



HAL
open science

Contributions to the design, study and implementation of Markov chain Monte Carlo methods applied to Bayesian inference

Aurélien Enfroy

► **To cite this version:**

Aurélien Enfroy. Contributions to the design, study and implementation of Markov chain Monte Carlo methods applied to Bayesian inference. Computational Geometry [cs.CG]. Institut Polytechnique de Paris, 2022. English. NNT : 2022IPPAS012 . tel-03752067

HAL Id: tel-03752067

<https://theses.hal.science/tel-03752067>

Submitted on 16 Aug 2022

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

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



INSTITUT
POLYTECHNIQUE
DE PARIS

NNT : 2022IPPAS012

Thèse de doctorat

TELECOM
SudParis



IP PARIS

Contributions à la conception, l'étude et la mise en oeuvre de méthodes de Monte Carlo par chaîne de Markov appliquées à l'inférence bayésienne

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à Télécom SudParis

École doctorale n°574 École doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 15 Juillet 2022, par

AURÉLIEN ENFROY

Composition du Jury :

Tony Lelièvre Professeur, École des Ponts ParisTech (CERMICS)	Président
Yves Atchade Professeur, Boston University	Rapporteur
Florent Malrieu Professeur, Université de Tours (Institut Denis Poisson)	Rapporteur
Nicolas Chopin Professeur, ENSAE (CREST)	Examineur
Dasha Loukianova Maîtresse de conférences, Université d'Évry Val d'Essonne (LaMME)	Examineur
Randal Douc Professeur, Télécom SudParis (SAMOVAR)	Directeur de thèse
Alain Durmus Maître de conférence, ENS Paris-Saclay (Centre Borelli)	Invité

Remerciements

Je souhaite tout d'abord remercier mes directeurs de thèse Randal Douc et Alain Durmus pour leur soutien et le temps qu'ils m'ont consacré.

Je remercie Florent Malrieu et Yves Atchade d'avoir accepté de rapporter ma thèse, ainsi que Dasha Loukianova, Nicolas Chopin et Tony Lelièvre d'avoir accepté de faire partie du jury.

Je tiens à remercier mes co-auteurs, Jimmy Olsson, Éric Moulines, Gabriel Stoltz, Andreas Eberle, Arnaud Guillin et Pierre Monmarché.

Je tiens enfin à remercier ma famille pour leur soutien inébranlable.

Contents

1	Introduction	1
1.1	Algorithmes de Monte Carlo par chaînes de Markov	1
1.1.1	Application de l'échantillonnage à l'inférence bayésienne	1
1.1.2	Algorithmes de Monte Carlo par chaînes de Markov	3
1.1.3	L'algorithme de Metropolis Hastings et ses dérivés	4
1.2	Convergence des chaînes de Markov et des méthodes MCMC associées	6
1.3	Les diffusions de Langevin sous/sur-amorties et leurs discrétisations	11
1.3.1	La diffusion de Langevin sur-amorties: discrétisation et processus collant associé	11
1.4	Résumé synthétique des contributions	17
1.4.1	Processus collant discret	17
1.4.2	Convergence des discrétisations du processus de Langevin sous-amorti	19
1.4.3	L'algorithme de téléportation Kick-Kac	21
2	Boost your favorite Markov Chain Monte Carlo sampler using Kac's theorem: the Kick-Kac teleportation algorithm	23
2.1	Introduction	24
2.2	Kac's formula	26
2.3	The Kick-Kac teleportation process	29
2.3.1	The memoryless Kick-Kac teleportation process	29
2.3.2	The Kick-Kac teleportation process	31
2.3.3	A law of large numbers for the KKT process	34
2.3.4	Geometric ergodicity of the KKT process	38
2.4	A general Kick-Kac teleportation process	41
2.4.1	A generalization of Kac's theorem	41
2.4.2	Revisiting the hybrid kernel of [20]	44
2.4.3	Revisiting the Metropolis–Hastings algorithm	45
2.5	Numerical illustrations	46
2.5.1	Two multi-modal distributions	48
2.5.2	Stochastic volatility model	49
2.5.3	The Ginzburg–Landau model	51

2.6	Kernel notation	52
2.7	Proofs and technical lemmas	52
2.7.1	Proof of Proposition 2.4	52
2.7.2	Proof of Proposition 2.8	54
2.7.3	Technical results for proving Theorem 2.7 and Theorem 2.9	54
3	Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics	59
3.1	Introduction	60
3.2	Setting and main results	62
3.2.1	Structural assumptions on the numerical schemes	62
3.2.2	Examples of numerical schemes of the form (3.2) satisfying A1-A2	64
3.2.3	Minorization, drift and convergence uniform in the timestep	69
3.3	Strategy of proof of Theorem 3.1 in a simple case	74
3.4	Proof of the minorization condition	77
3.4.1	Algebraic structure of the equations	77
3.4.2	Structure and properties of the noise	79
3.4.3	Stability estimates	85
3.4.4	Proof of Theorem 3.1	94
3.5	Proof of Theorem 3.4	97
3.5.1	Conditioned drift inequality	97
3.5.2	Proof of Theorem 3.4	101
3.5.3	Supporting lemmas	103
3.6	Postponed proofs	104
3.6.1	Proof of Lemma 6	104
3.6.2	Proof of Lemma 8	105
3.6.3	Proof of Lemma 16	105
3.6.4	Proof of Lemma 17	106
3.6.5	Proof of Lemma 18	106
3.6.6	Proof of Lemma 19	107
3.6.7	Proof of Proposition 3.2	107
3.7	Complementary and technical results	109
4	Discrete sticky couplings of functional autoregressive processes	113
4.1	Introduction	114
4.2	Sticky reflection coupling	116
4.2.1	Main result	116
4.2.2	The discrete sticky kernel	119
4.2.3	Analysis of the auxiliary Markov chain	121
4.3	Continuous-time limit	124
4.4	An application in Bayesian statistics: parameter estimation in an ODE	127
4.4.1	Setting and verifying the assumptions	127
4.4.2	Numerical results	131
4.5	Postponed proofs	133

Contents

4.5.1	Postponed proofs of Section 4.2	133
4.5.2	Postponed proofs of Section 4.3	150
4.5.3	Postponed proofs of Section 4.4	160
Bibliographie		162

CHAPTER 1

Introduction

Mes travaux de thèse portent sur les méthodes de Monte Carlo par chaîne de Markov et leurs applications à l'inférence bayésienne. C'est donc tout naturellement que j'introduis ici ces deux domaines ce qui m'amènera à introduire les principaux problèmes auxquels j'ai tâché de répondre dans mes contributions.

1.1 Algorithmes de Monte Carlo par chaînes de Markov

Les méthodes de Monte Carlo forment maintenant une classe d'algorithmes très populaires dont la finalité est d'approximer des intégrales $\pi(f) = \int f d\pi$, où $f : X \rightarrow \mathbb{R}$ est une fonction intégrable par rapport à une distribution π , dite distribution cible, sur un espace mesurable (X, \mathcal{X}) . En effet, les théorèmes limites usuels justifient d'estimer cette quantité à partir d'un échantillon $\{X_i\}_{i=1}^n$ indépendantes et identiquement distribuées (i.i.d.) de loi π par

$$\pi(f) \approx \hat{\pi}_n^{\text{MC}}(f) = n^{-1} \sum_{i=1}^n f(X_i).$$

Cependant, comme nous allons le voir tout de suite, dans la plupart des applications qui nous intéressent, π bien qu'admettant une densité par rapport à une mesure de référence σ -finie, cette densité n'est connue que à une constante de normalisation près.

1.1.1 Application de l'échantillonnage à l'inférence bayésienne

Nous présentons ici succinctement les statistiques bayésiennes. Nous renvoyons par exemple à [59, 93, 118] pour de nombreuses ressources et références sur le sujet. Notons que le problème de l'échantillonnage d'une distribution ne se limite pas aux statistiques bayésiennes. Cependant, l'une des motivations de cette thèse est de développer et d'analyser des algorithmes d'échantillonnage pour l'appliquer à des problèmes issus de ce domaine ce qui est un sujet suffisamment vaste pour illustrer nos résultats.

L'inférence bayésienne d'observations Y à valeurs dans un espace mesurable (Y, \mathcal{Y}) , est basée sur un modèle probabiliste $(Y, \mathcal{Y}, (\mathbb{P}_\theta)_{\theta \in \Theta})$. Pour simplifier l'exposition, nous

supposons ici que Θ est un mesurable de \mathbb{R}^d et que le modèle est dominé par une mesure sigma finie μ_Y sur (Y, \mathcal{Y}) . Notons la vraisemblance du modèle $\vartheta \mapsto f_\vartheta(y) = [d\mathbb{P}_\vartheta/d\mu_Y](y)$. De plus, nous supposons également que $(\vartheta, y) \mapsto f_\vartheta(y)$ est bi-mesurable sur l'espace mesurable produit $(\Theta \times Y, \mathcal{B}(\Theta) \otimes \mathcal{Y})$. Un exemple de base est le modèle gaussien où Y est un vecteur de \mathbb{R}^n dont chaque composante est indépendante et identiquement distribuée de loi gaussienne unidimensionnelle de moyenne $\vartheta \in \mathbb{R}$ et de variance $\sigma^2 > 0$ fixée. Alors le modèle est dominé par la mesure de Lebesgue et admet pour vraisemblance, les fonctions définies pour tout $y \in \mathbb{R}^n$ et $\vartheta \in \mathbb{R}$ par

$$f_\vartheta(y) = (2\pi\sigma^2)^{-n/2} \prod_{i=1}^n \exp\left(-\frac{(y_i - \vartheta)^2}{2\sigma^2}\right),$$

où pour $y \in \mathbb{R}^n$, y_i est la i -ème coordonnée de y dans la base canonique. Alors que en statistique fréquentiste, on inférerait le paramètre ϑ en maximisant la vraisemblance, les statistiques bayésiennes considèrent le paramètre ϑ comme lui-même une réalisation d'une variable aléatoire θ . Ainsi, à partir d'une loi de probabilité a priori, que l'on notera Π_0 , la loi jointe du modèle, notée Π , qui est la distribution de (θ, Y) sur $(\Theta \times Y, \mathcal{B}(\Theta) \otimes \mathcal{Y})$, est définie pour tout $A \in \mathcal{B}(\Theta) \otimes \mathcal{Y}$ par

$$\Pi(A) = \int_{\Theta \times Y} \mathbb{1}_A(\vartheta, y) f_\vartheta(y) \Pi_0(d\vartheta) \mu_Y(dy),$$

et la loi marginal Π_Y de Y est définie pour tout $\tilde{A} \in \mathcal{Y}$ par

$$\Pi_Y(\tilde{A}) = \int_{\Theta \times Y} \mathbb{1}_{\Theta \times \tilde{A}}(\vartheta, y) f_\vartheta(y) \Pi_0(d\vartheta) \mu_Y(dy).$$

Cette distribution admet une densité par rapport à μ_Y :

$$f_Y(y) = \int_{\Theta} f_\vartheta(y) \Pi_0(d\vartheta).$$

L'expression de la loi conditionnelle de θ sachant Y , appelée loi a posteriori, découle alors du théorème de Bayes [124, Théorème 1.31]. La loi a posteriori que l'on notera Π_{post} sur $(\Theta, \mathcal{B}(\Theta))$ admet une densité par rapport à Π_0 qui est donnée pour Π_0 -presque tout ϑ par

$$\Pi_{\text{post}}(\vartheta) = \frac{f_\vartheta(Y)}{f_Y(Y)}. \tag{1.1}$$

La théorie de la décision permet alors de construire des estimateurs dits de Bayes ont qui des propriétés d'optimalité très intéressantes. Ces estimateurs sont le plus souvent sous la forme d'intégrale par rapport à la loi a posteriori du modèle. Par exemple la moyenne a posteriori $\int_{\Theta} \vartheta \Pi_{\text{post}}(d\vartheta)$ est un estimateur de Bayes pour le paramètre θ . Mais on peut souhaiter obtenir plus d'informations sur la loi a posteriori, comme ses quantiles, on peut généraliser ces problèmes en disant qu'il s'agit de calculer des intégrales par rapport à Π_{post} , $\int g(\vartheta) \Pi_{\text{post}}(d\vartheta)$. Les algorithmes de Monte Carlo permettent d'estimer cette quantité. Cependant, on peut observer que la densité a posteriori (1.1) n'est connue qu'à une constante multiplicative près car $f_Y(Y)$ est la plupart du temps inconnu. En effet, sauf dans le cas de lois dites conjuguées [120, section 3.3], f_Y , la densité marginale, n'a pas de forme analytique.

1.1.2 Algorithmes de Monte Carlo par chaînes de Markov

Le fait de ne connaître la densité de la loi cible qu'à une constante de normalisation près rend impossible, la plupart du temps, l'utilisation des méthodes d'échantillonnage simple comme la méthode par inversion ou par rejet. Pour répondre à ce problème, des version auto-normalisée de l'échantillonnage préférentiel peuvent être mise à l'oeuvre mais il est connu que ce type d'algorithmes n'est pas efficaces lorsque X est un espace euclidien de grande dimension ou un ensemble fini de cardinal grand. Une autre solution est de remplacer les échantillons i.i.d. de loi π par une chaîne de Markov associée à un noyau de Markov K ayant pour loi stationnaire, la loi visée π . Nous rappelons rapidement la définition des notions en jeu.

Un noyau de Markov K est une application de $\mathsf{X} \times \mathcal{X}$ dans $[0, 1]$ vérifiant que 1) $x \mapsto K(x, \mathsf{A})$ est $\mathcal{X}/\mathcal{B}(\mathbb{R})$ -mesurable pour tout $\mathsf{A} \in \mathcal{X}$ et 2) $\mathsf{A} \mapsto K(x, \mathsf{A})$ est une mesure de probabilité sur $(\mathsf{X}, \mathcal{X})$ pour tout $x \in \mathsf{X}$. Pour toute fonction mesurable $f : \mathsf{X} \rightarrow \mathbb{R}$, on définira dès que cela fait sens $Kf(x) = \int K(x, dy)f(y)$ et pour toute mesure de probabilité μ , $\mu K(\mathsf{A}) = \int \mu(dx)K(x, \mathsf{A})$. Enfin pour tous noyaux K_1, K_2 sur $\mathsf{X} \times \mathcal{X}$ nous définissons le noyau $K_1 K_2(x, \mathsf{A}) = \int K_1(x, dy)K_2(y, \mathsf{A})$.

