm{d}s$$
$$\leq L^2\left(d\gamma_{n+1}^2 + dL^2\gamma_{n+1}^4/12 + (L^2\gamma_{n+1}^3/2)\left\|Y_{\Gamma_n} - x^\star\right\|^2\right) .$$

Plugging this bound in (B.22) concludes the proof. □

Lemma B.9. *Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence of positive numbers. Let $\varpi, \beta > 0$ be positive constants satisfying $\varpi^2 \leq 4\beta$ and $\tau > 0$. Assume there exists $N \geq 1$, $\gamma_N \leq \tau$ and $\gamma_N \varpi \leq 1$. Then for all $n \geq 0$, $j \geq 2$*

(i)

$$\sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi + \gamma_k \beta (\gamma_k - \tau)_+\right) \gamma_i^j \leq \sum_{i=N}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j + \left\{\beta^{-1} \gamma_1^{j-2} \prod_{k=1}^{N-1} \left(1 + \gamma_k^2 \beta\right)\right\} \prod_{k=N}^{n+1} \left(1 - \varpi \gamma_k\right) .$$

(ii) For all $\ell \in \{N, \ldots, n\}$,

$$\sum_{i=N}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j \leq \exp\left(-\sum_{k=\ell}^{n+1} \varpi \gamma_k\right) \sum_{i=N}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi} .$$

Proof. By definition of N we have

$$\begin{aligned}
&\sum_{i=1}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi + \gamma_k \beta (\gamma_k - \tau)_+\right) \gamma_i^j \\
&\leq \sum_{i=N}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j + \left\{\sum_{i=1}^{N-1} \prod_{k=i+1}^{N-1} \left(1 + \gamma_k^2 \beta\right) \gamma_i^j\right\} \prod_{k=N}^{n+1} \left(1 - \gamma_k \varpi\right) .
\end{aligned} \tag{B.23}$$

Using that $(\gamma_k)_{k\geq 1}$ is nonincreasing, we have

$$\begin{aligned}
\sum_{i=1}^{N-1} \prod_{k=i+1}^{N-1} \left(1 + \gamma_k^2 \beta\right) \gamma_i^j &\leq \sum_{i=1}^{N-1} \frac{\gamma_i^{j-2}}{\beta} \left\{\prod_{k=i}^{N-1} \left(1 + \gamma_k^2 \beta\right) - \prod_{k=i+1}^{N-1} \left(1 + \gamma_k^2 \beta\right)\right\} \\
&\leq \beta^{-1} \gamma_1^{j-2} \prod_{k=1}^{N-1} \left(1 + \gamma_k^2 \beta\right) .
\end{aligned}$$

Plugging this inequality in (B.23) concludes the proof of (i). Let $\ell \in \{N, \ldots, n+1\}$. Since $(\gamma_k)_{k\geq 1}$ is nonincreasing and for every $x \in \mathbb{R}$, $(1 + x) \leq e^x$, we get

$$\begin{aligned}
\sum_{i=N}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j &= \sum_{i=N}^{\ell-1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j + \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i^j \\
&\leq \sum_{i=N}^{\ell-1} \exp\left(-\sum_{k=i+1}^{n+1} \varpi \gamma_k\right) \gamma_i^j + \gamma_\ell^{j-1} \sum_{i=\ell}^{n+1} \prod_{k=i+1}^{n+1} \left(1 - \gamma_k \varpi\right) \gamma_i \\
&\leq \exp\left(-\sum_{k=\ell}^{n+1} \varpi \gamma_k\right) \sum_{i=N}^{\ell-1} \gamma_i^j + \frac{\gamma_\ell^{j-1}}{\varpi} .
\end{aligned}$$

□

Lemma B.10. *Let $(\gamma_k)_{k\geq 1}$ be a nonincreasing sequence of positive numbers, $\varpi, \beta, \tau > 0$ be positive real numbers, and $N \geq 1$ satisfying the assumptions of Lemma B.9. Let $\mathsf{P} \in \mathbb{N}^*$, $C_i \geq 0$, $i = 0, \ldots, \mathsf{P}$ be positive constants and $(u_n)_{n\geq 0}$ be a sequence of real numbers with $u_0 \geq 0$ satisfying for all $n \geq 0$*

$$u_{n+1} \leq (1 - \gamma_{n+1}\varpi + \beta\gamma_{n+1}(\gamma_{n+1} - \tau)_+)u_n + \sum_{i=0}^{\mathsf{P}} C_j \gamma_{n+1}^{j+2} \,.$$

Then for all $n \geq 1$,

$$u_n \leq \left\{\prod_{k=1}^{N-1}\left(1 + \beta\gamma_k^2\right)\right\} \prod_{k=N}^{n} (1 - \gamma_k\varpi)\, u_0 + \sum_{j=0}^{\mathsf{P}} C_j \sum_{i=N}^{n} \prod_{k=i+1}^{n} (1 - \gamma_k\varpi)\, \gamma_i^{j+2}$$
$$+ \left\{\sum_{j=0}^{\mathsf{P}} C_j \beta^{-1}\gamma_1^j \prod_{k=1}^{N-1}\left(1 + \gamma_k^2\beta\right)\right\} \prod_{k=N}^{n} (1 - \varpi\gamma_k) \;.$$

Proof. This is a consequence of a straightforward induction and Lemma B.9-(i). □

Proposition B.11. *Assume* ***L*** *1,* ***H*** *11 and* ***G*** *6. Let $x^\star$ be the unique minimizer of U. Let $\zeta_0 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, $(Y_t, \overline{Y}_t)_{t\geq 0}$ such that $(Y_0, \overline{Y}_0)$ is distributed according to ζ_0 and given by (5.12). Then for all $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1}]$:*

$$\mathbb{E}\left[\left\|Y_t - \overline{Y}_t\right\|^2\right] \leq \tilde{u}_n^{(1)}\mathbb{E}\left[\left\|Y_0 - \overline{Y}_0\right\|^2\right] + \tilde{u}_n^{(4)} + \tilde{u}_{t,n}^{(5)} \,,$$

where

$$\tilde{u}_n^{(1)}(\gamma) \overset{def}{=} \left\{\prod_{k=1}^{n_1-1}\left(1 + L^2(1+\rho)\gamma_k^2\right)\right\} \prod_{k=n_1}^{n} (1 - \kappa\gamma_k/2) \;, \tag{B.24}$$

$$\tilde{u}_n^{(4)}(\gamma) \overset{def}{=} \sum_{i=n_1}^{n} \gamma_i^2 \mathsf{f}(\gamma_i) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2)$$
$$+ \mathsf{f}(\gamma_1)(L^2(1+\rho))^{-1}\left\{\prod_{k=1}^{n_1-1} (1 + \gamma_k^2(1+\rho)L^2)\right\} \prod_{k=n_1}^{n} (1 - \kappa\gamma_k/2) \,,$$

where for $\gamma > 0$,

$$\mathsf{f}(\gamma) = \left\{2\kappa^{-1} + (1+\rho^{-1})\gamma\right\}(dL^2 + \delta L^4\gamma/2 + dL^4\gamma^2/12) \,,$$
$$\delta = \max_{i\geq 1}\left\{e^{-2m\Gamma_{i-1}}\mathbb{E}\left[\|Y_0 - x^\star\|^2\right] + (1 - e^{-2m\Gamma_{i-1}})(d/m)\right\} \;,$$

and

$$\tilde{u}_{t,n}^{(5)}(\gamma) \overset{def}{=} \frac{m+L}{2}\left(\frac{(t-\Gamma_n)^3 L^2}{3} E_{1,n}(\mu_0, \gamma) + (t-\Gamma_n)^2 d\right) \;,$$

where $E_{1,n}(\mu_0, \gamma)$ is given by (B.16) and μ_0 is the initial distribution of $\overline{Y}_0$.

Proof. Lemma B.8 with $\epsilon_1 = \kappa/4$ and $\epsilon_2 = \rho$, **G**6, Lemma B.9, ((i)) imply for all $n \geq 0$

$$\mathbb{E}\left[\left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2\right] \leq \tilde{u}_n^{(1)}(\gamma)\mathbb{E}\left[\left\|Y_0 - \overline{Y}_0\right\|^2\right] + \tilde{u}_n^{(4)}(\gamma)\,. \tag{B.25}$$

Now let $n \geq 0$ and $t \in [\Gamma_n, \Gamma_{n+1}]$. By (5.12),

$$\left\|Y_t - \overline{Y}_t\right\|^2 = \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 - 2\int_{\Gamma_n}^t \left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), Y_s - \overline{Y}_s \right\rangle \mathrm{d}s\,. \tag{B.26}$$

Moreover for all $s \in [\Gamma_n, \Gamma_{n+1}]$, by (6.4) we get

$$\begin{aligned}\left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), Y_s - \overline{Y}_s \right\rangle &= \left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), Y_s - \overline{Y}_{\Gamma_n} + \overline{Y}_{\Gamma_n} - \overline{Y}_s \right\rangle \\ &\geq (m+L)^{-1}\left\|\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\right\|^2 + \left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), \overline{Y}_{\Gamma_n} - \overline{Y}_s \right\rangle\,.\end{aligned} \tag{B.27}$$

Since $|\langle a, b\rangle| \leq (m+L)^{-1}\|a\|^2 + (m+L)\|b\|^2/4$, we have

$$\left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), \overline{Y}_{\Gamma_n} - \overline{Y}_s \right\rangle \geq -(m+L)^{-1}\left\|\nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n})\right\|^2 - (m+L)\left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2/4\,.$$

Using this inequality in (B.27), we get

$$\left\langle \nabla U(Y_s) - \nabla U(\overline{Y}_{\Gamma_n}), Y_s - \overline{Y}_s \right\rangle \geq -(m+L)\left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2/4\,,$$

and (B.26) becomes

$$\left\|Y_t - \overline{Y}_t\right\|^2 \leq \left\|Y_{\Gamma_n} - \overline{Y}_{\Gamma_n}\right\|^2 + ((m+L)/2)\int_{\Gamma_n}^t \left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2 \mathrm{d}s \tag{B.28}$$

Therefore, by the previous inequality, it remains to bound the expectation of $\left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2$. By (5.12) and using $\nabla U(x^\star) = 0$,

$$\left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2 = \left\|-(s-\Gamma_n)(\nabla U(\overline{Y}_{\Gamma_n}) - \nabla U(x^\star)) + \sqrt{2}(B_s - B_{\Gamma_n})\right\|^2\,.$$

Then taking the expectation, using the Markov property of $(B_t)_{t\geq 0}$ and **L**1, we have

$$\mathbb{E}\left[\left\|\overline{Y}_s - \overline{Y}_{\Gamma_n}\right\|^2\right] \leq (s-\Gamma_n)^2 L^2 \mathbb{E}\left[\left\|\overline{Y}_{\Gamma_n} - x^\star\right\|^2\right] + 2(s-\Gamma_n)d\,. \tag{B.29}$$

The proof follows from taking the expectation in (B.28), combining (B.25)-(B.29), and using Theorem B.7. □

Theorem B.12. *Assume **L**1, **H**11 and **G**6. Then for all $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ and $n \geq 1$,*

$$W_2^2(\mu_0 Q_\gamma^n, \pi) \leq \tilde{u}_n^{(1)}(\gamma) W_2^2(\mu_0, \pi) + \tilde{u}_n^{(2)}(\gamma) \,, \tag{B.30}$$

where $(\tilde{u}_n^{(1)})_{n \geq 0}$ is given by (B.24) and

$$\tilde{u}_n^{(2)}(\gamma) \overset{def}{=} \sum_{i=n_1}^{n} \gamma_i^2 \phi(\gamma_i) \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) + \phi(\gamma_1)(L^2(1+\rho))^{-1} \left\{ \prod_{k=1}^{n_1-1} (1 + \gamma_k^2(1+\rho)L^2) \right\} \prod_{k=n_1}^{n} (1 - \kappa\gamma_k/2) \,, \tag{B.31}$$

where

$$\phi(\gamma) = \left\{ 2\kappa^{-1} + (1+\rho^{-1})\gamma \right\} (dL^2 + dL^4\gamma/(2m) + dL^4\gamma^2/12) \,.$$

Proof. Let ζ_0 be an optimal transference plan of μ_0 and π. Let $(Y_t, \overline{Y}_t)_{t \geq 0}$ with $(Y_0, \overline{Y}_0)$ distributed according to ζ_0 and defined by (5.12). By definition of W_2 and since for all $t \geq 0$, π is invariant for P_t, $W_2^2(\mu_0 Q^n, \pi) \leq \mathbb{E}[\|Y_{\Gamma_n} - X_{\Gamma_n}\|^2]$. Then the proof follows from Proposition B.11 since Y_0 is distributed according to π and by (6.7), which shows that $\delta \leq d/m$. □

B.3.1 Explicit bound based on Theorem B.12 for $\gamma_k = \gamma_1 k^\alpha$ with $\alpha \in (0,1]$

We give here a bound on the sequences $(\tilde{u}_n^{(1)}(\gamma))_{n \geq 1}$ and $(\tilde{u}_n^{(2)}(\gamma))_{n \geq 1}$ for $(\gamma_k)_{k \geq 1}$ defined by $\gamma_1 > 0$ and $\gamma_k = \gamma_1 k^{-\alpha}$ for $\alpha \in (0,1]$. Recall that $\boldsymbol{\psi}_\beta$ is given by (B.5). First note, since $(\gamma_k)_{k \geq 1}$ is nonincreasing, for all $n \geq 1$, we have

$$\tilde{u}_n^{(2)}(\gamma) \leq \sum_{j=0}^{1} \mathsf{C}_j \sum_{i=n_1}^{n} \gamma_i^{j+2} \prod_{k=i+1}^{n} (1 - \kappa\gamma_k/2) + \sum_{j=0}^{1} \mathsf{C}_j (L^2(1+\rho))^{-1} \gamma_1^j \left\{ \prod_{k=1}^{n_1-1} (1 + \gamma_k^2(1+\rho)L^2) \right\} \prod_{k=n_1}^{n} (1 - \kappa\gamma_k/2) \,, \tag{B.32}$$

where

$$\mathsf{C}_1 = \mathsf{b} dL^2 \,, \mathsf{C}_2 = \mathsf{b}(dL^4/(2m) + \gamma_1 dL^4/12) \,, \mathsf{b} = 2\kappa^{-1} + (1+\rho^{-1})\gamma_1 \,.$$

1. For $\alpha = 1$ and $n_1 = 1$, by (B.6) and (B.7), we have

$$\tilde{u}_n^{(1)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2}$$

$$\tilde{u}_n^{(2)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \sum_{j=0}^{1} \mathsf{C}_j \left\{ \gamma_1^{j+2} (\boldsymbol{\psi}_{\kappa\gamma_1/2-1-j}(n+1) + 1) + (L^2(1+\rho))^{-1} \gamma_1^j \right\} .$$

For $n_1 > 1$, since $(\gamma_k)_{k \geq 0}$ is non increasing, using again (B.6), (B.7), and the bound for $t \in \mathbb{R}$, $(1+t) \leq e^t$

$$\tilde{u}_n^{(1)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \exp\left\{\kappa\gamma_1 \boldsymbol{\psi}_0(n_1)/2 + L^2(1+\rho)\gamma_1^2(\boldsymbol{\psi}_{-1}(n_1-1)+1)\right\}$$

$$\tilde{u}_n^{(2)}(\gamma) \leq (n+1)^{-\kappa\gamma_1/2} \sum_{j=0}^{1} \mathsf{C}_j \left(\gamma_1^{j+2}(\boldsymbol{\psi}_{\kappa\gamma_1/2-1-j}(n+1) - \boldsymbol{\psi}_{\kappa\gamma_1/2-1-j}(n_1)+1)\right.$$
$$\left. +(\gamma_1^j/(L^2(1+\rho)) \exp\left\{\kappa\gamma_1 \boldsymbol{\psi}_0(n_1)/2 + L^2(1+\rho)\gamma_1^2(\boldsymbol{\psi}_{-1}(n_1-1)+1)\right\}\right) .$$

Thus, for $\gamma_1 > \kappa/2$, we get a bound in $\mathcal{O}(n^{-1})$.

2. For $\alpha \in (0,1)$ and $n_1 = 1$, by (B.6) and Lemma B.9-(ii) applied with $\ell = \lceil n/2 \rceil$, we have

$$\tilde{u}_n^{(1)}(\gamma) \leq \exp\left(-(\kappa\gamma_1/2)\boldsymbol{\psi}_{\alpha-1}(n+1)\right)$$

$$\tilde{u}_n^{(2)}(\gamma) \leq \sum_{j=0}^{1} \mathsf{C}_j \left\{\gamma_1^{j+2}\left(\boldsymbol{\psi}_{1-\alpha(j+2)}(\lceil n/2 \rceil)+1\right) \exp\left\{-(\kappa\gamma_1/2)(\boldsymbol{\psi}_{1-\alpha}(n+1) - \boldsymbol{\psi}_{1-\alpha}(\lceil n/2 \rceil))\right\}\right.$$
$$\left. +2\kappa^{-1}\gamma_1^{j+1}(n/2)^{-\alpha(j+1)} + \gamma_1^j \exp\left\{-(\kappa\gamma_1/2)\boldsymbol{\psi}_{\alpha-1}(n+1)\right\}\right\} .$$

For $n_1 > 1$ and $\lceil n/2 \rceil \geq n_1$, since $(\gamma_k)_{k \geq 0}$ is non increasing, using again (B.6), and Lemma B.9-(ii) applied with $\ell = \lceil n/2 \rceil$, and the bound for $t \in \mathbb{R}$, $(1+t) \leq e^t$, we get

$$\tilde{u}_n^{(1)}(\gamma) \leq \exp\left\{-\kappa\gamma_1(\boldsymbol{\psi}_{\alpha-1}(n+1) - \boldsymbol{\psi}_{1-\alpha}(n_1))/2 + L^2(1+\rho)\gamma_1^2(\boldsymbol{\psi}_{1-2\alpha}(n_1-1)+1)\right\}$$

$$\tilde{u}_n^{(2)}(\gamma) \leq \sum_{j=0}^{1} \mathsf{C}_j \left\{2\kappa^{-1}\gamma_1^{j+1}(n/2)^{-\alpha(j+1)}\right.$$
$$+ \gamma_1^{j+2}\left(\boldsymbol{\psi}_{1-\alpha(j+2)}(\lceil n/2 \rceil) - \boldsymbol{\psi}_{1-\alpha(j+2)}(n_1)+1\right) \exp\left\{-(\kappa\gamma_1/2)(\boldsymbol{\psi}_{1-\alpha}(n+1) - \boldsymbol{\psi}_{1-\alpha}(\lceil n/2 \rceil))\right\}$$
$$\left. +(\gamma_1^j/(L^2(1+\rho)) \exp\left\{-\kappa\gamma_1(\boldsymbol{\psi}_{\alpha-1}(n+1) - \boldsymbol{\psi}_{1-\alpha}(n_1))/2 + L^2(1+\rho)\gamma_1^2(\boldsymbol{\psi}_{1-2\alpha}(n_1-1)+1)\right\}\right\} .$$

B.4 Explicit bounds on the MSE

Without loss of generality, assume that $\|f\|_{\mathrm{Lip}} = 1$. In the following, denote by $\Omega(x) \stackrel{\text{def}}{=} \|x - x^\star\|^2 + d/m$ and C a constant (which may take different values upon each appearance), which does not depend on m, L, γ_1, α and $\|x - x^\star\|$.