Le processus $(Y_i)_{i \in \mathbb{N}}$ est une chaîne de Markov à valeurs dans $(\mathsf{X}, \mathcal{X})$ sur un espace filtré $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{i \in \mathbb{N}}, \mathbb{P})$ si il existe un noyau de Markov K sur $\mathsf{X} \times \mathcal{X}$ tel que $\mathbb{P}(Y_{i+1} \in \mathsf{A} | \mathcal{F}_i) = K(Y_i, \mathsf{A})$ pour tout $i \in \mathbb{N}$ et $\mathsf{A} \in \mathcal{X}$. Étant donné un noyau de Markov K et une loi initiale μ_0 sur $(\mathsf{X}, \mathcal{X})$ il est toujours possible de construire une mesure sur l'espace canonique $(\mathsf{X}^{\mathbb{N}}, \mathcal{X}^{\otimes \mathbb{N}})$ telle que les applications canoniques $Y_i : (y_i)_{i \in \mathbb{N}} \mapsto y_i$ forment une chaîne de Markov de noyau K et de loi initiale μ_0 . Cette mesure sera notée \mathbb{P}_{μ_0} et son espérance \mathbb{E}_{μ_0} . Dans le cas où $\mu_0 = \delta_x$, la mesure de Dirac à $x \in \mathsf{X}$, on notera simplement \mathbb{P}_x et \mathbb{E}_x .

La loi cible π est dite invariante pour K si $\pi K = \pi$. De plus, si π est l'unique loi invariante pour K , alors pour π -presque tout $x \in \mathsf{X}$, \mathbb{P}_x -presque sûrement

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{i=1}^n f(Y_i) = \pi(f), \quad (1.2)$$

Sous la condition que P est Harris récurrent, (1.2) est aussi valable pour tout $x \in \mathsf{X}$. Cette condition supplémentaire signifie simplement que pour tout $\mathsf{A} \in \mathcal{X}$, et $x \in \mathsf{A}$,

$$\mathbb{P}_x \left(\sum_{i=1}^{\infty} \mathbb{1}_{\mathsf{A}}(Y_i) = \infty \right) = 1.$$

Ces résultats justifie de remplacer les échantillons i.i.d. de loi π par une chaîne de Markov associée à K . Le problème qui se pose alors est de construire un noyau K ayant π pour loi invariante. Les méthodes de Monte Carlo par chaîne de Markov (MCMC) répondent justement à ce problème : ils consistent en des procédés d'échantillonnage de chaîne de Markov ayant une loi cible ayant le plus souvent une densité par rapport à une mesure de référence connu uniquement à une constante de normalisation près. De plus, bien que la loi forte des grands nombres soit un résultat déjà très intéressant, elle ne permet pas vraiment de quantifier la convergence de la chaîne de Markov $(Y_i)_{i \in \mathbb{N}}$ ou de son

noyau vers π . Cependant, avant de rentrer plus en détail sur l'étude des chaînes de Markov et de leur convergence, nous donnons dans la suite quelques exemples importants d'algorithmes MCMC.

1.1.3 L'algorithme de Metropolis Hastings et ses dérivés

Un premier algorithme MCMC qui est maintenant considéré comme l'une des méthodes numériques ayant eu le plus d'impact est l'algorithme de Metropolis-Hastings. Il a été proposé dans [98], et généralisé dans [61] et [129]. Ce dernier se base sur un noyau de Markov Q sur $(\mathbf{X}, \mathcal{X})$, appelé noyau de proposition et une fonction $\alpha_H : \mathbf{X}^2 \rightarrow [0, 1]$, dit taux d'acceptation pour définir un noyau de transition P sur $(\mathbf{X}, \mathcal{X})$ pour tout $x \in \mathbf{X}$ et $A \in \mathcal{X}$,

$$P(x, A) = \int_A \alpha_H(x, y)Q(x, dy) + \delta_x(A) \int_{\mathbf{X}} (1 - \alpha_H(x, y))Q(x, dy), \quad (1.3)$$

L'échantillonnage d'une chaîne de noyau P se fait de manière simple dans le cas où on peut simuler suivant Q . Étant donné X_n , un candidat Y_{n+1} de loi $Q(X_n, \cdot)$ et indépendant du passé est généré ainsi qu'une variable uniforme U_{n+1} sur $[0, 1]$ de manière indépendante. Ensuite, si $U_{n+1} \leq \alpha_H(X_n, Y_{n+1})$, alors le candidat est dit accepté et on pose $X_{n+1} = Y_{n+1}$ sinon il est dit rejeté et on pose $X_{n+1} = X_n$. Il est aisé de montrer que le processus $(X_n)_{n \in \mathbb{N}}$ ainsi défini est bien une chaîne de Markov ayant pour noyau P donné par (1.3). Cependant, pour que P ait pour loi invariante π , étant donné Q , il convient de choisir convenablement α_H . Un critère qui peut être plus simple à vérifier que l'invariance est la réversibilité de P par rapport à π : pour toute fonction $f : \mathbf{X}^2 \rightarrow \mathbb{R}$ mesurable bornée,

$$\int_{\mathbf{X}^2} f(x, y)\pi(dx)P(x, dy) = \int_{\mathbf{X}^2} f(x, y)\pi(dy)P(y, dx), \quad (1.4)$$

C'est sur cette condition que se base l'algorithme de Metropolis-Hastings comme nous allons le voir.

Définissons les deux mesures sur $(\mathbf{X}^2, \mathcal{X}^{\otimes 2})$, μ_H et μ_H^T pour tout $A \in \mathcal{X}^{\otimes 2}$ par

$$\mu_H(A) = \int_{\mathbf{X} \times \mathbf{X}} \mathbb{1}_A(y, z)\pi(dy)Q(y, dz), \quad \mu_H^T(A) = \int_{\mathbf{X} \times \mathbf{X}} \mathbb{1}_A(z, y)\pi(dy)Q(y, dz).$$

Il est aisé de remarquer que P défini par (1.3) est réversible par rapport à π si et seulement si

$$\int_{\mathbf{X}^2} f(x, y)\alpha_H(x, y)\mu_H(dx, dy) = \int_{\mathbf{X}^2} f(x, y)\alpha_H(y, x)\mu_H^T(dx, dy).$$

Cette identité est la clé de [129, Theorem 3] qui établit que P est réversible par rapport à π si et seulement si il existe $S \in \mathcal{X}^{\otimes 2}$, qui soit symétrique, tel que $\alpha(x, y) = 0$ pour tout $(x, y) \in S^c$, $\mu_H(\cdot \cap S)$ soit absolument continue par rapport à $\mu_H^T(\cdot \cap S)$ de densité φ_H satisfaisant μ_H -presque partout

$$\alpha_H(x, y)\varphi_H(x, y) = \alpha_H(y, x). \quad (1.5)$$

On dira que le noyau P est de type Metropolis-Hastings si il est sous la forme (1.3) et vérifie (1.5). Trois exemples importants sont les suivants.

(1) Supposons qu'il existe une mesure ν sur (X, \mathcal{X}) telle que π admet une densité encore notée π par rapport à ν et Q admet une densité de transition q par rapport à ν . De plus définissons

$$S = \left\{ (x, y) \in X^2 : \pi(x)q(x, y)\pi(y)q(y, x) \neq 0 \right\},$$

et

$$\alpha_H(x, y) = \begin{cases} \min \left(1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \right) & \text{pour } (x, y) \in S \\ 0 & \text{sinon,} \end{cases}$$

la condition (1.5) pour que P soit réversible est satisfaite avec

$$\varphi_H(x, y) = \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)},$$

défini sur S .

Un choix simple pour le noyau de proposition Q est de prendre simplement une distribution de densité q par rapport à ν . Alors les variables aléatoires proposées au cours de l'algorithme constituent une suite i.i.d. distribuée suivant ν . Observons alors que dans ce cas,

$$\alpha_H(x, y) = \min(1, \pi(y)q(x)/\pi(x)q(y)).$$

La méthode correspondant à ce choix de α_H et Q est appelée l'algorithme de Metropolis-Hastings indépendant. Une autre possibilité pour le noyau de proposition Q est celui d'une marche aléatoire lorsque $(X, +)$ est un groupe et que ν est invariante par translation. Considérons, une loi de densité \tilde{q} , symétrique $\tilde{q}(x) = \tilde{q}(-x)$, par rapport à ν . Dans ce cas, on peut considérer le noyau de proposition

$$Q(x, A) = \int_A \tilde{q}(x - y) d\nu(y).$$

La simulation suivant $Q(x, \cdot)$ est simple, puisque elle consiste simplement à simuler X suivant \tilde{q} et de considérer $x + X$. Comme \tilde{q} est supposée symétrique α_H est sous la forme

$$\alpha_H(x, y) = \min(1, \pi(y)/\pi(x)).$$

Cette méthode est appelée l'algorithme de Metropolis à marche aléatoire symétrique. Un exemple concret de l'algorithme de Metropolis à marche aléatoire symétrique est dans le cas $X = \mathbb{R}^d$ muni de la mesure de Lebesgue et où

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

la proposition est la marche aléatoire avec incréments gaussiens centrés et matrice de covariance σId , $\sigma > 0$.

(2) Si π admet une densité positive par rapport à une mesure ν , notée π , et Q est réversible par rapport à ν , alors en posant $S = X$ et pour tout $(x, y) \in X^2$,

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

(1.3) définit toujours un noyau de Metropolis-Hastings réversible par rapport à π . Notons que dans ce cadre contrairement à (1), il n'est pas nécessaire que Q soit dominé par rapport à ν .

(3) Une dernière possibilité que nous présentons ici est de considérer des noyaux déterministes $Q(x, A) = \delta_{\Phi(x)}(A)$, dans le cas $X = \mathbb{R}^d$ et où $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est un C^1 -difféomorphisme involutif. Alors, on peut montrer que la condition (1.5) est satisfaite avec

$$\alpha_H(x, y) = \min \left(1, \frac{\pi(\Phi(x)) |\text{Jac}_\Phi(x)|}{\pi(x)} \right),$$

où Jac_Φ et le déterminant jacobien de Φ . Cette construction est à la base d'un algorithme MCMC très populaire, l'algorithme Hamiltonian Monte Carlo.

Malgré le succès indéniable des méthodes MCMC, l'un des obstacles majeurs auxquels elles sont confrontées est le problème de leur convergence. Plus précisément, comme nous allons le voir, il est important de quantifier la convergence des noyaux de Markov associés aux méthodes MCMC vers la loi cible pour être capable de dériver des bornes de convergence similaires aux méthodes de Monte Carlo classiques.

1.2 Convergence des chaînes de Markov et des méthodes MCMC associées

Ici nous nous intéressons pour simplifier la présentation à l'erreur quadratique moyenne des méthodes MCMC associée à un noyau de Markov K définie par

$$\text{MSE}_n(f) = \mathbb{E}_{\mu_0} \left[\left(n^{-1} \sum_{i=1}^n \{f(Y_i) - \pi(f)\} \right)^2 \right].$$

Cette quantité peut alors se décomposer aisément suivant un terme de biais et de variance

$$\begin{aligned} \text{MSE}_n(f) &= B_n^2(f) + V_n(f), \\ B_n(f) &= n^{-1} \sum_{i=1}^n \mu_0 K^i f - \pi(f) \quad , \quad V_n(f) = \text{Var}_{\mu_0} \left(n^{-1} \sum_{i=1}^n f(Y_i) \right). \end{aligned} \tag{1.6}$$

Contrairement aux méthodes de Monte Carlo "classique" qui sont sans biais, les méthodes MCMC sont biaisées $B_n(f) \neq 0$ et pour analyser ce terme, la convergence du noyau K vers π doit être étudiée.

Pour établir la convergence de la séquence de mesures de probabilité $(\mu_0 K^k)_{k \in \mathbb{N}}$ vers π pour toute mesure de probabilité initiale μ_0 , on peut utiliser plusieurs distances sur l'ensemble des mesures de probabilité sur $(\mathsf{X}, \mathcal{X})$ noté $\mathcal{P}(\mathsf{X})$, ou un de ses sous-ensembles. Ces dernières sont le plus souvent définies comme

$$(\mu, \nu) \mapsto \sup_{f \in \mathcal{F}} |\mu(f) - \nu(f)|,$$

pour une classe de fonction \mathcal{F} de X dans \mathbb{R} , ou de manière duale

$$\mathbf{W}_{\mathbf{c}}(\mu, \nu) = \inf_{\xi \in \Pi(\mu, \nu)} \int_{\mathsf{X} \times \mathsf{X}} \mathbf{c}(x, y) d\xi(x, y), \quad (1.7)$$

où $\mathbf{c} : \mathsf{X}^2 \rightarrow \mathbb{R}_+$ est une fonction bi-mesurable, symétrique telle que $\mathbf{c}(x, y) = 0$ si et seulement si $x = y$ et $\Pi(\mu, \nu)$ est l'ensemble des plans de transfert de μ et ν , c'est à dire l'ensemble des mesures de probabilité sur $(\mathsf{X} \times \mathsf{X}, \mathcal{X} \otimes \mathcal{X})$ telles que pour tout ensemble mesurable $\mathbf{A} \in \mathcal{X}$, $\zeta(\mathbf{A} \times \mathsf{X}) = \mu(\mathbf{A})$ et $\zeta(\mathsf{X} \times \mathbf{A}) = \nu(\mathbf{A})$. Nous appelons $\mathbf{W}_{\mathbf{c}}$ la distance de Wasserstein associée au coût \mathbf{c} . Observons que la plupart du temps la distance de Wasserstein n'est pas une métrique mais simplement une semi-métrique. De plus, si on considère

$$\mathcal{F}_{\mathbf{c}} = \{f : \mathsf{X} \rightarrow \mathbb{R} : [f]_{\mathbf{c}} < +\infty\}, \quad [f]_{\mathbf{c}} = \sup_{x, y \in \mathsf{X}, x \neq y} \frac{|f(x) - f(y)|}{\mathbf{c}(x, y)},$$

Alors pour tout $f \in \mathcal{F}_{\mathbf{c}}$,

$$|\mu(f) - \nu(f)| \leq [f]_{\mathbf{c}} \mathbf{W}_{\mathbf{c}}(\mu, \nu).$$

Ainsi la convergence d'un noyau K en la semi-métrique $\mathbf{W}_{\mathbf{c}}$ permet de borner le terme de biais $B_n^2(f)$ dans la décomposition de l'erreur quadratique de manière uniforme sur la classe de fonction $\mathcal{F}_{\mathbf{c}}$. Ainsi si pour une loi initiale μ_0 , $\mathbf{W}_{\mathbf{c}}(\mu_0 K^k, \pi) \leq C(\mu_0) u_k$ avec $\sum_k u_k < +\infty$, $B_n^2(f)$ est d'ordre n^{-2} . Nous donnons maintenant deux exemples fondamentaux de classes de fonctions de coût qui mènent à des métriques très utilisées en probabilités et statistiques.

Soit $V : \mathsf{X} \rightarrow [1, \infty)$ est une fonction mesurable. Nous définissons la V -norme entre deux mesures de probabilité μ et ν comme

$$\|\mu - \nu\|_V = \sup_{\|f\|_V \leq 1} |\mu(f) - \nu(f)|, \quad \text{où } \|f\|_V = \sup_{\mathsf{X}} \{|f|/V\}. \quad (1.8)$$

Si $V \equiv 1$, alors $\|\cdot\|_V$ est la variation totale que l'on note par $\|\cdot\|_{\text{TV}}$. Il peut être établi que les V -normes correspondent aux distances de Wasserstein associées aux coûts $(x, y) \mapsto \mathbb{1}_{\Delta_{\mathsf{X}}^c}(x, y) \{V(x) + V(y)\}/2$ où $\Delta_{\mathsf{X}} = \{(x, x) : x \in \mathsf{X}\}$.

La convergence des chaînes de Markov a été étudiée de manière exhaustive en utilisant la distance de variation totale et la V -norme [99], en particulier, introduisons ici la notion d'ergodicité géométrique uniforme V afin de présenter un exemple concret de bornes de convergence, et parce que certains des résultats de cette thèse sont exprimés à travers cette notion. Soit $V : \mathsf{X} \rightarrow [1, \infty)$ est une fonction mesurable. On dit que le noyau

de Markov K , de loi invariante π , est V -uniformément géométriquement ergodique s'il existe $C > 0$ et $\rho \in [0, 1)$ tels que pour tout $x \in \mathsf{X}$ et $n \in \mathbb{N}$,

$$\|\delta_x K^n - \pi\|_V \leq C \rho^n V(x). \quad (1.9)$$

Il s'agit d'une notion forte de convergence et elle peut être assouplie de multiples façons. Nous renvoyons à [37, 99] pour une présentation plus développée des résultats de convergence sur les chaînes de Markov. Pour prouver (1.9), deux conditions doivent être vérifiées. Tout d'abord, on doit pouvoir s'assurer de l'existence d'un ensemble $\mathsf{C} \in \mathcal{X}$ dit *small*, c'est à dire vérifiant pour un $\epsilon > 0$ et pour tout $x, x' \in \mathsf{C}$,

$$\|K(x, \cdot) - K(x', \cdot)\|_{\text{TV}} \leq 1 - \epsilon. \quad (1.10)$$

Cette condition assure que à l'intérieur de cet ensemble C , on puisse coupler deux chaînes de Markov de noyau K , partant de deux points différents dans C . Pour contrôler le temps passé en dehors de C , nous pouvons utiliser la condition de dérive : il existe $\lambda \in [0, 1)$, $b \geq 0$ tels que

$$KV \leq \lambda V + b \mathbb{1}_{\mathsf{C}}. \quad (1.11)$$

Sous certaines conditions sur C, λ, b (1.10) et (1.11) permettent d'établir (1.9).

Un autre type de métriques maintenant populaire sont celles qui sont basées sur les distance de Wasserstein à poids ; cf. par exemple [37, Section 20.7] pour des références sur le sujet. Plus exactement, cela consiste à considérer des semi-métriques définies par (1.7) avec $\mathbf{c} = \mathbf{d}_{\tilde{V}}(x, y)$ où

$$\mathbf{d}_{\tilde{V}}(x, y) = \mathbf{d}^p(x, y) \tilde{V}(x, y),$$

$p \geq 1$, $\tilde{V} : \mathsf{X}^2 \rightarrow [1, +\infty)$ et $\mathbf{d} : \mathsf{X} \rightarrow \mathsf{X}$ est une distance sur X qui le rend séparable et complet pour la topologie associée et tel que \mathcal{X} est la tribu borélienne correspondante. Dans le cas $\tilde{V} \equiv 1$, $\mathsf{X} = \mathbb{R}^d$ et \mathbf{d} est la distance euclidienne, nous retrouvons la distance de Wasserstein d'ordre p usuelle. De plus, l'ensemble des fonctions vérifiant $[f]_{\mathbf{d}_{\tilde{V}}} < +\infty$ correspond à l'ensemble des fonctions Lipschitz à poids pour la distance \mathbf{d} . Remarquons que dans le cas où $\mathbf{d}_{\tilde{V}}$ est semi-continue inférieurement, par exemple lorsque \tilde{V} est continue, [131, Théorème 4.1] assure que l'infimum dans la distance de Wasserstein (1.7) associée à $\mathbf{d}_{\tilde{V}}$ par un plan de transfert $\zeta^* \in \Pi(\mu, \nu)$. De même que dans le cas de la V -norme, on dit que le noyau de Markov K , de loi invariante π , est uniformément géométriquement ergodique pour $\mathbf{W}_{\mathbf{d}_{\tilde{V}}}$ s'il existe $C > 0$ et $\rho \in [0, 1)$ tels que pour tout $x \in \mathsf{X}$ et $n \in \mathbb{N}$,

$$\mathbf{W}_{\mathbf{d}_{\tilde{V}}}(\delta_x K^n, \pi) \leq C \rho^n \int \tilde{V}(x, y) d\pi(y). \quad (1.12)$$

Voyons maintenant comment les convergences (1.9) et (1.12) peuvent être utilisées pour contrôler $\mathbf{B}_n(f)$ dans la décomposition de l'erreur (1.6). Par définition (1.8), si K est V -uniformément géométriquement ergodique, nous obtenons que si $\|f\|_V < +\infty$,

$$|\mathbf{B}_n(f)| \leq \frac{C \|f\|_V \mu_0(V)}{n(1 - \rho)}.$$

Si K est uniformément géométriquement ergodique pour la distance de Wasserstein associée à $\mathbf{d}_{\tilde{V}}$, semi-continue inférieurement, nous pouvons retrouver une telle convergence. En effet, comme nous l'avons remarqué précédemment, il existe dans ce cas pour tout $k \in \mathbb{N}$, un plan de transfert entre π et $\mu_0 K^k$ ce qui permet d'en déduire que

$$|\mathbb{E}_{\mu_0}[f(Y_k)] - \pi(f)| \leq [f]_{\mathbf{d}_{\tilde{V}}} \mathbf{W}_{\mathbf{d}_{\tilde{V}}}(\mu_0 K^k, \pi).$$

Ainsi, si (1.12) est vraie, nous obtenons que pour toute fonction f vérifiant $[f]_{\mathbf{d}_{\tilde{V}}} < +\infty$,

$$|B_n(f)| \leq \frac{C[f]_{\mathbf{d}_{\tilde{V}}} \int \tilde{V}(x, y) d\pi(y) d\mu_0(x)}{n(1 - \rho)}.$$

Tournons nous maintenant à l'étude du terme de fluctuation $V_n(f)$ dans (1.6). Pour cela, une première possibilité est de prendre un point de vue asymptotique et établir un théorème central limite de la forme, sous \mathbb{P}_{μ_0} pour une distribution initiale μ_0 ,

$$n^{-1/2} \left\{ \sum_{k=0}^{n-1} (f(Y_k) - \pi(f)) \right\} \Rightarrow N(0, \sigma^2(f)), \text{ lorsque } n \rightarrow +\infty, \quad (1.13)$$

où \Rightarrow désigne la convergence en loi et $N(m, \sigma^2)$ est la loi gaussienne unidimensionnelle de moyenne m et de variance σ^2 . La quantité $\sigma^2(f)$ est appelée la variance asymptotique du noyau K pour la fonction f . Ce type de résultat permet de conclure grossièrement que le terme de fluctuation $V_n(f)$ se comporte comme $V_n(f) \approx \sigma^2(f)/n$. Ce point de vue asymptotique est un moyen pratique de construire des intervalles de confiance pour $\pi(f)$.

Pour montrer un théorème centrale limite du type (1.13), une approche qui s'est montrée intéressante est basée sur les solutions de Poisson associées à K . Soit f une fonction mesurable, alors on dit que \hat{f} est solution de l'équation de Poisson associée à f si,

$$\hat{f} - K\hat{f} = f - \pi(f). \quad (1.14)$$

Notons que il n'y a pas unicité de la solution de Poisson car pour tout $c \in \mathbb{R}$, $\hat{f} + c$ est également une solution de Poisson. [37, Theorem 21.2.5] montre que si $f \in L^2(\pi)$, où π est l'unique probabilité invariante pour K , qu'il existe une solution de Poisson \hat{f} associée à f , vérifiant $\hat{f} \in L^2(\pi)$, alors (1.13) est satisfait à stationnarité, c'est à dire pour $\mu_0 = \pi$ et on peut identifier la variance asymptotique:

$$\sigma^2(f) = \mathbb{E}_{\pi}[\{\hat{f}(X_1) - K\hat{f}(X_0)\}^2]. \quad (1.15)$$

De plus, [37, Proposition 21.1.3] assure alors que si pour toutes mesures de probabilité μ_0, ν_0

$$\lim_{n \rightarrow +\infty} \|\mu_0 K^n - \nu_0 K^n\|_{TV} = 0, \quad (1.16)$$

alors si le théorème centrale limite (1.13) est vrai pour une fonction f à stationnarité alors il est vrai pour toute distribution initiale μ_0 avec la même variance asymptotique. Á partir de ces deux résultats, on peut montrer que (1.13) est vraie pour toute fonction

f vérifiant $\|f\|_{V^{1/2}} < +\infty$ si K est V -uniformément ergodique. En effet, si K vérifie (1.9), alors (1.16) est automatiquement vraie et il est facile de vérifier que π est l'unique mesure de probabilité stationnaire pour K car on peut utiliser le théorème suivant.

Proposition 1.1. *Supposons que K est un noyau de Markov ergodique dans le sens où pour tout $x \in \mathsf{X}$, $\lim_{n \rightarrow +\infty} \|\delta_x K^n - \pi\|_{\text{TV}} = 0$. Si $\tilde{\pi}$ est une autre distribution stationnaire pour K . Alors nécessairement $\tilde{\pi} = \pi$.*

Preuve: Par définition et en utilisant l'invariance de $\tilde{\pi}$, il est aisé de montrer que pour tout $n \in \mathbb{N}$, $\|\tilde{\pi} - \pi\|_{\text{TV}} = \|\tilde{\pi} K^n - \pi\|_{\text{TV}} \leq \int \|\delta_x K^n - \pi\|_{\text{TV}} d\tilde{\pi}(x)$. Or, pour tout $x \in \mathsf{X}$, $\lim_{n \rightarrow +\infty} \|\delta_x K^n - \pi\|_{\text{TV}} = 0$ par hypothèse et la distance en variation totale est bornée par 2. Ainsi, une application directe du théorème de convergence dominée permet de conclure. \square

Ensuite, il reste à montrer que pour f mesurable vérifiant $\|f\|_{V^{1/2}} < +\infty$, il existe une solution de Poisson \hat{g}_f (1.14). Pour cela, on définit la fonction \hat{g}_f pour tout $x \in \mathsf{X}$ par

$$\hat{g}_f(x) = \sum_{i=0}^{\infty} \{K^i f(x) - \pi(f)\}.$$

Cette fonction est bien définie car $\pi(f) \leq \|f\|_{V^{1/2}} \pi(V) < +\infty$ en utilisant que V est plus grande que 1 et [37, Lemma 14.1.10]. De plus,

$$\sum_{i=0}^{\infty} |K^i f(x) - \pi(f)| \leq \|f\|_{V^{1/2}} C^{1/2} V^{1/2}(x) / (1 - \rho^{1/2}),$$

d'après (1.9) et l'inégalité de Cauchy-Schwarz. Cela conclut la preuve que \hat{g}_f est bien définie et de plus vérifie que pour tout $x \in \mathsf{X}$,

$$\hat{g}_f \leq \|f\|_{V^{1/2}} C^{1/2} V^{1/2}(x) / (1 - \rho^{1/2}). \quad (1.17)$$

Ainsi, $\hat{g}_f \in L^2(\pi)$ car $\pi(V) < +\infty$, et on peut conclure que (1.13) est vraie pour toute distribution initiale et on peut également avoir une estimée de la variance asymptotique à partir de (1.15) et (1.17).

Pour améliorer la convergence des méthodes MCMC, il a été proposé d'en combiner plusieurs. À partir de noyaux $\{K_i\}_{i=1}^I$ invariants pour la densité cible, le plus souvent, les auteurs considèrent des noyaux de la forme

$$K_1 \cdots K_I \text{ ou } \sum_{i=1}^I \varpi_i K_i,$$

où $\{\varpi_i\}_{i=1}^I \in (\mathbb{R}_+)^I$ est une suite de combinaison convexe de poids : $\sum_{i=1}^I \varpi_i = 1$. Cependant, ce type de procédure n'est que partiellement satisfaisante car elles ne permettent pas d'exploiter certaines informations sur les noyaux $\{K_i\}_{i=1}^I$, et en particulier leur comportement local sur certaines régions de l'espace. Dans une contribution [36] développée en Section 1.4.3, nous proposons une nouvelles méthodes permettant une meilleure combinaison de noyaux MCMC suivant leurs propriétés de convergence locales.