B.5 Explicit bounds based on Theorem 6.4

1. First for $\alpha = 0$, recall by (B.1), (6.19) and (6.20) we have for all $p \geq 1$,

$$u_p^{(1)}(\gamma) W_2^2(\delta_x, \pi) + u_p^{(2)}(\gamma) \leq 2\Omega(x)(1-\kappa\gamma_1/2)^p + 2\kappa^{-1}(\mathsf{A}_0\gamma_1 + \mathsf{A}_1\gamma_1^2) ,$$

where A_0 and A_1 are given by (B.2) and (B.3) respectively. So by Proposition 6.9 and Lemma 6.24, we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq C\left(\frac{\kappa^{-1}\exp(-\kappa N\gamma_1/2)\Omega(x)}{\gamma_1 n} + \kappa^{-1}\mathsf{A}_0\gamma_1\right) .$$

Therefore plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C\left(\kappa^{-1}\mathsf{A}_0\gamma_1 + \frac{\kappa^{-2} + \kappa^{-1}\exp(-\kappa N\gamma_1/2)\Omega(x)}{n\gamma_1}\right) . \tag{B.33}$$

So with fixed γ_1 this bound is of order γ_1. If we fix the number of iterations n, we can optimize the choice of γ_1. Set

$$\gamma_{\star,0}(n) = (\kappa^{-1}\mathsf{A}_0)^{-1}(C_{\mathrm{MSE},0}/n)^{1/2} \ , \ \text{where } C_{\mathrm{MSE},0} \stackrel{\text{def}}{=} \kappa^{-3}\mathsf{A}_0 \ ,$$

and (B.33) becomes if $\gamma_1 \leftarrow \gamma_{\star,0}(n)$,

$$\mathrm{MSE}_f(N,n) \leq C(C_{\mathrm{MSE},0}n)^{-1/2}\left(\kappa^{-1}\exp(-\kappa N\gamma_{\star,0}(n)/2)\Omega(x) + C_{\mathrm{MSE},0}\right) .$$

Setting $N_0(n) = 2(\kappa\gamma_{\star,0}(n))^{-1}\log(\Omega(x))$, we end up with

$$\mathrm{MSE}_f(N_0(n),n) \leq C(C_{\mathrm{MSE},0}/n)^{1/2} .$$

Note that $N_0(n)$ is of order $n^{1/2}$.

2. For $\alpha \in (0,1/2)$, recall by (B.1), (6.19) and (6.20) we have for all $p \geq 1$,

$$u_p^{(1)}(\gamma)W_2^2(\delta_x,\pi) \leq 2\Omega(x)\prod_{i=1}^{p}(1-\kappa\gamma_i/2) . \tag{B.34}$$

So by Proposition 6.9, Lemma 6.24, (B.6) and (B.8), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{(1-2\alpha)n^\alpha} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x)}{\gamma_1 n^{1-\alpha}}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{(1-2\alpha)n^\alpha} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x) + \kappa^{-2}}{\gamma_1 n^{1-\alpha}}\right) . \tag{B.35}$$

At fixed γ_1, this bound is of order $n^{-\alpha}$, and is better than (B.33) for $(\gamma_k)_{k\geq 1}$ constant. If we fix the number of iterations n, we can optimize the choice of γ_1 again. Set

$$\gamma_{\star,\alpha}(n) = (\kappa^{-1}\mathsf{A}_0/(1-2\alpha))^{-1}(C_{\mathrm{MSE},\alpha}/n^{1-2\alpha})^{1/2} \ , \ \text{where } C_{\mathrm{MSE},\alpha} \stackrel{\text{def}}{=} \kappa^{-3}\mathsf{A}_0/(1-2\alpha) \ ,$$

(B.35) becomes with $\gamma_1 \leftarrow \gamma_{\star,\alpha}(n)$,

$$\begin{aligned}&\mathrm{MSE}_f(N,n)\\&\le C(C_{\mathrm{MSE},\alpha}n)^{-1/2}\left(\kappa^{-1}\exp\left\{-\kappa N^{1-\alpha}\gamma_{\star,\alpha}(n)/(2(1-\alpha))\right\}\Omega(x)+C_{\mathrm{MSE},\alpha}\right)\ .\end{aligned}$$

Setting $N_\alpha(n) = \{2(1-\alpha)(\kappa\gamma_{\star,\alpha}(n))^{-1}\log(\Omega(x))\}^{1/(1-\alpha)}$, we end up with

$$\mathrm{MSE}_f(N_\alpha(n),n) \le C(C_{\mathrm{MSE},\alpha}/n)^{1/2}\ .$$

It is worthwhile to note that the order of $N_\alpha(n)$ in n is $n^{(1-2\alpha)/(2(1-\alpha))}$, and $C_{\mathrm{MSE},\alpha}$ goes to infinity as $\alpha \to 1/2$.

3. If $\alpha = 1/2$, (B.34) still holds. By Lemma 6.24, (B.6) and (B.8), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)]-\pi(f)\right\}^2 \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1\log(n)}{n^{1/2}}+\frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1N^{1/2}/4\right\}\Omega(x)}{\gamma_1 n^{1/2}}\right)\ .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1\log(n)}{n^{1/2}}+\frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1N^{1/2}/4\right\}\Omega(x)+\kappa^{-2}}{\gamma_1 n^{1/2}}\right)\ . \tag{B.36}$$

At fixed γ_1, the order of this bound is $\log(n)n^{-1/2}$, and is the best bound for the MSE. Fix the number of iterations n, and we now optimize the choice of γ_1. Set

$$\gamma_{\star,1/2}(n) = (\kappa^{-1}\mathsf{A}_0)^{-1}(C_{\mathrm{MSE},1/2}/\log(n))^{1/2}\ ,\ \text{where } C_{\mathrm{MSE},1/2} \overset{\text{def}}{=} \kappa^{-3}\mathsf{A}_0\ ,$$

and (B.36) becomes with $\gamma_1 \leftarrow \gamma_{\star,1/2}(n)$,

$$\begin{aligned}&\mathrm{MSE}_f(N,n)\\&\quad\le C\left(\frac{\log(n)}{nC_{\mathrm{MSE},1/2}}\right)^{1/2}\left(\kappa^{-1}\exp\left\{-\kappa N^{1/2}\gamma_{\star,1/2}(n)/4\right\}\Omega(x)+\frac{C_{\mathrm{MSE},1/2}}{\log(n)}\right)\ .\end{aligned}$$

Setting $N_{1/2}(n) = (4(\kappa\gamma_{\star,1/2}(n))^{-1}\log(\Omega(x)))^2$, we end up with

$$\mathrm{MSE}_f(N_{1/2}(n),n) \le C\left(\frac{\log(n)C_{\mathrm{MSE},1/2}}{n}\right)^{1/2}\ .$$

We can see that we obtain a worse bound than for $\alpha = 0$ and $\alpha \in (0,1/2)$.

4. For $\alpha \in (1/2, 1]$, (B.34) still holds. By Lemma 6.24, (B.6) and (B.8), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{n^{1-\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x)}{\gamma_1 n^{1-\alpha}}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{n^{1-\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x) + \kappa^{-2}}{\gamma_1 n^{1-\alpha}}\right) .$$

For fixed γ_1, the MSE is of order $n^{1-\alpha}$, and is worse than for $\alpha = 1/2$. For a fixed number of iteration n, optimizing γ_1 would imply to choose $\gamma_1 \to +\infty$ as $n \to +\infty$. Therefore, in that case, the best choice of γ_1 is the largest possible value $1/(m+L)$.

5. For $\alpha = 1$, (B.34) still holds. By Lemma 6.24, (B.6) and (B.8), the bias is upper bounded by

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{\log(n)} + \frac{\kappa^{-1}N^{-\kappa\gamma_1/2}\Omega(x)}{\gamma_1\log(n)}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \le C\left(\frac{\kappa^{-1}\mathsf{A}_0\gamma_1}{\log(n)} + \frac{\kappa^{-1}N^{-\kappa\gamma_1/2}\Omega(x) + \kappa^{-2}}{\gamma_1\log(n)}\right) .$$

For fixed γ_1, the order of the MSE is $(\log(n))^{-1}$. For a fixed number of iterations, the conclusions are the same than for $\alpha \in (1/2, 1)$.

B.5.1 Explicit bound based on Theorem 6.7

1. First for $\alpha = 0$, recall by (B.9), (6.19) and (6.20) we have for all $p \ge 1$,

$$u_p^{(1)}(\gamma)W_2^2(\delta_x, \pi) + u_p^{(3)}(\gamma) \le 2\Omega(x)(1-\kappa\gamma_1/2)^p + 2\kappa^{-1}(\mathsf{B}_0\gamma_1 + \mathsf{B}_1\gamma_1^2) ,$$

where B_0 and B_1 are given by (B.10) and (B.11) respectively. So by Proposition 6.9 and Lemma 6.24, we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \le C\left(\frac{\kappa^{-1}\exp(-\kappa N\gamma_1/2)\Omega(x)}{\gamma_1 n} + \kappa^{-1}\mathsf{B}_0\gamma_1^2\right) .$$

Therefore plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \le C\left(\kappa^{-1}\mathsf{B}_0\gamma_1^2 + \frac{\kappa^{-2} + \kappa^{-1}\exp(-\kappa N\gamma_1/2)\Omega(x)}{n\gamma_1}\right) . \tag{B.37}$$

So with fixed γ_1 this bound is of order γ_1. If we fix the number of iterations n, we can optimize the choice of γ_1. Set

$$\gamma_{\star,0}(n) = (\kappa \mathsf{B}_0 n)^{-1/3} ,$$

and (B.33) becomes if $\gamma_1 \leftarrow \gamma_{\star,0}(n)$,

$$\mathrm{MSE}_f(N,n) \leq C(\mathsf{B}_0^{-1/2} n)^{-2/3} \left(\kappa^{-4/3} \exp(-\kappa N \gamma_{\star,0}(n)/2)\Omega(x) + \kappa^{-5/3}\right) .$$

Setting $N_0(n) = 2(\kappa\gamma_{\star,0}(n))^{-1}\log(\Omega(x))$, we end up with

$$\mathrm{MSE}_f(N_0(n),n) \leq C(\mathsf{B}_0^{-1/2}\kappa^{5/2} n)^{-2/3} .$$

Note that $N_0(n)$ is of order $n^{1/3}$.