1.3 Les diffusions de Langevin sous/sur-amorties et leurs discrétisations

Nous avons pu observer que, bien que l'algorithme de Metropolis-Hastings donne un moyen générique de définir des méthodes MCMC, le choix du noyau de proposition peut être problématique. Idéalement, nous souhaiterions pouvoir utiliser des noyaux qui soient proches en un certain sens d'être réversibles par rapport à π pour qu'un nombre non négligeable de points soient acceptés au cours de l'exécution de l'algorithme. Pour obtenir de tels noyaux, il a été proposé de considérer des dynamiques continues pour lesquelles la mesure cible π est invariante. Les dynamiques de Langevin sont aujourd'hui l'une des dynamiques par défaut pour échantillonner des configurations de systèmes moléculaires en physique statistique computationnelle [2, 58, 90]. Elles ont gagné également en popularité en statistiques computationnelles et en apprentissage [10, 118] pour obtenir des échantillons approximatifs de la distribution a posteriori d'un modèle statistique. Ici, nous considérons deux telles dynamiques : les dynamiques de Langevin sous et sur-amorties,

Nous supposons dans toute cette section que l'espace d'état est \mathbb{R}^d et que la mesure cible π admet une densité par rapport à la mesure de Lebesgue de la forme

$$x \mapsto \exp(-U(x)) / \int e^{-U(y)} dy.$$

1.3.1 La diffusion de Langevin sur-amorties: discrétisation et processus collant associé

Ici, la fonction $U : \mathbb{R}^d \rightarrow \mathbb{R}$ est supposée continûment différentiable. La première dynamique de Langevin que nous considérons est la diffusion de Langevin sur-amorties telle que proposée par [52, 107, 121]. Cette dynamique est régie par l'équation différentielle stochastique (EDS) suivante :

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t, \tag{1.18}$$

où ∇U est le gradient de U et $(B_t)_{t \geq 0}$ est un mouvement brownien d -dimensionnel. Sous de faibles hypothèses sur U , par exemple ∇U Lipschitz, (1.18) admet des solutions fortes uniques. De plus, ces solutions permettent de définir le semi-groupe de Markov associé à (1.18) par $P_t(x, \mathbf{A}) = \mathbb{P}(X_t^x \in \mathbf{A})$, où $(X_t^x)_{t \geq 0}$ est une solution de (1.18) partant de $x \in \mathbb{R}^d$ et $\mathbf{A} \in \mathcal{B}(\mathbb{R}^d)$. On peut montrer que le semi-groupe $(P_t)_{t \geq 0}$ est réversible par rapport à π , *i.e.*, pour tout $t \geq 0$, P_t est réversible comme noyau de Markov (1.4). Pour cela, le chemin le plus simple est de passer par son générateur défini pour toute fonction $f \in C^2(\mathbb{R}^d)$ par

$$\mathcal{A}f = -\langle \nabla U, \nabla f \rangle + \Delta f, \tag{1.19}$$

où Δf est le Laplacien de f . On peut montrer que \mathcal{A} vu comme un opérateur sur $L^2(\pi)$ de domaine $D(\mathcal{A})$ est auto-adjoint : pour toutes fonctions $f, g \in C^2(\mathbb{R}^d)$ à support compact,

$$\int f \mathcal{A}g d\pi = \int \mathcal{A}f g d\pi.$$

En conséquence, cela entraîne que P_t est lui-même auto-adjoint dans $L^2(\pi)$ d'après [94, Proposition 2.8.5] et [54, Proposition 2.9]. De plus, sous des conditions faibles de régularité, $(P_t)_{t \geq 0}$ est ergodique :

$$\lim_{t \rightarrow +\infty} \|\mu P_t - \pi\|_{\text{TV}} = 0,$$

pour toute distribution initiale μ . Par ailleurs des résultats de convergence en différentes métriques, plus ou moins fins, ont pu être établis dans la littérature [32, 44]. Le problème principal qui se pose alors est la simulation de solutions de (1.18). Ce qui la plupart du temps est très coûteux ou impossible. Pour répondre à ce problème, il est maintenant très courant d'approximer des solutions de (1.18) à partir de discrétisation. Ici, nous nous intéressons à la discrétisation d'Euler-Maruyama qui définit la chaîne de Markov $(Y_k)_{k \in \mathbb{N}}$ par la récurrence :

$$Y_{k+1} = Y_k - \gamma \nabla U(Y_k) + \sqrt{2\gamma} G_{k+1}, \quad (1.20)$$

où $\gamma > 0$ est le pas de discrétisation et $(G_k)_{k \in \mathbb{N}}$ est une suite de variables aléatoires i.i.d. de loi normale standard. Notons R_γ le noyau de Markov associé à cette chaîne. La méthode MCMC utilisant la chaîne de Markov $(Y_k)_{k \in \mathbb{N}}$ comme des échantillons de lois approximant π est appelée l'algorithme de Langevin non-ajusté (ULA) ou Monte Carlo Langevin (LMC). Ici comme nous allons le voir, nous indiquons le noyau par le pas γ car c'est le paramètre essentiel qui contrôle la qualité d'approximation de la discrétisation. Dans le cas où U est Lipschitz et satisfait des conditions de courbures à l'infini du type

$$\lim_{\|x\| \rightarrow +\infty} \langle \nabla U(x), x \rangle / \|x\|^\beta \geq \eta,$$

pour $\eta > 0$ et $\beta \geq 0$ alors il est facile de vérifier que R_γ satisfait une condition de dérive et qu'il est fortement apériodique et irréductible par rapport à la mesure de Lebesgue. Ainsi, R_γ admet alors une unique mesure stationnaire π_γ . Malheureusement, cette loi est le plus souvent différente de π . Les premiers résultats initiés par [127] ont permis de montrer que si f est suffisamment régulières alors un développement de l'erreur était possible, à savoir :

$$\int f d\pi - \int f d\pi_\gamma = C_f \gamma + \mathcal{O}(\gamma^2). \quad (1.21)$$

Pour justifier rapidement ce résultat, nous introduisons le générateur à la chaîne de Markov à temps discret de noyau R_γ qui est défini par $R_\gamma f - f$. Un calcul simple montre que les deux générateurs sont très proches. Plus précisément, soit Y_1 le premier itéré de (1.20) avec $Y_0 = x$. Un développement de Taylor de f autour de x implique que

$$\mathbb{E}[f(X_1)] = f(x) + \gamma \mathcal{L}f(x) + \mathcal{O}(\gamma^2),$$

\mathcal{A} (1.19) est le générateur de la dynamique continue. Soit alors \hat{f} une solution de l'équation de Poisson associée à f et $\mathcal{L} : \mathcal{L}\hat{f} = -(f - \pi(f))$. Nous obtenons de manière

informelle en notant $(\mathcal{F}_k)_{k \in \mathbb{N}}$ la filtration associée à $(Y_k)_{k \in \mathbb{N}}$,

$$\begin{aligned} \sum_{k=0}^{n-1} \{f(Y_k) - \pi(f)\} &= \gamma^{-1} \sum_{k=0}^{n-1} \left\{ \hat{f}(Y_k) - \mathbb{E}^{\mathcal{F}_k} [\hat{f}(Y_{k+1})] \right\} + O(n\gamma) \\ &= \gamma^{-1} \left\{ \hat{f}(Y_0) - \mathbb{E}^{\mathcal{F}_{n-1}} [\hat{f}(Y_n)] \right\} + \gamma^{-1} \sum_{k=1}^{n-1} \left\{ \hat{f}(Y_k) - \mathbb{E}^{\mathcal{F}_{k-1}} [\hat{f}(Y_k)] \right\} + O(n\gamma). \end{aligned}$$

Le premier terme provient des conditions initiales et le second est une somme d'incrémentes de martingale par rapport à la filtration $(\mathcal{F}_k)_{k \in \mathbb{N}}$. En prenant l'espérance, nous obtenons alors

$$\left| \mathbb{E} \left[\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \right] - \pi(f) \right| \lesssim \frac{1}{n\gamma} + \gamma. \quad (1.22)$$

Ainsi, une conclusion partielle de (1.21) s'en déduit en prenant $n \rightarrow +\infty$.

Aussi, il existe maintenant de nombreux résultats permettant de borner la distance entre les itérés de R_γ et π en différentes métriques sous diverses conditions sur le potentiel U [32, 43, 44]. Plus précisément, ces résultats s'énoncent sous la forme de bornes de complexité au sens où, pour tout $\epsilon > 0$, il existe $\gamma_\epsilon > 0$ et $n_\epsilon \in \mathbb{N}^*$ tels que

$$\mathbf{d}(\mu_0 R_{\gamma_\epsilon}^{n_\epsilon}, \pi) \leq \epsilon, \quad (1.23)$$

la distribution initiale μ_0 étant le plus souvent "bien choisie". Ces résultats permettent de contrôler le terme de biais dans la décomposition de l'erreur quadratique (1.6). Il existe également de nouveaux résultats des bornes quantitatives de convergence de R_γ vers sa loi stationnaire π_γ [33], qui permettent comme nous l'avons vu de contrôler le terme de fluctuation.

Dans cette thèse, nous nous intéressons à un autre type d'erreur qui peut se produire lors de l'implémentation de ULA. Nous nous intéressons dans [40] au cas où le gradient du potentiel ∇U n'est possible que à une erreur numérique près. Cette situation a déjà été étudiée dans [51] pour le pendant continu sous l'hypothèse que il existe $\mathfrak{m} > 0$ et $R \geq 0$ tel que pour tout $x, y \in \mathbb{R}^d$, $\|x - y\| \geq R$,

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq \mathfrak{m} \|x - y\|^2. \quad (1.24)$$

Plus précisément, considérons

$$d\tilde{X}_t = -\nabla \tilde{U}(\tilde{X}_t) dt + \sqrt{2} dB_t, \quad (1.25)$$

avec $\nabla \tilde{U} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ et

$$\sup_{x \in \mathbb{R}^d} \|\nabla U(x) - \nabla \tilde{U}(x)\| \leq c_\infty < +\infty. \quad (1.26)$$

Il est à noter que $\nabla \tilde{U}$ n'est pas nécessairement issu du gradient d'un potentiel. Sur des conditions sur $\nabla \tilde{U}$ similaires à ∇U , (1.25) admet une unique solution forte et il existe un semi-groupe de Markov associé à (1.25), noté $(\tilde{P}_t)_{t \geq 0}$ qui admet une unique

mesure invariante $\tilde{\pi}$. Nous aimerions alors pouvoir mesurer l'écart entre μP_t et $\mu \tilde{P}_t$ pour une distribution initiale μ donné et de façon uniforme en temps, c'est-à-dire trouver $\varphi_{c_\infty} : \mathbb{R}_+ \rightarrow +\infty$ telle que

$$\mathbf{d}(\mu P_t, \mu \tilde{P}_t) \leq \varphi_{c_\infty}(t), \quad (1.27)$$

et idéalement $\varphi_{c_\infty}(t) \equiv 0$ si $c_\infty = 0$. Cela est fait dans [51] pour le cas de la distance en variation totale. L'objet essentiel qu'ils introduisent est un couplage $(X_t, \tilde{X}_t)_{t \geq 0}$ entre les deux EDS (1.18) et (1.25) et qui satisfait \mathbb{P} presque sûrement,

$$\|X_t - \tilde{X}_t\| \leq r_t, \quad (1.28)$$

où $(r_t)_{t \geq 0}$ est un processus dit collant car solution de l'EDS

$$dr_t = \{r_t \kappa(r_t) + c_\infty\} dt + 2\sqrt{2} \mathbb{1}_{\mathbb{R}_+^*}(r_t) dB_t^1, \quad (1.29)$$

où $(B_t^1)_{t \geq 0}$ est un mouvement brownien standard de dimension 1 et pour $r > 0$,

$$r^2 \kappa(r) = \sup_{\|x-y\|=r} \{ \langle \nabla U(x) - \nabla U(y), x-y \rangle \}.$$

L'existence du couplage $(X_t, \tilde{X}_t)_{t \geq 0}$ et du processus $(r_t)_{t \geq 0}$ se fait essentiellement en considérant deux fonctions continûment différentiables $\phi_\delta, \tilde{\phi}_\delta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ vérifiant

$$\begin{aligned} \phi_\delta(0) &= 0, \phi_\delta(t) > 0 \text{ si } t \in (0, \delta], \phi_\delta(t) = 1 \text{ si } t \geq \delta, \\ \phi_\delta^2(t) + (\tilde{\phi}_\delta(t))^2 &= 1 \text{ pour tout } t \geq 0, \end{aligned}$$

et la famille de processus $(X_t^\delta, \tilde{X}_t^\delta, r_t^\delta)_{t \geq 0}$ indexée par $\delta > 0$ tel que

$$\begin{aligned} dX_t^\delta &= -\nabla U(X_t^\delta) dt + \sqrt{2} \{ \phi_\delta(r_t) dB_t + \tilde{\phi}_\delta(r_t) d\tilde{B}_t \} \\ d\tilde{X}_t^\delta &= -\nabla \tilde{U}(\tilde{X}_t^\delta) dt + \sqrt{2} \{ \phi_\delta(r_t) (\text{Id} - 2e_t e_t^T) dB_t + \tilde{\phi}_\delta(r_t) d\tilde{B}_t \}, \\ r_t^\delta &= \|X_t^\delta - \tilde{X}_t^\delta\| \end{aligned} \quad (1.30)$$

où $e_t = [X_t^\delta - \tilde{X}_t^\delta] / \|X_t^\delta - \tilde{X}_t^\delta\|$ si $r_t \neq 0$ et $e_t = 0$ sinon, et $(B_t)_{t \geq 0}$ et $(\tilde{B}_t)_{t \geq 0}$ sont deux mouvements brownien standard indépendants. Alors, sous les conditions que ∇U est Lipschitz et (1.26), on peut montrer l'existence et l'unicité faible de $(X_t^\delta, \tilde{X}_t^\delta, r_t^\delta)_{t \geq 0}$ vérifiant (1.30) et si $\nabla \tilde{U}$ est lui-même Lipschitz, l'existence est même forte. Il est à noter que pour tout $\delta > 0$, $(X_t^\delta, \tilde{X}_t^\delta)_{t \geq 0}$ est un couplage de (1.18) et (1.25) d'après la caractérisation de Levy du mouvement brownien vérifiant $\|X_t^\delta - \tilde{X}_t^\delta\| \leq r_t^\delta$. Aussi, on peut vérifier que la famille de processus $(X_t^\delta, \tilde{X}_t^\delta, r_t^\delta)_{t \geq 0}$ indexée par $\delta > 0$ est tendue dans l'ensemble des mesures de probabilités sur $\mathbb{R}^{2d} \times \mathbb{R}_+$ et que chaque point limite $(X_t, \tilde{X}_t, r_t)_{t \geq 0}$ satisfait les propriétés escomptées par un passage à la limite en quelque sorte naturel, c'est-à-dire $(X_t, \tilde{X}_t)_{t \geq 0}$ est un couplage de (1.18) et (1.25) vérifiant (1.28) et $(r_t)_{t \geq 0}$ est solution faible de (1.29).

Il est à noter que la preuve de la tension dans [51] est un peu rapide. Les auteurs invoquent [65, 66], mais il s'avère que les résultats de ces deux papiers ne permettent

pas de conclure. Dans un papier récent, [47] les auteurs proposent une preuve plus élémentaire qui s'inspire en fait de l'une des contributions détaillées en Section 1.4.1. En effet, dans cette dernière mes co-auteurs et moi-même étudions un processus discret qui se révèle être une discrétisation de $(X_t, \tilde{X}_t, r_t)_{t \geq 0}$ et dérivons des propriétés de convergence pour ce dernier.

Si nous revenons au processus $(X_t, \tilde{X}_t, r_t)_{t \geq 0}$, on peut observer que d'après les propriétés que nous avons énoncés sur ce dernier, pour tout $t \geq 0$,

$$\|\mu P_t - \mu \tilde{P}_t\|_{\text{TV}} \leq \mathbb{P}(X_t \neq \tilde{X}_t) \leq \mathbb{P}(r_t > 0).$$

Ainsi, on a bien répondu à l'objectif annoncé (1.27) avec $\varphi_{c_\infty}(t) = \mathbb{P}(r_t > 0)$. Enfin, en étudiant les propriétés ergodiques de $(r_t)_{t \geq 0}$, [51] établissent une borne sur $\lim_{t \rightarrow +\infty} \mathbb{P}(r_t > 0)$ et ainsi une borne sur $\|\pi - \tilde{\pi}\|_{\text{TV}}$ où $\tilde{\pi}$ est la mesure stationnaire pour $(\tilde{P}_t)_{t \geq 0}$. De plus, les auteurs vérifient que $\lim_{t \rightarrow +\infty} \mathbb{P}(r_t > 0) \propto c_\infty$ lorsque $c_\infty \rightarrow 0$.

En réalité, un résultat du même type peut s'obtenir en distance de Wasserstein d'ordre 2 dans le cas où (1.24) est vraie pour tout $x, y \in \mathbb{R}^d$. En effet, considérons le couplage synchrone

$$\begin{aligned} dX_t &= -\nabla U(X_t)dt + \sqrt{2}dB_t \\ d\tilde{X}_t &= -\nabla \tilde{U}(\tilde{X}_t)dt + \sqrt{2}dB_t. \end{aligned}$$

On en déduit alors facilement que

$$\begin{aligned} d\|X_t - \tilde{X}_t\|^2 &= -2 \langle X_t - \tilde{X}_t, \nabla U(X_t) - \nabla \tilde{U}(\tilde{X}_t) \rangle dt \\ &\leq -2\mathfrak{m}\|X_t - \tilde{X}_t\|^2 dt + 2c_\infty\|X_t - \tilde{X}_t\| dt \\ &\leq -\mathfrak{m}\|X_t - \tilde{X}_t\|^2 dt + (c_\infty^2/\mathfrak{m})dt. \end{aligned}$$

On en déduit par une application directe du lemme de Gronwall que

$$\|X_t - \tilde{X}_t\|^2 \leq e^{-\mathfrak{m}t}\|X_0 - \tilde{X}_0\|^2 + (1 - e^{-\mathfrak{m}t})(c_\infty/\mathfrak{m})^2.$$

On en déduit alors que pour tout $t \geq 0$,

$$\mathbf{W}_2(\mu P_t, \mu \tilde{P}_t) \leq (1 - e^{-\mathfrak{m}t})^{1/2} c_\infty/\mathfrak{m}.$$

1.3.1.1 La diffusion de Langevin sous-amortie et ses discrétisations

Dans cette section, nous nous intéressons à la dynamique de Langevin sous-amortie, ou parfois appelée cinétique, qui décrit l'évolution de la position $(X_t)_{t \geq 0}$ et de la vitesse $(V_t)_{t \geq 0}$ d'un système par l'EDS:

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= [-\nabla U(X_t) - \kappa V_t] dt + \sigma dB_t. \end{aligned} \tag{1.31}$$

Ici, $\kappa, \sigma > 0$ sont des coefficients de friction et de diffusion respectivement, et $(B_t)_{t \geq 0}$ est un mouvement brownien standard. On montre que l'unique mesure de probabilité

invariante de (1.31) est la mesure de probabilité de Boltzmann–Gibbs, dont la densité par rapport à la mesure de Lebesgue est proportionnelle à $\bar{\pi} : (x, v) \mapsto \exp(-\kappa[2U(x) + \|v\|^2]/\sigma^2)$. Lorsque le paramètre de friction tend vers $+\infty$, alors on peut montrer [108, Section 6.5] que une version de $(X_t)_{t \geq 0}$ remise à l'échelle, tend vers une solution de Langevin sur-armorti

$$X_t = -\nabla U(X_t)dt + \sigma dB_t.$$

Nous sommes également intéressés par les situations où le terme de dérive ne provient pas d'un gradient, comme dans les simulations de dynamique moléculaire hors équilibre [29, 55], [86, Section 5] ce qui nous amène à considérer l'EDS

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= [b(X_t) - \kappa V_t] dt + \sigma dB_t, \end{aligned} \quad (1.32)$$

pour un champ $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$.

Malheureusement, tout comme la diffusion (1.18), la simulation exacte de solutions de (1.32) n'est en général pas possible et alors des discrétisations sont utilisées à la place pour approcher une trajectoire et son comportement asymptotique. La discrétisation la plus aisée est encore une fois la discrétisation d'Euler-Maruyama correspondant à (1.32) :

$$\begin{aligned} X_{k+1} &= X_k + \gamma V_k, \\ V_{k+1} &= V_k + \gamma [b(X_k) - \kappa V_k] + \sigma \gamma^{1/2} G_{k+1}. \end{aligned} \quad (1.33)$$

Ici γ est encore le pas de discrétisation et $(G_k)_{k \in \mathbb{N}^*}$ une suite i.i.d. de variables aléatoires gaussiennes standards. Une autre discrétisation qui est aussi explicite consiste à geler en quelque sorte la position X_k à chaque itération et à intégrer l'EDS de type Ornstein–Uhlenbeck correspondant :

$$\begin{aligned} d\tilde{X}_t &= \tilde{V}_t dt, \\ d\tilde{V}_t &= [-\nabla U(X_k) - \kappa \tilde{V}_t] dt + \sigma dB_t. \end{aligned}$$

On peut résoudre cette équation en utilisant l'intégration par parties ([80, p.116]),

$$d(e^{\kappa t} \tilde{V}_t) = -\nabla U(\tilde{X}_{k\gamma}) e^{\kappa t} dt + \sigma e^{\kappa t} dB_t,$$

pour tout $t \in (k\gamma, (k+1)\gamma]$,

$$\begin{aligned} \tilde{X}_t &= \tilde{X}_0 - \frac{1}{\kappa} (e^{-\kappa t} - 1) \tilde{V}_0 - \frac{1}{\kappa} \left(t + \frac{e^{-\kappa t} - 1}{\kappa} \right) \nabla U(X_k) \\ &\quad - \frac{\sigma}{\kappa} \int_0^t (e^{\kappa(s-t)} - 1) dB_s, \\ \tilde{V}_t &= e^{-\kappa t} \tilde{V}_0 + \frac{1}{\kappa} (e^{-\kappa t} - 1) \nabla U(X_k) + \sigma \int_0^t e^{\kappa(s-t)} dB_s. \end{aligned} \quad (1.34)$$

On pose alors $(X_{k+1}, V_{k+1}) = (\tilde{X}_\gamma, \tilde{V}_\gamma)$ avec $(\tilde{X}_0, \tilde{V}_0) = (X_k, V_k)$. Malheureusement comme le cas des discrétisations du Langevin sur-armortie, les discrétisations de (1.31)

même si elles sont ergodiques et possèdent une unique mesure stationnaire $\bar{\pi}_\gamma$ elle est différente de $\bar{\pi}$. Cependant, il existe des résultats de convergence faible du type (1.22) pour de nombreux schémas numériques. En ce qui concerne des estimées de la forme (1.23), la discrétisation définie à partir de (1.34) a déjà fait l'objet de plusieurs publications, par exemple [26, 31]. En particulier, en notant $\mu_{n,\gamma}$ la loi de X_n , ces deux études montrent que $\mathbf{W}_2(\mu_{n,\gamma}, \bar{\pi}) \leq \epsilon$ pour $n \gtrsim \sqrt{d}/\epsilon$, lorsque U est fortement convexe. Des schémas du second ordre sont aussi proposés dans [30, Section 4] et [31, Section 4].

Pour ce qui est de la convergence des schémas vers leur mesure stationnaire, nous pouvons déjà nous intéresser aux résultats existants pour la dynamique continue. Il existe diverses méthodes, par exemple celles basées sur l'hypocoercivité [126, 130] ou par couplage et existence d'une fonction de Lyapunov [49, 95, 117, 137]. De plus, bien que la dynamique soit dégénérée, *i.e.* la matrice de covariance associée à (1.32) ne soit pas inversible, on peut montrer [117], que, pour tout $t_0 > 0$ et condition initiale $(X_0, V_0) = (x_0, v_0) \in \mathbb{R}^{2d}$, la variable aléatoire (X_{t_0}, V_{t_0}) a une distribution avec une densité positive $(x_{t_0}, v_{t_0}) \mapsto p_{t_0}((x_0, v_0), (x_{t_0}, v_{t_0}))$ par rapport à la mesure de Lebesgue sur \mathbb{R}^{2d} , ce qui implique que le semigroupe de Markov associé à (1.32) est irréductible. Enfin, des bornes quantitatives sur p_{t_0} pour tout $t_0 > 0$ peuvent être établies par exemple en utilisant le calcul de Malliavin [8], des principes de comparaison [111, 112], ou encore avec la méthode dite parametrix [76, 85].

L'obtention de taux de convergence quantitatifs au niveau discret pour les schémas de discrétisation associés à (1.32) est plus difficile. Cela fait l'objet de l'une des contributions qui est résumée en Section 1.4.2.

1.4 Résumé synthétique des contributions

1.4.1 Processus collant discret

Nous nous intéressons dans cet article aux chaînes de Markov $(Y_k)_{k \in \mathbb{N}}$ partant de $y \in \mathbb{R}^d$ et définies par récurrence de la forme :

$$Y_{k+1} = T_\gamma(Y_k) + \sigma\sqrt{\gamma}Z_{k+1}, \quad (1.35)$$

où $\sigma > 0$, $\gamma \in (0, \bar{\gamma}]$, pour un certain $\bar{\gamma} > 0$, $\{T_\gamma : \gamma \in (0, \bar{\gamma}]\}$ est une famille de fonctions continues de \mathbb{R}^d à \mathbb{R}^d et $(Z_k)_{k \geq 1}$ est une séquence de i.i.d. de variables aléatoires gaussiennes sur \mathbb{R}^d . Notons que (1.35) comprends entre autre la discrétisation d'Euler-Maruyama des diffusions de Langevin sur-amorties. Dans ce cas, γ correspond au pas de discrétisation et pour tout $\gamma \in (0, \bar{\gamma}]$, $T_\gamma(y) = y - \gamma \nabla U(y)$ pour un certain potentiel $U : \mathbb{R}^d \rightarrow \mathbb{R}$. Cependant, dans certaines applications, le calcul explicite de ∇U n'est pas possible et d'autres méthodes numériques doivent être mises en œuvre pour l'estimer. Aussi, l'utilisation de ces algorithmes entraînent le plus souvent un biais supplémentaire. Dans [40], nous étudions précisément cette source d'erreur supplémentaire. En particulier, en se basant sur une chaîne définie par (1.35), nous considérons une seconde chaîne de Markov $(\tilde{Y}_k)_{k \in \mathbb{N}}$ définie par récurrence par

$$\tilde{Y}_{k+1} = \tilde{T}_\gamma(\tilde{Y}_k) + \sigma\sqrt{\gamma}\tilde{Z}_{k+1}, \quad (1.36)$$

où $\{\tilde{T}_\gamma : \gamma \in (0, \bar{\gamma}]\}$ est une famille de fonctions mesurables de \mathbb{R}^d dans \mathbb{R}^d telle que pour tout γ , \tilde{T}_γ est une approximation de T_γ dans le sens où

$$\sup_{y \in \mathbb{R}^d} \left\| \tilde{T}_\gamma(y) - T_\gamma(y) \right\| \leq c_\infty, \quad (1.37)$$

pour $c_\infty > 0$. Dans (1.36), $(\tilde{Z}_k)_{k \geq 1}$ est une séquence de i.i.d. d -dimensionnelle de variables aléatoires gaussiennes standard potentiellement corrélées avec $(Z_k)_{k \geq 1}$ apparaissant dans (1.35).