2. For $\alpha \in (0,1/3)$, recall by (B.9), (6.19) and (6.20) we have for all $p \geq 1$,

$$u_p^{(1)}(\gamma) W_2^2(\delta_x,\pi) \leq 2\Omega(x) \prod_{i=1}^{p} (1-\kappa\gamma_i/2) . \tag{B.38}$$

So by Proposition 6.9, Lemma 6.24, (B.6) and (B.14), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1^2}{(1-3\alpha)n^{2\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x)}{\gamma_1 n^{1-\alpha}}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1^2}{(1-3\alpha)n^{2\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x) + \kappa^{-2}}{\gamma_1 n^{1-\alpha}}\right) . \tag{B.39}$$

If we fix the number of iterations n, we can optimize the choice of γ_1 again. Set

$$\gamma_{\star,\alpha}(n) = (n^{1-3\alpha}\kappa\mathsf{B}_0/(1-3\alpha))^{-1/3} ,$$

(B.39) becomes with $\gamma_1 \leftarrow \gamma_{\star,\alpha}(n)$,

$$\mathrm{MSE}_f(N,n) \leq C(\mathsf{B}_0^{-1/2} n)^{-2/3} \left(\kappa^{-4/3}\exp(-\kappa N\gamma_{\star,0}(n)/(2(1-\alpha)))\Omega(x) + \kappa^{-5/3}(1-3\alpha)^{-1/3}\right) .$$

Setting $N_\alpha(n) = \{(\kappa\gamma_{\star,\alpha}(n))^{-1}\log(\Omega(x))\}^{1/(1-\alpha)}$, we end up with

$$\mathrm{MSE}_f(N_\alpha(n),n) \leq C(\mathsf{B}_0^{-1/2}\kappa^{5/2} n)^{-2/3} .$$

It is worthwhile to note that the order of $N_\alpha(n)$ in n is $n^{(1-3\alpha)/(3(1-\alpha))}$.

3. If $\alpha = 1/3$, (B.38) still holds. By Lemma 6.24, (B.6) and (B.14), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1^2\log(n)}{n^{2/3}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{2/3}/4\right\}\Omega(x)}{\gamma_1 n^{2/3}}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1^2\log(n)}{n^{2/3}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{2/3}/4\right\}\Omega(x) + \kappa^{-2}}{\gamma_1 n^{2/3}}\right) . \tag{B.40}$$

At fixed γ_1, the order of this bound is $\log(n)n^{-2/3}$, and is the best bound for the MSE. Fix the number of iterations n, and we now optimize the choice of γ_1. Set

$$\gamma_{\star,1/2}(n) = (\kappa\mathsf{B}_0\log(n))^{-1/3} ,$$

and (B.40) becomes with $\gamma_1 \leftarrow \gamma_{\star,1/2}(n)$,

$$\mathrm{MSE}_f(N,n) \leq C\left(\frac{\log(n)\mathsf{B}_0}{n^2}\right)^{1/3}\left(\kappa^{-4/3}\exp\left\{-\kappa N^{1/2}\gamma_{\star,1/2}(n)/4\right\}\Omega(x) + \kappa^{-5/3}\right) .$$

Setting $N_{1/2}(n) = (4(\kappa\gamma_{\star,1/2}(n))^{-1}\log(\Omega(x)))^{3/2}$, we end up with

$$\mathrm{MSE}_f(N_{1/2}(n),n) \leq C\left(\frac{\log(n)\mathsf{B}_0}{\kappa^5 n^2}\right)^{1/3} .$$

We can see that we obtain a worse bound than for $\alpha = 0$ and $\alpha \in (0,1/3)$.

4. For $\alpha \in (1/3, 1]$, (B.38) still holds. By Lemma 6.24, (B.6) and (B.14), we have the following bound for the bias

$$\left\{\mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f)\right\}^2 \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1}{n^{1-\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x)}{\gamma_1 n^{1-\alpha}}\right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C\left(\frac{\kappa^{-1}\mathsf{B}_0\gamma_1}{n^{1-\alpha}} + \frac{\kappa^{-1}\exp\left\{-\kappa\gamma_1 N^{1-\alpha}/(2(1-\alpha))\right\}\Omega(x) + \kappa^{-2}}{\gamma_1 n^{1-\alpha}}\right) .$$

For fixed γ_1, the MSE is of order $n^{1-\alpha}$, and is worse than for $\alpha = 1/2$. For a fixed number of iteration n, optimizing γ_1 would imply to choose $\gamma_1 \to +\infty$ as $n \to +\infty$. Therefore, in that case, the best choice of γ_1 is the largest possible value $1/(m+L)$.

5. For $\alpha = 1$, (B.34) still holds. By Lemma 6.24, (B.6) and (B.8), the bias is upper bounded by

$$\left\{ \mathbb{E}_x[\hat{\pi}_n^N(f)] - \pi(f) \right\}^2 \leq C \left(\frac{\kappa^{-1} \mathsf{B}_0 \gamma_1}{\log(n)} + \frac{\kappa^{-1} N^{-\kappa\gamma_1/2} \Omega(x)}{\gamma_1 \log(n)} \right) .$$

Plugging this inequality and the one given by Theorem 6.10 in (6.22) implies:

$$\mathrm{MSE}_f(N,n) \leq C \left(\frac{\kappa^{-1} \mathsf{B}_0 \gamma_1}{\log(n)} + \frac{\kappa^{-1} N^{-\kappa\gamma_1/2} \Omega(x) + \kappa^{-2}}{\gamma_1 \log(n)} \right) .$$

For fixed γ_1, the order of the MSE is $(\log(n))^{-1}$. For a fixed number of iterations, the conclusions are the same than for $\alpha \in (1/2, 1)$.

Appendix C

Appendix of Chapter 8

Expressions of $C_5^{\bullet}(x_1,\xi_1)$

$$\begin{aligned}C_5^{\mathrm{fM}}(x_1,\xi_1) = \frac{\ell^5}{720}\Big(&\xi_1^5 g^{(5)}(x_1) + 5\xi_1^3 g^{(5)}(x_1) + 15\xi_1^3 g^{(4)}(x_1)g'(x_1)\\ &+ 15\xi_1 g^{(4)}(x_1)g'(x_1) + 30\xi_1^3 g^{(3)}(x_1)g''(x_1)\\ &+10\xi_1 g^{(3)}(x_1)g''(x_1) + 30\xi_1 g^{(3)}(x_1)g'(x_1)^2 + 35\xi_1 g'(x_1)g''(x_1)^2\Big)\end{aligned}$$

$$\begin{aligned}C_5^{\mathrm{mO}}(x_1,\xi_1) = \ell^5\Bigg(&\frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1)\\ &+ \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{29}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) - \frac{7}{48}\xi_1 g^{(3)}(x_1)g''(x_1)\\ &+ \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 + \frac{1}{6}\xi_1 g'(x_1)g''(x_1)^2\Bigg)\ .\end{aligned}$$

$$\begin{aligned}C_5^{\mathrm{bO}}(x_1,\xi_1) = \ell^5\Bigg(&\frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1)\\ &+ \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{29}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) - \frac{19}{144}\xi_1 g^{(3)}(x_1)g''(x_1)\\ &+ \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 + \frac{1}{6}\xi_1 g'(x_1)g''(x_1)^2\Bigg)\ .\end{aligned}$$

$$C_5^{\mathrm{gbO}}(x_1,\xi_1) = \ell^5 \left(\frac{1}{720}\xi_1^5 g^{(5)}(x_1) + \frac{1}{144}\xi_1^3 g^{(5)}(x_1) + \frac{1}{48}\xi_1^3 g^{(4)}(x_1)g'(x_1) \right.$$
$$+ \frac{1}{48}\xi_1 g^{(4)}(x_1)g'(x_1) + \frac{1}{72}a_3\xi_1 g^{(3)}(x_1)g''(x_1) + \frac{1}{6}a_4^2\xi_1^3 g^{(3)}(x_1)g''(x_1)$$
$$- \frac{1}{6}a_4^2\xi_1 g^{(3)}(x_1)g''(x_1) + \frac{5}{144}\xi_1^3 g^{(3)}(x_1)g''(x_1) + \frac{1}{48}\xi_1 g^{(3)}(x_1)g''(x_1)$$
$$+ \frac{1}{24}\xi_1 g^{(3)}(x_1)g'(x_1)^2 - \frac{1}{24}a_1^2\xi_1 g'(x_1)g''(x_1)^2 + \frac{1}{6}a_4^2\xi_1 g'(x_1)g''(x_1)^2$$
$$\left. + \frac{1}{24}\xi_1 g'(x_1)g''(x_1)^2 \right) .$$

Expressions of $K^\bullet$

We provide here the expressions of the quantities $K^\bullet$ involved in Theorems 8.4, 8.5, 8.7, 8.8. Let X be a random variable distributed according to π_1.