Notons maintenant R_γ et \tilde{R}_γ le noyau de Markov associé à $(Y_k)_{k \in \mathbb{N}}$ et $(\tilde{Y}_k)_{k \in \mathbb{N}}$ respectivement. Supposons maintenant que la famille $\{T_\gamma : \gamma \in (0, \bar{\gamma}]\}$ satisfasse la condition uniforme en le pas de temps suivante : il existe $\tilde{m} > 0$ tel que pour tout $\gamma \in (0, \bar{\gamma}]$, $x, y \in \mathbb{R}^d$,

$$\|T_\gamma(x) - T_\gamma(y)\| \leq (1 - \tilde{m}\gamma) \|x - y\|. \quad (1.38)$$

En particulier, lorsque $T_\gamma = x - \gamma \nabla U(x)$, (1.38) est satisfait si U vérifie (1.24) et $\bar{\gamma}$ est pris suffisamment petit. Sous (1.38), il est aisé de montrer que R_γ et \tilde{R}_γ sont uniformément géométriquement ergodiques. Notons alors leur loi invariante par π_γ et $\tilde{\pi}_\gamma$ respectivement. Sous (1.37), le processus global $(\tilde{Y}_k)_{k \in \mathbb{N}}$ peut être vu comme une version perturbée de $(Y_k)_{k \in \mathbb{N}}$, et on s'attend alors à ce que $\tilde{\pi}_\gamma$ soit proche de π_γ . Le but principal de [40] est d'établir des bornes quantitatives sur des distances de Wasserstein entre ces deux lois. L'étude de la perturbation des processus de Markov a fait l'objet de nombreux travaux, comme par exemple [71, 97, 102, 123, 125]. Cependant, il s'avère que ces résultats existants ne s'appliquent pas tels quels. Plus précisément, nous portons dans notre étude, une attention particulière à la dépendance des bornes que nous obtenons vis-à-vis du pas de discrétisation γ . C'est-à-dire que nous obtenons des bornes qui sont uniformes en le pas de temps γ de la forme: pour tout $\gamma \in (0, \bar{\gamma}]$,

$$\mathbf{W}_c(\pi_\gamma, \tilde{\pi}_\gamma) \leq \varphi(c_\infty),$$

pour une fonction φ vérifiant $\varphi(c_\infty) \sim Cc_\infty$, pour une constante $C \geq 0$ dépendante uniquement des paramètres du problème. En particulier, nous retrouvons dans le cas de la variation totale les résultats de [51] pour le processus collant continu décrit en Section 1.3.1 et étendons ces résultats à d'autres distances de Wasserstein.

Comme dans [51], nos résultats sont basés sur la construction d'un couplage approprié des processus, c'est-à-dire une construction simultanée d'une paire $(Y_k, \tilde{Y}_k)_{k \in \mathbb{N}}$ de chaînes non-indépendantes qui suivent marginalement (4.1) et (4.2) respectivement et sont conçues pour se rapprocher et rester proches les unes des autres. Les estimations sur les lois des chaînes découlent alors de l'étude de $(\|Y_k - \tilde{Y}_k\|)_{k \in \mathbb{N}}$, qui est elle-même basée sur l'analyse d'une chaîne de Markov $(W_k)_{k \in \mathbb{N}}$ sur $[0, +\infty)$ qui est construite de telle sorte à borner cette différence : $\|Y_k - \tilde{Y}_k\|_k \leq W_k$ pour tout $k \in \mathbb{N}$. Ainsi, l'étude de la distance entre les lois marginales des deux chaînes de Markov se réduit à l'étude de la chaîne de Markov $(W_k)_{k \in \mathbb{N}}$ qui possède de nombreuses propriétés intéressantes. Par exemple, elle est stochastiquement monotone, c'est-à-dire si $(W'_k)_{k \in \mathbb{N}}$ est une chaîne de Markov associée au même noyau de Markov que $(W_k)_{k \in \mathbb{N}}$ et telle que $W_0 \leq W'_0$,

alors pour tout $k \in \mathbb{N}$, W'_k est stochastiquement dominé par W_k , *i.e.* pour tout $t \geq 0$, $\mathbb{P}(W_k \leq t) \geq \mathbb{P}(W'_k \leq t)$. L'analyse de cette chaîne de Markov auxiliaire constitue une des contributions de [40]. Les principaux résultats de cette analyse sont les suivants. Premièrement, nous prouvons que $(W_k)_{k \in \mathbb{N}}$ admet une mesure de probabilité invariante unique et que, indépendamment de γ , les moments et la masse sur $(0, +\infty)$ de cette loi d'équilibre sont d'ordre c_∞ lorsque $c_\infty \rightarrow 0$. Ensuite, nous établissons la convergence géométrique de la chaîne vers son équilibre, à un taux explicite (uniforme en le pas de temps γ).

Dans le cas où (1.35) correspond à la discrétisation d'Euler-Maruyama (1.20), nous prouvons également que lorsque $\gamma \rightarrow 0$, la chaîne $(W_k)_{k \in \mathbb{N}}$, remise à l'échelle de manière appropriée, converge en loi vers la diffusion collante en temps continu (1.29). Ce résultat apporte un éclairage nouveau sur le processus collant continu, qui, dans [51], est construit comme la limite de diffusions en temps continu avec des coefficients de diffusion qui tendent vers zéro. Ici, nous montrons que dans un certain sens, $(W_k)_{k \in \mathbb{N}}$ peut alors être vu comme un schéma de discrétisation pour le processus collant.

Outre l'évidente différence temps continu/temps discret entre [51] et [40], soulignons d'autres distinctions importantes. Tout d'abord, dans [51], le processus collant unidimensionnel possède une mesure invariante explicite. Ce n'est pas le cas dans notre cadre, ce qui rend la dérivation des bornes sur les moments de l'équilibre bien plus compliquée. Deuxièmement, dans [51], bien qu'il soit prouvé que la masse à zéro et le premier moment de la loi de la diffusion collante convergent vers leur valeur à l'équilibre (ce qui est suffisant pour obtenir des estimations sur les lois des deux processus initiaux d -dimensionnels), la question de la convergence à long terme n'est pas abordée pour la diffusion collante, alors que nos résultats de convergence à long terme pour $(W_k)_{k \geq 0}$ ainsi que sa convergence en $\gamma \rightarrow +\infty$ fournissent un taux de convergence explicite pour la diffusion collante. Enfin, les résultats de [51] ne concernent que la variation totale, alors que nous considérons une classe plus large de distances.

1.4.2 Convergence des discrétisations du processus de Langevin sous-amorti

Dans [41], nous analysons le comportement de chaînes de Markov $(X_k, V_k)_{k \in \mathbb{N}}$, de la forme suivante, pour tout $k \in \mathbb{N}$, $\gamma \in (0, \bar{\gamma}]$

$$\begin{aligned} X_{k+1} &= X_k + \gamma V_k + \gamma f_\gamma \left(X_k, \gamma^\delta V_k, \gamma^{\delta+1/2} \sigma_\gamma Z_{k+1}, W_{k+1} \right) + \gamma^{\delta+1/2} \sigma_\gamma \mathbf{D}_\gamma Z_{k+1}, \\ V_{k+1} &= \tau_\gamma V_k + \gamma g_\gamma \left(X_k, \gamma^\delta V_k, \gamma^{\delta+1/2} \sigma_\gamma Z_{k+1}, W_{k+1} \right) + \sqrt{\gamma} \sigma_\gamma Z_{k+1}, \end{aligned} \quad (1.39)$$

où $\bar{\gamma}, \delta > 0$, $(Z_{k+1})_{k \in \mathbb{N}}$ est une suite de vecteurs gaussiens d -dimensionnels centrés réduits indépendants identiquement distribués, $(W_{k+1})_{k \in \mathbb{N}}$ est une suite de vecteurs gaussiens¹ m -dimensionnels centrés réduits indépendants identiquement distribués, indépendant de $(Z_{k+1})_{k \in \mathbb{N}}$. Á ma connaissance, tous les schémas explicites de discrétisation de (1.32)

¹en fait la loi des $(W_{k+1})_{k \in \mathbb{N}}$ peut être bien plus générale mais nous nous restreignons à ce cas pour simplifier la présentation

peuvent se mettre sous cette forme, en particulier celui d'Euler-Maruyama (1.33) et celui défini par (1.34). De plus, cette famille comprend également les schémas par division proposés dans [18, 84], qui deviennent de plus en plus populaires en dynamique moléculaire (voir [2, Section 12.2]) Plus d'exemples sont fournis dans le Chapter 3. Notons \bar{R}_γ le noyau de Markov associé à la chaîne de Markov $(X_k, V_k)_{k \in \mathbb{N}}$.

En ce qui concerne la convergence de \bar{R}_γ , il est primordial d'établir des taux de convergence ρ_γ pour les schémas de discrétisation avec un pas de temps $\gamma > 0$ qui sont similaires à leur homologue continu, c'est-à-dire telle que $\log(\rho_\gamma)$ est linéaire en le pas de temps γ , c'est-à-dire $\log(\rho_\gamma) \propto \gamma$ lorsque $\gamma \rightarrow 0$. Plus précisément, si nous désignons par \bar{R}_γ le noyau de Markov associé à un schéma de discrétisation donné, il est raisonnable de s'attendre à ce que, sous des conditions appropriées et pour tout $\gamma > 0$ suffisamment petit, ce noyau admette une distribution invariante $\bar{\pi}_\gamma$ et

$$\mathbf{d}(\mu_0 \bar{R}_\gamma^k, \bar{\pi}_\gamma) \leq C \bar{\rho}^{\gamma k}, \quad (1.40)$$

pour une distribution initiale μ_0 , où $C \geq 0$ et $\bar{\rho} \in [0, 1)$, tandis que \mathbf{d} est une certaine distance sur l'ensemble des mesures de probabilité sur \mathbb{R}^{2d} . Actuellement, l'une des principales options pour obtenir une telle convergence est de prouver des estimations de Lyapunov et des conditions de minorisation qui sont uniformes en le pas de temps, c'est-à-dire que de telles conditions tiennent pour $\bar{R}_\gamma^{\lceil t_0/\gamma \rceil}$ pour un certain $t_0 > 0$ pour des constantes qui ne dépendent pas de γ . Cette stratégie a été utilisée pour la dynamique de Langevin sur-amortie dans [16] et [33]. Alors que les conditions de Lyapunov sont basées sur des calculs algébriques directs, et peuvent nécessiter de considérer des schémas implicites [77, 95], montrer une condition de minorisation uniforme dans le pas de temps est le principal point faible de cette approche. Pour les dynamiques stochastiques non dégénérées, ce type de résultats peut être établi en s'appuyant sur le calcul de Malliavin [9], mais la preuve qui en résulte est assez complexe. Il peut également être possible d'écrire une preuve directe comme dans [16] à partir du théorème de Girsanov, mais ceci n'est toutefois possible que pour des dynamiques avec un bruit additif. Enfin, l'approche par couplage développée dans [33, 44] ne peut pas être appliquée aux discrétisations de l'EDS dégénérée (1.32). L'une des contributions qui est résumée en Section 1.4.2 est d'établir (1.40) pour la plupart des schémas existants. Dans [41], nous considérons une autre approche pour obtenir des conditions de minorisation uniformes dans le pas de temps : l'idée est de considérer le schéma numérique sur de petits temps physiques comme une perturbation d'un processus gaussien donné.

Soit P un noyau de Markov. On dit qu'un ensemble \mathbf{C} est de Doeblin (pour le noyau P) s'il existe $m \in \mathbb{N}^*$ et $\varepsilon > 0$ tel que pour tout $z, z' \in \mathbf{C} \times \mathbf{C}$,

$$\|P^m(z, \cdot) - P^m(z', \cdot)\|_{\text{TV}} \leq 2(1 - \varepsilon).$$

Le premier résultat de [41] est de montrer que sous des conditions faibles sur le schéma (1.39) tous les compacts de \mathbb{R}^{2d} sont des ensembles de Doeblin R_γ mais avec m et ε uniforme en γ *i.e.* il existe $\bar{t}_0, \bar{\gamma} > 0$ tel que pour tout $M > 0$ et $t_0 \in (0, \bar{t}_0]$ il existe $\varepsilon_{t_0, M}$ tel que pour tout $\gamma \in (0, \bar{\gamma}]$ et $z, z' \in \mathbb{R}^{2d}$, $\|z\| \leq M$, $\|z'\| \leq M$,

$$\|R_\gamma^{\lceil t_0/\gamma \rceil + 1}(z, \cdot) - R_\gamma^{\lceil t_0/\gamma \rceil + 1}(z', \cdot)\|_{\text{TV}} \leq 2(1 - \varepsilon_{t_0, M}). \quad (1.41)$$

1.4. Résumé synthétique des contributions

Le deuxième résultat principal est d'établir des conditions de dérivées pour \bar{R}_γ également uniforme en le pas de temps, c'est-à-dire de la forme : il existe $V : \mathbb{R}^{2d} \rightarrow [1, +\infty)$, $\lambda \in (0, 1)$ et $M, b \geq 0$, tels que pour tout $\gamma \in (0, \bar{\gamma}]$ et $x, v \in \mathbb{R}^d$,

$$\bar{R}_\gamma V(x, v) \leq \lambda^\gamma V(x, v) + \gamma b \mathbb{1}_{[0, M]}(\|x\| + \|v\|). \quad (1.42)$$

La combinaison des deux résultats (1.41) et (1.42) permettent l'application direct de [37, Théorème 19.4.1], ce qui implique qu'il existe $\bar{\gamma}_U > 0$ tel que, pour tout $\gamma \in (0, \bar{\gamma}_U]$, \bar{R}_γ admet une unique mesure de probabilité invariante $\bar{\pi}_\gamma$ et il existe $A_U \geq 0$ et $\rho_U \in (0, 1)$ tels que pour tout $\gamma \in (0, \bar{\gamma}_U]$,

$$\text{pour tout } x, v \in \mathbb{R}^d, \quad \|\delta_{(x,v)} R_\gamma^k - \bar{\pi}_\gamma\|_V \leq A_U \rho_U^{k\gamma} V(x, v).$$

Nous avons alors bien répondu à notre objectif initial à savoir (1.40).