$$K^{\mathrm{fM}} = \mathbb{E}\left[\frac{79g^{(5)}(\mathrm{X})^2}{17280} + \frac{11g^{(4)}(\mathrm{X})^2 g'(\mathrm{X})^2}{1152} + \frac{77g^{(3)}(\mathrm{X})^2 g''(\mathrm{X})^2}{2592} + \frac{1}{576}g^{(3)}(\mathrm{X})^2 g'(\mathrm{X})^4 \right.$$
$$+ \frac{49g'(\mathrm{X})^2 g''(\mathrm{X})^4}{20736} + \frac{7}{576}g^{(4)}(\mathrm{X})g^{(5)}(\mathrm{X})g'(\mathrm{X}) + \frac{19}{864}g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})g''(\mathrm{X})$$
$$+ \frac{1}{288}g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})g'(\mathrm{X})^2 + \frac{7g^{(5)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X})^2}{1728}$$
$$+ \frac{1}{144}g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})g'(\mathrm{X})^3 + \frac{7}{864}g^{(4)}(\mathrm{X})g'(\mathrm{X})^2 g''(\mathrm{X})^2 + \frac{7g^{(3)}(\mathrm{X})g'(\mathrm{X})^3 g''(\mathrm{X})^2}{1728}$$
$$\left. + \frac{5}{432}g^{(3)}(\mathrm{X})^2 g'(\mathrm{X})^2 g''(\mathrm{X}) + \frac{35g^{(3)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X})^3}{2592} + \frac{29}{864}g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X}) \right] .$$

$$K^{\mathrm{mO}} = \mathbb{E}\left[\frac{79g^{(5)}(\mathrm{X})^2}{17280} + \frac{11g^{(4)}(\mathrm{X})^2 g'(\mathrm{X})^2}{1152} + \frac{1567g^{(3)}(\mathrm{X})^2 g''(\mathrm{X})^2}{3456} \right.$$
$$+ \frac{1}{576}g^{(3)}(\mathrm{X})^2 g'(\mathrm{X})^4 + \frac{1}{36}g'(\mathrm{X})^2 g''(\mathrm{X})^4 + \frac{7}{576}g^{(4)}(\mathrm{X})g^{(5)}(\mathrm{X})g'(\mathrm{X})$$
$$+ \frac{17}{192}g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})g''(\mathrm{X}) + \frac{1}{288}g^{(3)}(\mathrm{X})g^{(5)}(\mathrm{X})g'(\mathrm{X})^2$$
$$+ \frac{1}{72}g^{(5)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X})^2 + \frac{1}{144}g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})g'(\mathrm{X})^3 +$$
$$\frac{1}{36}g^{(4)}(\mathrm{X})g'(\mathrm{X})^2 g''(\mathrm{X})^2 + \frac{1}{72}g^{(3)}(\mathrm{X})g'(\mathrm{X})^3 g''(\mathrm{X})^2$$
$$+ \frac{11}{288}g^{(3)}(\mathrm{X})^2 g'(\mathrm{X})^2 g''(\mathrm{X}) + \frac{11}{72}g^{(3)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X})^3 +$$
$$\left. \frac{73}{576}g^{(3)}(\mathrm{X})g^{(4)}(\mathrm{X})g'(\mathrm{X})g''(\mathrm{X}) \right] .$$

$$
\begin{aligned}
K^{\mathrm{gbO}} = \mathbb{E}\Big[& \frac{1}{36} g'(\mathrm{X})^2 g''(\mathrm{X})^4 a_4^4 + \frac{5}{18} g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 a_4^4 \\
& + \frac{1}{9} g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) a_4^4 - \frac{1}{72} a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^4 a_4^2 + \frac{1}{72} g'(\mathrm{X})^2 g''(\mathrm{X})^4 a_4^2 \\
& + \frac{11}{72} g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 a_4^2 + \frac{1}{108} a_3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 a_4^2 \\
& + \frac{1}{36} g'(\mathrm{X})^2 g''(\mathrm{X}) g^{(3)}(\mathrm{X})^2 a_4^2 - \frac{1}{36} a_1^2 g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) a_4^2 \\
& + \frac{5}{72} g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) a_4^2 + \frac{1}{216} a_3 g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) a_4^2 \\
& + \frac{1}{72} g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X}) a_4^2 + \frac{1}{36} g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X}) a_4^2 \\
& + \frac{7}{72} g'(\mathrm{X}) g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(4)}(\mathrm{X}) a_4^2 + \frac{1}{72} g'(\mathrm{X}) g''(\mathrm{X})^2 g^{(5)}(\mathrm{X}) a_4^2 \\
& + \frac{5}{72} g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(5)}(\mathrm{X}) a_4^2 + \frac{1}{576} a_1^4 g'(\mathrm{X})^2 g''(\mathrm{X})^4 \\
& - \frac{1}{288} a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^4 + \frac{1}{576} g'(\mathrm{X})^2 g''(\mathrm{X})^4 + \frac{1}{576} g'(\mathrm{X})^4 g^{(3)}(\mathrm{X})^2 \\
& + \frac{a_3^2 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2}{5184} + \frac{1}{288} a_3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2 \\
& + \frac{79 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X})^2}{3456} + \frac{1}{96} g'(\mathrm{X})^2 g''(\mathrm{X}) g^{(3)}(\mathrm{X})^2 \\
& + \frac{1}{864} a_3 g'(\mathrm{X})^2 g''(\mathrm{X}) g^{(3)}(\mathrm{X})^2 + \frac{11 g'(\mathrm{X})^2 g^{(4)}(\mathrm{X})^2}{1152} \\
& + \frac{79 g^{(5)}(\mathrm{X})^2}{17280} - \frac{1}{96} a_1^2 g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) \\
& + \frac{1}{96} g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) - \frac{1}{864} a_1^2 a_3 g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) \\
& + \frac{1}{864} a_3 g'(\mathrm{X}) g''(\mathrm{X})^3 g^{(3)}(\mathrm{X}) - \frac{1}{288} a_1^2 g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X}) \\
& + \frac{1}{288} g'(\mathrm{X})^3 g''(\mathrm{X})^2 g^{(3)}(\mathrm{X}) - \frac{1}{144} a_1^2 g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X}) \\
& + \frac{1}{144} g'(\mathrm{X})^2 g''(\mathrm{X})^2 g^{(4)}(\mathrm{X}) + \frac{1}{144} g'(\mathrm{X})^3 g^{(3)}(\mathrm{X}) g^{(4)}(\mathrm{X}) \\
& + \frac{17}{576} g'(\mathrm{X}) g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(4)}(\mathrm{X}) + \frac{1}{432} a_3 g'(\mathrm{X}) g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(4)}(\mathrm{X}) \\
& - \frac{1}{288} a_1^2 g'(\mathrm{X}) g''(\mathrm{X})^2 g^{(5)}(\mathrm{X}) + \frac{1}{288} g'(\mathrm{X}) g''(\mathrm{X})^2 g^{(5)}(\mathrm{X}) + \\
& \frac{1}{288} g'(\mathrm{X})^2 g^{(3)}(\mathrm{X}) g^{(5)}(\mathrm{X}) + \frac{11}{576} g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(5)}(\mathrm{X}) \\
& + \frac{1}{864} a_3 g''(\mathrm{X}) g^{(3)}(\mathrm{X}) g^{(5)}(\mathrm{X}) + \frac{7}{576} g'(\mathrm{X}) g^{(4)}(\mathrm{X}) g^{(5)}(\mathrm{X}) \Big] .
\end{aligned}
$$

Bibliography

[AC93] J. H. Albert and S. Chib. "Bayesian analysis of binary and polychotomous response data". In: *Journal of the American statistical Association* 88.422 (1993), pp. 669–679.

[AFV14] C. Andrieu, G. Fort, and M. Vihola. "Quantitative convergence rates for sub-geometric Markov chains". In: *Adv. Appl. Probab.* (2014). accepted for publication.

[And+03] C. Andrieu, N. De Freitas, A. Doucet, and M. I Jordan. "An introduction to MCMC for machine learning". In: *Machine learning* 50.1-2 (2003), pp. 5–43.

[AVZ14] A. Abdulle, G. Vilmart, and K. C. Zygalakis. "High order numerical approximation of the invariant measure of ergodic SDEs". In: *SIAM J. Numer. Anal.* 52.4 (2014), pp. 1600–1622. ISSN: 0036-1429.

[Bak+08] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin. "A simple proof of the Poincaré inequality for a large class of probability measures." eng. In: *Electronic Communications in Probability [electronic only]* 13 (2008), pp. 60–66.

[Bax05] P. H. Baxendale. "Renewal theory and computable convergence rates for geometrically ergodic Markov chains". In: *Ann. Appl. Probab.* 15.1B (2005), pp. 700–738. ISSN: 1050-5164.

[BCG08] D. Bakry, P. Cattiaux, and A. Guillin. "Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré". In: *J. Funct. Anal.* 254.3 (2008), pp. 727–759. ISSN: 0022-1236.

[BDJ98] R. Bubley, M. Dyer, and M. Jerrum. "An elementary analysis of a procedure for sampling points in a convex body". In: *Random Structures Algorithms* 12.3 (1998), pp. 213–235. ISSN: 1042-9832.

[Bes+08] A. Beskos, G. Roberts, A. Stuart, and J. Voss. "MCMC methods for diffusion bridges". In: *Stoch. Dyn.* 8.3 (2008), pp. 319–350. ISSN: 0219-4937.

[BGG12] F. Bolley, I. Gentil, and A. Guillin. "Convergence to equilibrium in Wasserstein distance for Fokker-Planck equations". In: *J. Funct. Anal.* 263.8 (2012), pp. 2430–2457. ISSN: 0022-1236.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552. ISBN: 978-3-319-00226-2; 978-3-319-00227-9.

[Bha78] R. N. Bhattacharya. "Criteria for Recurrence and Existence of Invariant Measures for Multidimensional Diffusions". In: *Ann. Probab.* 6.4 (Aug. 1978), pp. 541–553.

[Bil95] P. Billingsley. *Probability and measure*. Third. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995, pp. xiv+593. ISBN: 0-471-00710-2.

[BLM13] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. Oxford University Press, Oxford, 2013, pp. x+481. ISBN: 978-0-19-953525-5.

[Bob99] S. G. Bobkov. "Isoperimetric and Analytic Inequalities for Log-Concave Probability Measures". In: *Ann. Probab.* 27.4 (Oct. 1999), pp. 1903–1921.

[Bog98] V.I. Bogachev. *Gaussian Measures*. Mathematical surveys and monographs. American Mathematical Society, 1998. ISBN: 9780821810545.

[BS09] A. Beskos and A. Stuart. "MCMC methods for sampling function space". In: *ICIAM 07—6th International Congress on Industrial and Applied Mathematics*. Eur. Math. Soc., Zürich, 2009, pp. 337–364.

[BTbk] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. New York: John Wiley & Sons, 1973 (reprinted in paperback 1992 ISBN: 0-471-57428-7 pbk.)

[But14] O. Butkovsky. "Subgeometric rates of convergence of Markov processes in the Wasserstein metric". In: *Ann. Appl. Probab.* 24.2 (2014), pp. 526–552. ISSN: 1050-5164.

[BV13] O. A. Butkovsky and A. Yu. Veretennikov. "On asymptotics for Vaserstein coupling of Markov chains". In: *Stochastic Process. Appl.* 123.9 (2013), pp. 3518–3541. ISSN: 0304-4149.

[CE05] A. S. Cherny and H.-J. Engelbert. *Singular stochastic differential equations*. Vol. 1858. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2005, pp. viii+128. ISBN: 3-540-24007-1.

[CG09] P. Cattiaux and A. Guillin. "Trends to equilibrium in total variation distance". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 45.1 (2009), pp. 117–145. ISSN: 0246-0203.

[CH13] H. M. Choi and J. P. Hobert. "The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic". In: *Electron. J. Statist.* 7 (2013), pp. 2054–2064.

[CH14] B. Cloez and M. Hairer. "Exponential ergodicity for Markov processes with random switching". In: *Bernoulli* (2014). To appear.

[CK91] E. Carlen and P. Krée. "L^p estimates on iterated stochastic integrals". In: *Ann. Probab.* 19.1 (1991), pp. 354–368. ISSN: 0091-1798.

[CL89] M. F. Chen and S. F. Li. "Coupling methods for multidimensional diffusion processes". In: *Ann. Probab.* 17.1 (1989), pp. 151–177. ISSN: 0091-1798.

[Cot+13] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. "MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster". In: *Statist. Sci.* 28.3 (Aug. 2013), pp. 424–446.

[CRR05] O. F. Christensen, G. O. Roberts, and J. S. Rosenthal. "Scaling limits for the transient phase of local Metropolis–Hastings algorithms". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005), pp. 253–268.

[Dal16] A. S. Dalalyan. "Theoretical guarantees for approximate sampling from smooth and log-concave densities". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2016), n/a–n/a. ISSN: 1467-9868.

[Das+13] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. "MAP estimators and their consistency in Bayesian nonparametric inverse problems". In: *Inverse Problems* 29.9 (2013), pp. 095017, 27. ISSN: 0266-5611.

[DFM] A. Durmus, G. Fort, and É. Moulines. *Subgeometric rates of convergence in Wasserstein distance for Markov chains.* Accepted for publication in Ann. Inst. H. Poincaré Probab. Statist.

[DGM08] R. Douc, A. Guillin, and E. Moulines. "Bounds on regeneration times and limit theorems for subgeometric Markov chains". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 44.2 (2008), pp. 239–257. ISSN: 0246-0203.

[DM] A. Durmus and É. Moulines. *Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm.* Accepted for publication in Ann. Appl. Probab.

[DM15a] A. Durmus and E. Moulines. *Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm.* Tech. rep. July 2015.

[DM15b] A. Durmus and É. Moulines. "Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm". In: *Stat. Comput.* 25.1 (2015), pp. 5–19. ISSN: 0960-3174.

[DM15c] A. Durmus and É. Moulines. "Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis adjusted Langevin algorithm". In: *Stat. Comput.* 25.1 (2015), pp. 5–19. ISSN: 0960-3174.

[DM15d] A. Durmus and É. Moulines. *Supplement to "Non-asymptotic convergence analysis for the Unadjusted Langevin Algorithm".* 2015.

[DM16] A. Durmus and É. Moulines. *Sampling from strongly log-concave distributions with the Unadjusted Langevin Algorithm.* First version : arXiv:1605.01559, in preparation. 2016.

[DMP] A. Durmus, É. Moulines, and M. Pereyra. "Sampling from convex non continuously differentiable functions, when Moreau meets Langevin". In preparation.

[DMS07] R. Douc, É. Moulines, and P. Soulier. "Computable convergence rates for sub-geometric ergodic Markov chains". In: *Bernoulli* (2007), pp. 831–848.

[DMS14] R. Douc, E. Moulines, and D. Stoffer. *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. Chapman & Hall/CRC Texts in Statistical Science, 2014.

[Dou+04] R. Douc, G. Fort, É. Moulines, and P. Soulier. "Practical drift conditions for subgeometric rates of convergence". In: *Ann. Appl. Probab.* 14.3 (2004), pp. 1353–1377.

[DT12] A. S. Dalalyan and A. B. Tsybakov. "Sparse regression learning by aggregation and Langevin Monte-Carlo". In: *J. Comput. System Sci.* 78.5 (2012), pp. 1423–1443.

[Dur+15] A. Durmus, G. O Roberts, G. Vilmart, and K. C Zygalakis. *Fast Langevin based algorithm for MCMC in high dimensions*. arXiv:1507.02166, submitted. 2015.

[Dur+16a] A. Durmus, S. Le Corff, É Moulines, and G. O. Roberts. *Optimal scaling of the Random Walk Metropolis algorithm under Lp mean differentiability.* arXiv:1604.06664, submitted. 2016.

[Dur+16b] A. Durmus, U. Simsekli, É. Moulines, R. Badeau, and G. Richard. "Stochastic Gradient Richardson-Romberg Markov Chain Monte Carlo". In: *Thirtieth Annual Conference on Neural Information Processing Systems (NIPS).* 2016.

[Ebe14] A. Eberle. "Error bounds for Metropolis–Hastings algorithms applied to perturbations of Gaussian measures in high dimensions". In: *The Annals of Applied Probability* 24.1 (Feb. 2014), pp. 337–377.

[Ebe15] A. Eberle. "Reflection couplings and contraction rates for diffusions". English. In: *Probab. Theory Related Fields* (2015), pp. 1–36. ISSN: 0178-8051.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Characterization and convergence. New York: John Wiley & Sons Inc., 1986, pp. x+534. ISBN: 0-471-08186-8.

[Erm75] D. L Ermak. "A computer simulation of charged particles in solution. I. Technique and equilibrium properties". In: *The Journal of Chemical Physics* 62.10 (1975), pp. 4189–4196.

[FF10] S. Frühwirth-Schnatter and R. Frühwirth. “Data Augmentation and MCMC for Binary and Multinomial Logit Models Statistical Modelling and Regression Structures”. In: *Statistical Modelling and Regression Structures*. Ed. by Thomas Kneib and Gerhard Tutz. Heidelberg: Physica-Verlag HD, 2010. Chap. 7, pp. 111–132. ISBN: 978-3-7908-2412-4.

[FM00] Gersende Fort and Eric Moulines. “V-Subgeometric ergodicity for a Hastings–Metropolis algorithm”. In: *Statistics & Probability Letters* 49.4 (2000), pp. 401–410. ISSN: 0167-7152.

[FM03] G. Fort and E. Moulines. “Polynomial ergodicity of Markov transition kernels”. In: *Stochastic Processes and their Applications* 103.1 (2003), pp. 57–99. ISSN: 0304-4149.

[FS15] M. Fathi and G. Stoltz. “Improving dynamical properties of stabilized discretizations of overdamped Langevin dynamics”. In: *preprint* (2015).

[GC11] M. Girolami and B. Calderhead. “Riemann manifold Langevin and Hamiltonian Monte Carlo methods”. In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 73.2 (2011). With discussion and a reply by the authors, pp. 123–214. ISSN: 1369-7412.

[GM94] U. Grenander and M. I. Miller. “Representations of knowledge in complex systems”. In: *J. Roy. Statist. Soc. Ser. B* 56.4 (1994). With discussion and a reply by the authors, pp. 549–603. ISSN: 0035-9246.

[GP12] R. B. Gramacy and N. G. Polson. “Simulation-based Regularized Logistic Regression”. In: *Bayesian Anal.* 7.3 (Sept. 2012), pp. 567–590.

[Gre83] U. Grenander. “Tutorial in pattern theory”. Division of Applied Mathematics, Brown University, Providence. 1983.

[Han03] N. R. Hansen. “Geometric ergodicity of discrete-time approximations to multivariate diffusions”. In: *Bernoulli* 9.4 (2003), pp. 725–743. ISSN: 1350-7265.

[Has70] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their application”. In: *Biometrika* 57 (1970), pp. 97–109.

[HBJ14] T. E. Hanson, A J. Branscum, and W. O. Johnson. “Informative g-priors for logistic regression”. In: *Bayesian Anal.* 9.3 (2014), pp. 597–611. ISSN: 1936-0975.