1.4.3 L'algorithme de téléportation Kick-Kac

Dans ce travail, nous nous proposons de viser une distribution cible π en proposant une nouvelle classe d'algorithmes: les algorithmes de téléportation Kick-Kac (les processus KKT). Avant d'entrer dans la description de cette classe, nous rappelons la formule de Kac sur laquelle reposent ces algorithmes. Soit P un noyau de Markov sur (X, \mathcal{E}) admettant une mesure de probabilité invariante π . La formule de Kac s'écrit ainsi: pour tout ensemble C π -accessible, i.e. tel que $\mathbb{P}_x(\sigma_C < \infty) > 0$ pour π -presque tout $x \in X$, on a pour toute fonction mesurable, positive ou bornée f ,

$$\pi(f) = \int_C \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} f(X_k) \right] = \int_C \pi(dx) \mathbb{E}_x \left[\sum_{k=1}^{\sigma_C} f(X_k) \right].$$

S'il s'avère que π est l'unique probabilité invariante pour P , alors tout ensemble C telle que $\pi(C) > 0$ est aussi π -accessible, ce qui permet d'identifier rapidement des ensembles C pour lequel la formule de Kac s'applique. Il est remarquable de constater dans cette formule que l'expression de $\pi(f)$ s'appuie sur la connaissance de π dans l'ensemble C uniquement.

Ceci nous a amené à proposer de viser π de la manière suivante. Supposons donné un noyau de Markov P π -invariant. On construit d'abord une chaîne auxiliaire de noyau Q définie sur C visant π_C (la restriction de π à l'ensemble C). A partir de chaque état de cette chaîne, nous proposons un candidat suivant le noyau P et acceptons le candidat uniquement s'il n'est pas dans C . Formellement, l'algorithme s'écrit sous la forme suivante:

- (i) **Initialization:** Simuler (Y_0, Z_0) suivant une loi arbitraire
- (ii) **Itération:** pour $k \leftarrow 1$ à n
 - (a) simuler $Y_k^* \sim P(Y_{k-1}, \cdot)$
 - (b) si $Y_k^* \notin C$, poser $(Y_k, Z_k) \leftarrow (Y_k^*, Z_{k-1})$

(c) Sinon simuler $Z_k \sim Q(Z_{k-1}, \cdot)$ et poser $Y_k \leftarrow Z_k$

La chaîne étendue $(Y_k, Z_k)_{k \in \mathbb{N}}$ s'avère être une chaîne de Markov dont la loi invariante possède π comme marginale sur la première composante. Cet algorithme permet ainsi de combiner astucieusement deux dynamiques différentes, celle associée à P permettant d'explorer la distribution cible à l'extérieur de C et celle associée à Q à l'intérieur de C . Cet algorithme peut particulièrement nous être utile lorsque nous constatons qu'une chaîne de noyau P n'explore bien que certaines zones de l'espace d'état. Si on appelle C les zones où P est moins efficace, alors nous pouvons utiliser un autre noyau, appelé Q pour visiter cet ensemble, et combiner ensuite les deux suivant l'algorithme de téléportation.

Nous avons alors établi que la chaîne étendue satisfait une loi des grands nombres partant de n'importe quel point si et seulement si (a) le noyau Q vérifie la même propriété et (b) l'ensemble C est un ensemble atteignable pour P à partir de n'importe quel point de l'espace. Ce résultat (qui n'est pas améliorable car les conditions sont nécessaires et suffisantes) est basé sur une nouvelle caractérisation (du moins, à notre connaissance) d'une loi des grands nombres partant de n'importe quel point en terme d'harmoniques bornées.

Nous avons également obtenu des conditions permettant d'assurer que la chaîne résultante est géométriquement ergodique. Comme pour la loi des grands nombres, nous avons établi ces résultats à partir essentiellement de conditions sur Q (ensemble small pour Q et drift géométrique en dehors de cet ensemble) tandis que les conditions pour P sont nettement moins contraignantes (on ne suppose pas d'ensemble small pour P). Ceci confirme bien que l'algorithme de téléportation peut être une alternative intéressante lorsque les propriétés ergodiques de P sont ralenties dans certains endroits de l'espace d'état.

Enfin, nous proposons une extension de l'algorithme de téléportation qui nous permet d'étendre notablement les lois invariantes pour le noyau auxiliaire Q . Le principe de l'extension est très simple: il suffit d'ajouter une composante indépendante de loi uniforme et considérer le processus KKT associé à des ensembles de la forme $\{(x, u) \in X \times [0, 1] ; u \leq \alpha(x)\}$ où α est à valeurs dans $[0, 1]$. Au lieu que Q vise la restriction de π à un ensemble C , nous pouvons désormais lui faire viser n'importe quelle distribution auxiliaire $\tilde{\pi}$ pourvu que $\frac{d\tilde{\pi}}{d\pi}(x) \leq M$ pour π -presque tout $x \in X$. Cette extension permet de placer tout algorithme de Metropolis Hasting comme une instance particulière de l'algorithme de téléportation ainsi que l'algorithme de renouvellement de Brockwell et Kadane ([20]) qui en est aussi un cas particulier. De nombreuses questions restent encore et seront sans doute l'objet d'autres études: comment combiner astucieusement plusieurs ensembles C_i imbriqués les uns dans les autres ? Peut-on viser des distributions auxiliaires sans utiliser des chaînes de Markov ? Quelles seraient des heuristiques pour déterminer automatiquement les ensembles C permettant l'implémentation des algorithmes de téléportation ?

CHAPTER 2

Boost your favorite Markov Chain Monte Carlo sampler using Kac's theorem: the Kick-Kac teleportation algorithm

RANDAL DOUC¹, ALAIN DURMUS², AURÉLIEN ENFROY³ AND JIMMY OLSSON⁴

The present paper focuses on the problem of sampling from a given target distribution π defined on some general state space. To this end, we introduce a novel class of non-reversible Markov chains, each chain being defined on an extended state space and having an invariant probability measure admitting π as a marginal distribution. The proposed methodology is inspired by a new formulation of Kac's theorem and allows global and local dynamics to be smoothly combined. Under mild conditions, the corresponding Markov transition kernel can be shown to be irreducible and Harris recurrent. In addition, we establish that geometric ergodicity holds under appropriate conditions on the global and local dynamics. Finally, we illustrate numerically the use of the proposed method and its potential benefits in comparison to existing Markov chain Monte Carlo (MCMC) algorithms.

¹Department CITI, Telecom SudParis, Evry, France. Email: randal.douc@telecom-sudparis.eu

²Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette, France.

Email: alain.durmus@ens-paris-saclay.fr

³Laboratoire de mathématiques d'Orsay, Université Paris-Saclay, Orsay, France. Email: aurelien.enfroy@ens-paris-saclay.fr

⁴Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden. Email: jimmyol@kth.se

2.1 Introduction

In this work, we develop a new Monte Carlo technique to sample from a given target distribution π on some general state space $(\mathcal{X}, \mathcal{E})$. Our construction is based on *Kac's theorem*, which is a fundamental result in Markov chain theory. More precisely, let P be some Markov kernel leaving π invariant and $\mathsf{C} \in \mathcal{E}$ some measurable set. Moreover, for every $x \in \mathcal{E}$, let \mathbb{P}_x be the law of the canonical Markov chain $(X_n)_{n \in \mathbb{N}}$ on $(\mathcal{X}, \mathcal{E})$ with kernel P and such that $X_0 = x$, and denote by \mathbb{E}_x the corresponding expectation. Under some conditions on P and C , Kac's theorem states that for every bounded Borel-measurable function h on $(\mathcal{X}, \mathcal{E})$, the Lebesgue integral of h with respect to π , denoted by $\pi(h)$, can be expressed as

$$\pi(h) = \pi(\mathsf{C}) \int \pi_{\mathsf{C}}(\mathrm{d}x) \mathbb{E}_x \left[\sum_{k=0}^{\sigma_{\mathsf{C}}-1} h(X_k) \right], \quad (2.1)$$

where $\sigma_{\mathsf{C}} = \inf\{k > 0 : X_k \in \mathsf{C}\}$ is the first return time to C and $\pi_{\mathsf{C}} := \pi(\cdot \cap \mathsf{C})/\pi(\mathsf{C})$ is the (normalized) restriction of π to C (see Theorem 2.1 below). Here we have assumed that $\pi(\mathsf{C}) > 0$. A remarkable feature of Kac's formula is that it characterizes completely π by $\pi(\mathsf{C})$, π_{C} and P . In particular, there is a large flexibility in the choice of these three elements. In the light of (2.1), a possible strategy for estimating $\pi(h)$ is to produce i.i.d. draws $(X_0^i)_{i=1}^N$ from π_{C} and let each draw X_0^i initialize a Markov chain $(X_k^i)_{k \in \mathbb{N}}$ evolving according to P ; once the chain returns to C , *i.e.*, after σ_{C}^i time steps, it is killed and the last state is discarded. By (2.1), the independent Markovian trajectories $(X_k^i)_{k=0}^{\sigma_{\mathsf{C}}^i-1}$, $i \in \{1, \dots, N\}$, produced in this manner can be used to form a consistent estimator $\sum_{i=1}^N \sum_{k=0}^{\sigma_{\mathsf{C}}^i-1} h(X_k^i) / \sum_{i=1}^N \sigma_{\mathsf{C}}^i$ of $\pi(h)$. More generally, instead of sampling i.i.d. draws from π_{C} , we may generate $(X_0^i)_{i \in \mathbb{N}}$ by simulating a Markov chain evolving on C according to some Markov kernel Q leaving π_{C} invariant. Then, as before, we let each state X_0^i initialize a Markovian excursion $(X_k^i)_{k=0}^{\sigma_{\mathsf{C}}^i-1}$ evolving according to P . After this, the—now dependent—trajectories $(X_k^i)_{k=0}^{\sigma_{\mathsf{C}}^i-1}$, $i \in \mathbb{N}$, can be used to form an estimator of $\pi(h)$ as previously. Now, concatenate the produced trajectories into a single sequence and denote by Y_n the $(n+1)$ th element of this sequence (or, formally, for every $n \in \mathbb{N}$, let $i(n) := \sup\{i \in \mathbb{N} : \sum_{\ell=1}^i \sigma_{\mathsf{C}}^{\ell} \leq n\}$, $k(n) := n - \sum_{\ell=1}^{i(n)} \sigma_{\mathsf{C}}^{\ell}$ and set $Y_n := X_{k(n)}^{i(n)}$). The resulting process $(Y_n)_{n \in \mathbb{N}}$, which we will refer to as the *Kick-Kac teleportation* (KKT) process and define carefully in Section 2.3, is non-Markovian and evolves by moving—or, *teleporting*—across C according to Q and excursing the space outside C according to P . After each Q transition, an excursion according to P is initiated; once the excursion returns to C , its last state is discarded, and a new teleportative move according to Q , picking up the end state of the previous teleportation, takes place, generating the new state of the process. When the density of π is known up to a normalizing constant, so is the density of π_{C} , which provides a variety of possible choices of Q , including the plethora of *Markov chain Monte Carlo* (MCMC) kernels, such as those of *Metropolis-Hastings* (MH) type [61, 98, 128]. Moreover, with an appropriate choice of C , *rejection sampling*

can be implemented to generate i.i.d. draws from $\pi_{\mathcal{C}}$; in that case, $Q(x, \cdot) = \pi_{\mathcal{C}}$ for all $x \in \mathcal{C}$, yielding a process lacking long-term memory.

The KKT sampling method that we propose has particular relevance to target distributions with multiple, isolated modes. Standard MCMC methods, such as the *random-walk Metropolis (RWM) algorithm*, the *Metropolis-adjusted Langevin algorithm (MALA)* [121] or the *Hamiltonian Monte Carlo (HMC) method* [104], generally exhibit poor performance when applied to such models due to difficulties in crossing low-probability barriers. Still, since such models arise in many applications [56, 69, 78], plenty of variations on the MCMC technology have been proposed to address this issue. The literature on this topic is extensive, why we here only provide a brief introduction to place our work into a context. A first family of methods is based on tempering strategies [60, 101, 105] and a second one on optimization methods trying first to identify the modes of the target distribution before applying adaptive MCMC techniques [4, 79]. Finally, exploratory schemes have been suggested in [14, 133]. In contrast to these works, our methodology belongs to a class of ‘resurrected’ or ‘regenerative’ processes, which return to a certain part of the state space once being killed [63, 64, 103]. In many cases, these processes have been constructed with the aim of coupling from the past and perfect simulation [57, 110]. Among existing methods, we found that the hybrid kernel approach introduced in [20] is the one that is most closely related to the KKT sampler proposed by us, and a detailed comparison is given in Section 2.4.2 below. Using regeneration strategies for designing efficient Monte Carlo samplers is still the subject of active research; see [70, 81, 132] for some recent work in this direction. In the case where π is multi-modal, the KKT process provides, especially in its most general form—the *general KKT process*—, a very generic framework. Using the teleportation-based sampling procedure described above, multi-modality can be straightforwardly handled by letting \mathcal{C} be a low-probability region surrounding the modes of the target and P some MCMC kernel used to explore locally the modes. For instance, \mathcal{C} can be defined as a level set of the target density (up to a constant of proportionality). Parameterized appropriately, the KKT process can be made to transit with ease across regions of the state space where the ergodic properties of the kernel P are poor, de-correlating significantly the produced samples.

In the present paper, which provides a basis for further methodological innovation, we

- provide, under mild assumptions, a novel and concise proof of Kac’s theorem (2.1) as well as a generalized version of the same identity where $\pi_{\mathcal{C}}$ is replaced by a general probability distribution $\tilde{\pi}$ on (X, \mathcal{E}) such that $\tilde{\pi}$ is absolutely continuous with respect to π with a π -a.s. bounded density $d\tilde{\pi}/d\pi$. These results are of independent interest.
- use Kac’s theorem to construct an increasingly general KKT process framework and provide a rigorous analysis of its theoretical properties.
- illustrate, in a simulation study, the potential of the proposed KKT sampler.

Outline of the paper and the results

As a preparation for coming developments, we first provide, in Section 2.2, our new proof of Kac's theorem (2.1) under the mild assumption that C is π -accessible (Theorem 2.1). Moreover, as established by Corollary 1, in the case where π is the unique invariant distribution of P , Kac's theorem holds true for any set $C \in \mathcal{E}$ such that $\pi(C) > 0$.