[HH06] C. C. Holmes and L. Held. “Bayesian auxiliary variable models for binary and multinomial regression”. In: *Bayesian Anal.* 1.1 (Mar. 2006), pp. 145–168.

[HK14] E. Hazan and S Kale. “Beyond the regret minimization barrier: optimal algorithms for stochastic strongly-convex optimization”. In: *Journal of Machine Learning Research* 15.1 (2014), pp. 2489–2512.

[HMS11] M. Hairer, J.C. Mattingly, and M. Scheutzow. "Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations". In: *Probability Theory and Related Fields* 149.1-2 (2011), pp. 223–259. ISSN: 0178-8051.

[HS87] R. Holley and D. Stroock. "Logarithmic Sobolev inequalities and stochastic Ising models". In: *Journal of statistical physics* 46.5 (1987), pp. 1159–1194.

[HSV14] M. Hairer, A.M. Stuart, and S.J. Vollmer. "Spectral gaps for Metropolis-Hastings algorithms in infinite dimensions". In: *Ann. Appl. Probab.* 24 (2014), pp. 2455–290.

[IW89] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library. Elsevier Science, 1989.

[JH00] Søren Fiig Jarner and Ernst Hansen. "Geometric ergodicity of Metropolis algorithms". In: *Stochastic processes and their applications* 85.2 (2000), pp. 341–361.

[JH01] G. Jones and J. Hobert. "Honest exploration of intractable probability distributions via Markov chain Monte Carlo". In: *Statist. Sci.* 16.4 (2001), pp. 312–334.

[JLM14] B. Jourdain, T Lelièvre, and B. Miasojedow. "Optimal scaling for the transient phase of Metropolis Hastings algorithms: the longtime behavior". In: *Bernoulli* 20.4 (2014), pp. 1930–1978.

[JLM15] B. Jourdain, T. Lelièvre, and B. Miasojedow. "Optimal scaling for the transient phase of the random walk Metropolis algorithm: The mean-field limit". In: *Ann. Appl. Probab.* 25.4 (Aug. 2015), pp. 2263–2300.

[JO10] A. Joulin and Y. Ollivier. "Curvature, concentration and error estimates for Markov chain Monte Carlo". In: *Ann. Probab.* 38.6 (2010), pp. 2418–2442. ISSN: 0091-1798.

[JR02] S. F. Jarner and G. O. Roberts. "Polynomial convergence rates of Markov chains". In: *Ann. Appl. Probab.* 12.1 (2002), pp. 224–247. ISSN: 1050-5164.

[JR07] S. F. Jarner and G. O. Roberts. "Convergence of heavy-tailed Monte Carlo Markov chain algorithms". In: *Scandinavian Journal of Statistics* 34.4 (2007), pp. 781–815.

[JT01a] S. F. Jarner and R. L. Tweedie. "Locally contracting iterated functions and stability of Markov chains". In: *J. Appl. Probab.* 38.2 (2001), pp. 494–507. ISSN: 0021-9002.

[JT01b] S.F. Jarner and R.L. Tweedie. "Locally contracting iterated functions and stability of Markov chains." English. In: *J. Appl. Probab.* 38.2 (2001), pp. 494–507.

[JT03] S. F. Jarner and R .L. Tweedie. "Necessary conditions for geometric and polynomial ergodicity of random-walk-type". In: *Bernoulli* 9.4 (Aug. 2003), pp. 559–578.

[Ken78] J. Kent. “Time-reversible diffusions”. In: *Advances in Applied Probability* 10 (04 Dec. 1978), pp. 819–835. ISSN: 1475-6064.

[Kha80] R. Z. Khasminskii. *Stochastic stability of differential equations*. Vol. 7. Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis. Translated from the Russian by D. Louvish. Sijthoff & Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980, pp. xvi+344. ISBN: 90-286-0100-7.

[KP92] P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Vol. 23. Applications of Mathematics (New York). Springer-Verlag, Berlin, 1992, pp. xxxvi+632. ISBN: 3-540-54062-8.

[KS05] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*. Vol. 160. Applied Mathematical Sciences. Springer-Verlag, New York, 2005, pp. xvi+339. ISBN: 0-387-22073-9.

[KS91] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991. ISBN: 9780387976556.

[Kul97] S. Kullback. *Information theory and statistics*. Reprint of the second (1968) edition. Dover Publications, Inc., Mineola, NY, 1997, pp. xvi+399. ISBN: 0-486-69684-7.

[Le 86] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. New York: Springer-Verlag New York, 1986.

[Lem05] V. Lemaire. “Estimation de la mesure invariante d’un processus de diffusion”. PhD thesis. Université Paris-Est, 2005.

[Liu08] J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York, 2008, pp. xvi+343. ISBN: 978-0-387-76369-9; 0-387-95230-6.

[LM10] V. Lemaire and S. Menozzi. “On some non asymptotic bounds for the Euler scheme”. In: *Electron. J. Probab.* 15 (2010), no. 53, 1645–1681. ISSN: 1083-6489.

[LP02] D. Lamberton and G. Pagès. “Recursive computation of the invariant distribution of a diffusion”. In: *Bernoulli* 8.3 (2002), pp. 367–405. ISSN: 1350-7265.

[LP03] D. Lamberton and G. Pagès. “Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift”. In: *Stoch. Dyn.* 3.4 (2003), pp. 435–451. ISSN: 0219-4937.

[LR04] G. J. Lord and J. Rougemont. “A numerical scheme for stochastic PDEs with Gevrey regularity.” English. In: *IMA J. Numer. Anal.* 24.4 (2004), pp. 587–604.

[LR86] T. Lindvall and L. C. G. Rogers. “Coupling of multidimensional diffusions by reflection”. In: *Ann. Probab.* 14.3 (1986), pp. 860–872. ISSN: 0091-1798.

[LV07] L. Lovász and S. Vempala. "The geometry of logconcave functions and sampling algorithms". In: *Random Structures Algorithms* 30.3 (2007), pp. 307–358. ISSN: 1042-9832.

[Met+53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equations of state calculations by fast computing machines". In: *Journal of Chemical Physics* 23.2 (1953), pp. 1087–1092.

[MS10] N. Madras and D. Sezer. "Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances". In: *Bernoulli* 16.3 (2010), pp. 882–908. ISSN: 1350-7265.

[MSH02] J. C. Mattingly, A. M. Stuart, and D. J. Higham. "Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise". In: *Stochastic Process. Appl.* 101.2 (2002), pp. 185–232. ISSN: 0304-4149.

[MST10] J. C. Mattingly, A. M. Stuart, and M. V. Tretyakov. "Convergence of numerical time-averaging and stationary measures via Poisson equations". In: *SIAM J. Numer. Anal.* 48.2 (2010), pp. 552–577. ISSN: 0036-1429.

[MT09] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability.* 2nd. New York, NY, USA: Cambridge University Press, 2009. ISBN: 0521731828, 9780521731829.

[MT93a] S. P. Meyn and R. L. Tweedie. "Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes". In: *Advances in Applied Probability* 25.3 (1993).

[MT93b] S. P. Meyn and R. L. Tweedie. "Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes". In: *Adv. in Appl. Probab.* 25.3 (1993), pp. 518–548. ISSN: 0001-8678.

[MT94] S. P. Meyn and R. L. Tweedie. "Computable Bounds for Convergence Rates of Markov Chains." In: *Annals of Applied Probability* 4 (1994), pp. 981–1011.

[MT96] K. L. Mengersen and R. L. Tweedie. "Rates of convergence of the Hastings and Metropolis algorithms". In: *Ann. Statist.* 24.1 (Feb. 1996), pp. 101–121.

[Nea93] R. M. Neal. "Bayesian Learning via Stochastic Dynamics". In: *Advances in Neural Information Processing Systems 5, [NIPS Conference].* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993, pp. 475–482. ISBN: 1-55860-274-7.

[Nes04] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course.* Applied Optimization. Springer, 2004. ISBN: 9781402075537.

[Num78] E. Nummelin. "A splitting technique for Harris recurrent Markov chains". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 43.4 (1978), pp. 309–318. ISSN: 1432-2064.

[Num84] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge tracts in mathematics. Includes index. Cambridge, Cambridgeshire, New York: Cambridge University Press, 1984. ISBN: 0-521-25005-6.

[Oza92] T. Ozaki. “A bridge between nonlinear time series models and nonlinear stochastic dynamical systems: a local linearization approach”. In: *Statist. Sinica* 2.1 (1992), pp. 113–135. ISSN: 1017-0405.

[Par81] G. Parisi. “Correlation functions and computer simulations”. In: *Nuclear Physics B* 180 (1981), pp. 378–384.

[Pet95] V. V. Petrov. *Limit theorems of probability theory*. Vol. 4. Oxford Studies in Probability. Sequences of independent random variables, Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1995, pp. xii+292. ISBN: 0-19-853499-X.

[PSW13] N. G. Polson, J. G. Scott, and J. Windle. “Bayesian Inference for Logistic Models Using Polya-Gamma Latent Variables”. In: *Journal of the American Statistical Association* 108.504 (2013), pp. 1339–1349.

[RC10] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York, 2010. ISBN: 9781441919397.

[RDF78] P. J. Rossky, J. D. Doll, and H. L. Friedman. “Brownian dynamics as smart Monte Carlo simulation”. In: *The Journal of Chemical Physics* 69.10 (1978), pp. 4628–4633.

[RGG97] G. O. Roberts, A. Gelman, and W. R. Gilks. “Weak convergence and optimal scaling of random walk Metropolis algorithms”. In: *Ann. Appl. Probab.* 7.1 (1997), pp. 110–120. ISSN: 1050-5164.

[Rob06] C. Robert. *Le choix bayésien: Principes et pratique*. Statistique et probabilités appliquées. Springer-Verlag France, 2006. ISBN: 9782287311208.

[Ros95] J. S. Rosenthal. “Rates of Convergence for Gibbs Sampling for Variance Component Models”. In: *Ann. Statist.* 23.3 (June 1995), pp. 740–761.

[RR01a] G. O Roberts and J. S Rosenthal. “Optimal scaling for various Metropolis-Hastings algorithms”. In: *Statistical science* 16.4 (2001), pp. 351–367.

[RR01b] G. O. Roberts and J. S. Rosenthal. “Small and pseudo-small sets for Markov chains”. In: *Stoch. Models* 17.2 (2001), pp. 121–145. ISSN: 1532-6349.

[RR04] G. O. Roberts and J. S. Rosenthal. “General state space Markov chains and MCMC algorithms”. In: *Probab. Surv.* 1 (2004), pp. 20–71. ISSN: 1549-5787.

[RR98] G. O. Roberts and J. S. Rosenthal. “Optimal scaling of discrete approximations to Langevin diffusions”. In: *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60.1 (1998), pp. 255–268. ISSN: 1369-7412.

[RS02] G. O. Roberts and O. Stramer. "Langevin diffusions and Metropolis-Hastings algorithms". In: *Methodol. Comput. Appl. Probab.* 4.4 (2002). International Workshop in Applied Probability (Caracas, 2002), 337–357 (2003). ISSN: 1387-5841.

[RT00] G. O. Roberts and R. L. Tweedie. "Rates of convergence of stochastically monotone and continuous time Markov models". In: *J. Appl. Probab.* 37.2 (2000), pp. 359–373. ISSN: 0021-9002.

[RT96a] G. O. Roberts and R. L. Tweedie. "Exponential convergence of Langevin distributions and their discrete approximations". In: *Bernoulli* 2.4 (1996), pp. 341–363. ISSN: 1350-7265.

[RT96b] G. O. Roberts and R. L. Tweedie. "Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms". In: *Biometrika* 83.1 (1996), pp. 95–110.

[RT99] G. O. Roberts and R. L. Tweedie. "Bounds on Regeneration Times and Convergence Rates for Markov Chains". In: *Stochastic Processes and Their Applications* 80 (1999), pp. 211–229.

[RW00] L.C.G. Rogers and D. Williams. *Diffusions, Markov processes and martingales. Vol 2: Ito calculus.* Cambridge University press, 2000.

[RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion.* Third. Vol. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, pp. xiv+602. ISBN: 3-540-64325-7.

[Sch95] M. J. Schervish. *Theory of statistics.* Springer series in statistics. Corrected second printing : 1997. New York, Berlin, Heidelberg: Springer, 1995. ISBN: 0-387-94546-6.

[SH11] D. Sabanés Bové and L. Held. "Hyper-g priors for generalized linear models". In: *Bayesian Anal.* 6.3 (2011), pp. 387–410. ISSN: 1936-0975.

[Stu10] A. M. Stuart. "Inverse problems: a Bayesian perspective". In: *Acta Numer.* 19 (2010), pp. 451–559. ISSN: 0962-4929.

[SV79] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes.* Die Grundlehren der mathematischen Wissenschaften. Berlin, Heidelberg, New York: Springer-Verlag, 1979. ISBN: 0-387-90353-4.

[SW67] C. Stone and S. Wainger. "One-sided error estimates in renewal theory". English. In: *Journal d'Analyse Mathématique* 20.1 (1967), pp. 325–352. ISSN: 0021-7670.

[Tie94] L. Tierney. "Markov chains for exploring posterior disiributions (with discussion)". In: *Ann. Statist.* 22.4 (1994), pp. 1701–1762.

[Tie98] L. Tierney. "A note on Metropolis-Hastings kernels for general state spaces". In: *Ann. Appl. Probab.* 8 (1998), pp. 1–9.

[TT90] D. Talay and L. Tubaro. "Expansion of the global error for numerical schemes solving stochastic differential equations". In: *Stochastic Anal. Appl.* 8.4 (1990), 483–509 (1991). ISSN: 0736-2994.

[TT94] P. Tuominen and R. L. Tweedie. "Subgeometric Rates of Convergence of f-Ergodic Markov Chains". In: *Advances in Applied Probability* 26.3 (1994), pp. 775–798. ISSN: 00018678.

[Ver97] A. Yu. Veretennikov. "On polynomial mixing bounds for stochastic differential equations". In: *Stochastic Process. Appl.* 70.1 (1997), pp. 115–127. ISSN: 0304-4149.

[Vil09] C. Villani. *Optimal transport : old and new*. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 2009. ISBN: 978-3-540-71049-3.

[Wol14] Inc. Wolfram Research. "Mathematica". In: (2014).

[WPS13] J. Windle, N. G. Polson, and J. G. Scott. *BayesLogit: Bayesian logistic regression*. http://cran.r-project.org/web/packages/BayesLogit/index.html R package version 0.2. 2013.

List of Figures

3.1 Evolution of the rate of convergence ρ given by Theorem 3.14 in function of the dimension d. 71

3.3 Trace plot and auto-correlation in function of the dimension on 10000 iterations with a 10000 burn-in iterations 73

3.5 Trace plot and auto-correlation in dimension 100 for EI-MALA, pCN and MALA on 50000 iterations with a 10000 burn-in iterations. 74

6.1 Empirical distribution comparison between the Polya-Gamma Gibbs Sampler and ULA. Left panel: constant step size $\gamma_k = \gamma_1$ for all $k \geq 1$; right panel: decreasing step size $\gamma_k = \gamma_1 k^{-1/2}$ for all $k \geq 1$ 170

6.2 Upper left: German credit data set. Upper right: Australian credit data set. Lower left: Heart disease data set. Lower right: Prima Indian diabetes data set 171

7.1 Expected square jumped distance for the beta distribution with parameters $a_1 = 10$ and $a_2 = 10$ as a function of the mean acceptance rate for $d = 10, 50, 100$. 214

8.1 First-order efficiency of the new fMALA and the standard MALA for the double well potential $g(x) = -\frac{1}{4}x^4 + \frac{1}{2}x^2$, as a function of the overall acceptance rates in dimensions $d = 10, 100, 500, 1000$. The solid line is the reference asymptotic curve of efficiency for the new fMALA, normalised to have the same maximum value as the finite dimensional fMALA. . . . 266

8.2 Trace plots of $\|X\|^2$ for the Gaussian target density in dimension $d = 1000$ when starting at the origin. Comparison of fMALA with $h \sim d^{-1/5}$ (solid lines), MALA with $h \sim d^{-1/3}$ (dashed lines), RWM with $h \sim d^{-1}$ (dotted lines). 267

8.3 Auto-correlation versus LAG for the Gaussian target density in dimension $d = 1000$. Comparison of fMALA with $h \sim d^{-1/5}$ (black), MALA with $h \sim d^{-1/3}$ (white), RWM with $h \sim d^{-1}$ (gray). 268

8.4 First-order efficiency of the new fMALA and the standard MALA as a function of the overall acceptance rates for the dimensions $d = 100$ (white points), $d = 500$ (gray points), $d = 1000$ (dark points), respectively, for the non product density (8.45) with $\alpha(x) = x/2$ (top pictures) and $\alpha(x) = \sin(x)$ (bottom pictures). 269

List of Tables

4.1 Rates of convergence when ϕ increases at a logarithmic rate, a polynomial rate and a subexponential rate, obtained from Theorem 4.3 and from [But14, Theorem 2.1] and [Dou+04, Section 2.3.]. 88

5.1 For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Corollary 5.19 129
5.2 For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Theorem 5.20 130
5.3 For constant step sizes, dependency of γ and p in d, ε and parameters of U to get $\|\delta_x Q_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ using Theorem 5.22 132

6.1 Order of convergence of $W_2^2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**10 and **H**11 158
6.2 Order of convergence of $W_2^2(\delta_x Q_\gamma^n, \pi)$ for $\gamma_k = \gamma_1 k^{-\alpha}$ under **H**10, **H**11 and **H**12 158
6.3 Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 6.4 159
6.4 Order of the optimal choice of γ_1 for the fixed horizon setting and implied bound on $W_2^2(\delta_x Q_\gamma^n, \pi)$ based on Theorem 6.7 159
6.5 Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**10 and **H**11 161
6.6 Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed γ_1 and N under **H**10, **H**11 and **H**12 161
6.7 Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**10, **H**11 and **H**12 161
6.8 Bound for the MSE for $\gamma_k = \gamma_1 k^{-\alpha}$ for fixed n under **H**10 and **H**11 . . . 162
6.9 Order of convergence of $\|\delta_x Q_\gamma^n - \pi\|_{\mathrm{TV}}$ for $\gamma_k = \gamma_1 k^{-\alpha}$ based on Theorem 6.14 164
6.10 Order of the bound between π and π_γ in total variation function of the step size $\gamma > 0$ and the dimension d. 165
6.11 Order of the step size $\gamma > 0$ and the number of iterations $p \in \mathbb{N}^*$ to get $\|\delta_x R_\gamma^p - \pi\|_{\mathrm{TV}} \leq \varepsilon$ for $\varepsilon > 0$ 165
6.12 Dimension of the data sets 170

8.1 Summary of ergodicity results for the Metropolis-Hastings algorithms for the class $\mathscr{E}(\beta, \gamma)$ 263
8.2 Summary of ergodicity results for the unadjusted proposals for the class $\mathscr{E}(\beta, \gamma)$. 265

B.1 Dependencies of γ X
B.2 Dependencies of n X
B.3 Dependencies of γ XII
B.4 Dependencies of n XII