In the first part of Section 2.3 we examine the KKT process in the memoryless case where $Q(x, \cdot) = \pi_C$ for all $x \in C$. In this case, the KKT process $(Y_n)_{n \in \mathbb{N}}$ is a Markov chain, and we establish that it allows the target π of interest as an invariant distribution. In addition, we provide necessary and sufficient conditions for the process to be π -reversible (Proposition 2.4), conditions that are not satisfied in general.

In the second part of Section 2.3 we turn to a general Q leaving π_C invariant, and define the KKT process $(Y_n)_{n \in \mathbb{N}}$ via an additional, auxiliary C -valued process $(Z_n)_{n \in \mathbb{N}}$ defining the states of $(Y_n)_{n \in \mathbb{N}}$ inside C . In this construction (which is detailed in Algorithm 2), the bivariate process $(Y_n, Z_n)_{n \in \mathbb{N}}$ is a Markov chain on the extended state space $X \times C$ with invariant probability measure $\tilde{\pi}$ admitting π as a marginal with respect to the first argument (Proposition 2.5). Moreover, if π and π_C are the unique invariant probability measures for P and Q , respectively, then $\tilde{\pi}$ is the unique invariant probability measure for the process $(Y_n, Z_n)_{n \in \mathbb{N}}$ (Proposition 2.6). Then, under the assumption that $\tilde{\pi}$ is the unique invariant probability measure, we show that $(Y_n, Z_n)_{n \in \mathbb{N}}$ satisfies a law of large numbers with respect to $\tilde{\pi}$ if and only if the durations of the excursions from C are a.s. finite and Q satisfies a law of large numbers with respect to π_C (Theorem 2.7).

In order to complete the analysis of the KKT process, we also show that $(Y_n, Z_n)_{n \in \mathbb{N}}$ is geometrically ergodic (with respect to $\tilde{\pi}$) if Q is geometrically ergodic (with respect to π_C) and P exhibits geometric drift toward C (Theorem 2.9). Importantly, P is not required to be geometrically ergodic, which supports the idea underpinning the construction, namely that the mixing of a Markov chain evolving according to P can be drastically improved by changing the dynamics of the chain in regions of bad mixing.

Finally, to facilitate reading, the notation used in this paper is collected in Section 2.6 and some technical results and proofs are postponed to Section 2.7.

2.2 Kac's formula

As explained above, the teleportation process introduced in Section 2.3 is based on Kac's theorem. Due to its importance to the construction of the teleportation process, we restate, after the introduction of adequate notation, this result and provide general assumptions under which the theorem holds true. In addition, we provide an effective and, up to our knowledge, novel proof.

Let (X, \mathcal{E}) be some measurable space and denote by $M(\mathcal{E})$ and $M_1(\mathcal{E})$ the sets of σ -finite nonnegative measures and probability measures, respectively, on (E, \mathcal{E}) . In addition, we denote by $F_b(\mathcal{E})$ and $F_+(\mathcal{E})$ the sets of bounded and nonnegative $\mathcal{E}/\mathcal{B}(\mathbb{R})$ -measurable functions, respectively. Furthermore, let $(X^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ be the associated canonical space and denote by $(X_n)_{n \in \mathbb{N}}$ the corresponding coordinate process. For every

2.2. Kac's formula

initial distribution $\mu \in \mathcal{M}_1(\mathcal{E})$ and Markov kernel P on $\mathsf{X} \times \mathcal{E}$ we let \mathbb{P}_μ be the probability measure on $(\mathsf{X}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ induced by P and μ and denote by \mathbb{E}_μ the corresponding expectation. As usual, the *shift operator* is the transformation on $\mathsf{X}^{\mathbb{N}}$ given by $\theta : (\omega_0, \omega_1, \omega_2, \dots) \mapsto (\omega_1, \omega_2, \dots)$. If π is an invariant probability measure for P , i.e. $\pi P = \pi$, then by the Markov property, for every $A \in \mathcal{E}^{\otimes \mathbb{N}}$,

$$\mathbb{E}_\pi[\mathbb{1}_A \circ \theta] = \mathbb{E}_\pi[\mathbb{1}_A], \quad (2.2)$$

which implies that $\mathbb{P}_{\pi \circ \theta^{-1}} = \mathbb{P}_\pi$; in other words, $(\mathsf{X}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$ is a measure-preserving dynamical system. Moreover, for every $C \in \mathcal{E}$ we define recursively

$$\sigma_C^\ell := \inf \left\{ k > \sigma_C^{\ell-1} : X_k \in C \right\}, \quad \ell \in \mathbb{N}^*,$$

with $\sigma_C^0 := 0$. Note that $\sigma_C := \sigma_C^1$ is the return time to the set C . In the following we say that $C \in \mathcal{E}$ is π -*accessible* if $\mathbb{P}_x(\sigma_C < +\infty) > 0$ for π -almost all $x \in \mathsf{X}$. Moreover, we simply say that C is *accessible* if $\mathbb{P}_x(\sigma_C < +\infty) > 0$ for all $x \in \mathsf{X}$.

We have now all the notation required for the statement of Kac's theorem.

Theorem 2.1 (Kac's theorem). *Let P be a Markov kernel on $\mathsf{X} \times \mathcal{E}$ with invariant probability measure π . Then for every π -accessible set $C \in \mathcal{E}$ it holds that*

$$\pi = \pi_C^0 = \pi_C^1, \quad (2.3)$$

where the measures π_C^0 and π_C^1 are defined by, for $f \in F_+(\mathcal{E})$,

$$\pi_C^0(f) := \int_C \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{\sigma_C-1} f(X_k) \right], \quad \pi_C^1(f) := \int_C \pi(dx) \mathbb{E}_x \left[\sum_{k=1}^{\sigma_C} f(X_k) \right]. \quad (2.4)$$

Even though the conclusion of Theorem 2.1 is the same as that of [100, Theorem 10.4.9], these results are obtained under different assumptions. Indeed, [100, Theorem 10.4.9] requires the Markov kernel P to be recurrent and irreducible, properties that are not assumed in Theorem 2.1. It can also be noted that [37, Theorem 3.6.5] provides another statement of Kac's theorem which is expressed in terms of subinvariant measures (instead of invariant probability measures). Consequently, in [37], the identity (2.3) holds under the assumptions that (i) C is π -accessible and (ii) $\mathbb{P}_x(\sigma_C < \infty) = 1$ for π -almost all $x \in C$, whereas, as we will see in Lemma 1 below, it turns out that that (ii) is already implied by (i) in the particular case of invariant probability measures. This allows the proof of (2.3) to be neatly simplified.

We provide here an intermediate result which will be used in the proof of Theorem 2.1

Lemma 1. *Let P be a Markov kernel on $\mathsf{X} \times \mathcal{E}$ with invariant probability measure π . Then $C \in \mathcal{E}$ is π -accessible if and only if $\mathbb{P}_x(\sigma_C < +\infty) = 1$ for π -almost all $x \in \mathsf{X}$.*

Proof. Assume that C is π -accessible and set $A := \{\sigma_C < +\infty\}$. Since $\{\sigma_C \circ \theta < +\infty\} \subset \{\sigma_C < +\infty\}$, it holds that $\mathbb{1}_A \circ \theta \leq \mathbb{1}_A$. Combining this with (2.2) yields $\mathbb{1}_A \circ \theta = \mathbb{1}_A$, \mathbb{P}_π -a.s., i.e., $\mathbb{1}_A$ is \mathbb{P}_π -a.s. an invariant random variable for the dynamical

system $(\mathbf{X}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}, \mathbb{P}_\pi, \theta)$. By [100, Theorem 17.1.1], $\mathbb{P}_{X_0}(\mathbf{A}) = \mathbb{1}_{\mathbf{A}}$, \mathbb{P}_π -a.s. Since \mathbf{C} is π -accessible by assumption, it holds that $\mathbb{P}_x(\mathbf{A}) > 0$ for π -almost all $x \in \mathbf{X}$; thus, we obtain

$$1 = \int \pi(dx) \mathbb{1}_{\{\mathbb{P}_x(\mathbf{A}) > 0\}} = \mathbb{P}_\pi(\mathbb{P}_{X_0}(\mathbf{A}) > 0) = \mathbb{P}_\pi(\mathbb{1}_{\mathbf{A}} > 0) = \mathbb{P}_\pi(\mathbf{A}) = \mathbb{P}_\pi(\sigma_{\mathbf{C}} < +\infty).$$

Consequently,

$$0 = 1 - \mathbb{P}_\pi(\sigma_{\mathbf{C}} < +\infty) = \int \pi(dx) (1 - \mathbb{P}_x(\sigma_{\mathbf{C}} < +\infty)),$$

from which we finally conclude that $\mathbb{P}_x(\sigma_{\mathbf{C}} < +\infty) = 1$ for π -almost all $x \in \mathbf{X}$. The converse is obvious. \square

Proof of Theorem 2.1. By the last-exit decomposition and the Markov property we have for all $f \in \mathbf{F}_+(\mathcal{E}) \cap \mathbf{F}_b(\mathcal{E})$ and $n \in \mathbb{N}^*$,

$$\begin{aligned} \pi(f) &= \sum_{\ell=1}^n \mathbb{E}_\pi \left[f(X_n) \mathbb{1}_{\mathbf{C}}(X_\ell) \prod_{k=\ell+1}^n \mathbb{1}_{\mathbf{C}^c}(X_k) \right] + \mathbb{E}_\pi[f(X_n) \mathbb{1}_{\{\sigma_{\mathbf{C}} > n\}}] \\ &= \sum_{\ell=1}^n \mathbb{E}_\pi \left[\mathbb{1}_{\mathbf{C}}(X_\ell) \mathbb{E}_{X_\ell} \left[f(X_{n-\ell}) \prod_{k=1}^{n-\ell} \mathbb{1}_{\mathbf{C}^c}(X_k) \right] \right] + \mathbb{E}_\pi[f(X_n) \mathbb{1}_{\{\sigma_{\mathbf{C}} > n\}}]. \end{aligned}$$

Noting that π is invariant and setting $k = n - \ell$, we get

$$\begin{aligned} \pi(f) &= \sum_{k=0}^{n-1} \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x[f(X_k) \mathbb{1}_{\{\sigma_{\mathbf{C}} > k\}}] + \mathbb{E}_\pi[f(X_n) \mathbb{1}_{\{\sigma_{\mathbf{C}} > n\}}] \\ &= \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{(n-1) \wedge (\sigma_{\mathbf{C}} - 1)} f(X_k) \right] + \mathbb{E}_\pi[f(X_n) \mathbb{1}_{\{\sigma_{\mathbf{C}} > n\}}]. \end{aligned} \quad (2.5)$$

We now let $n \rightarrow \infty$ in (2.5). Since f is bounded and $\mathbb{P}_\pi(\sigma_{\mathbf{C}} = +\infty) = 0$ by Lemma 1, the dominated convergence theorem implies that the second term on the right-hand side of (2.5) tends to zero. Moreover, by applying the monotone convergence theorem to the first term, we obtain that $\pi = \pi_{\mathbf{C}}^0$. Finally, this implies the last equality of (2.3), since by the Markov property, $\pi_{\mathbf{C}}^1 = \pi_{\mathbf{C}}^0 P = \pi P = \pi$. \square

Remark 2.2. If $\mathbf{C} \in \mathcal{E}$ is arbitrary (and not necessarily π -accessible), then by (2.5), $\pi(f) \geq \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{(n-1) \wedge (\sigma_{\mathbf{C}} - 1)} f(X_k) \right]$ for every $f \in \mathbf{F}_+(\mathcal{E})$. Thus, the monotone convergence theorem implies that

$$\pi(f) \geq \lim_{n \rightarrow \infty} \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{(n-1) \wedge (\sigma_{\mathbf{C}} - 1)} f(X_k) \right] = \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x \left[\sum_{k=0}^{\sigma_{\mathbf{C}} - 1} f(X_k) \right].$$

Remark 2.3. If $\mathbf{C} \in \mathcal{E}$ is π -accessible, then Theorem 2.1 implies that $\pi(\mathbf{C}) > 0$; indeed, otherwise $\pi_{\mathbf{C}}^0$, and hence π , would be the null measure, which is not possible.

2.3. The Kick-Kac teleportation process

To apply Theorem 2.1, we need to check that C is π -accessible. If the Markov kernel P has a unique invariant probability measure π , it turns out that this condition can be simplified.

Lemma 2. *Let P be a Markov kernel on $X \times \mathcal{E}$ with a unique invariant probability measure π . Then every set $C \in \mathcal{E}$ such that $\pi(C) > 0$ is π -accessible.*

Before proving Lemma 2, we combine it with Theorem 2.1 to obtain the following corollary, which allows Kac's theorem to be applied to every set $C \in \mathcal{E}$ such that $\pi(C) > 0$.

Corollary 1. *Let P be a Markov kernel on $X \times \mathcal{E}$ with a unique invariant probability measure π . Then for all $C \in \mathcal{E}$ such that $\pi(C) > 0$, $\pi = \pi_C^0 = \pi_C^1$, where the measures π_C^0 and π_C^1 are defined in (2.4).*

Proof of Lemma 2. Let $C \in \mathcal{E}$ be such that $\pi(C) > 0$. Since P has a unique invariant probability measure, [37, Theorem 5.2.6] implies that the associated dynamical system is ergodic. Therefore, by the Birkhoff ergodic theorem,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} \mathbb{1}_C(X_k) = \pi(C) > 0, \quad \mathbb{P}_\pi\text{-a.s.}$$

Thus, $\mathbb{P}_\pi(\sigma_C < +\infty) = 1$, implying that $\mathbb{P}_x(\sigma_C < +\infty) = 1 > 0$ for π -almost all $x \in X$. This shows that C is π -accessible. \square

2.3 The Kick-Kac teleportation process

Given $\pi \in M_1(\mathcal{E})$, we will now use Kac's theorem (Theorem 2.1) to construct a stochastic process targeting π . In the following we let P be some π -invariant Markov kernel on $X \times \mathcal{E}$; this kernel will be referred to as the *base kernel*.

2.3.1 The memoryless Kick-Kac teleportation process

As we saw in the previous section, the following assumption guarantees that Kac's theorem (Theorem 2.1) applies.

(A1) The Markov kernel P allows an invariant probability measure π such that C is π -accessible.

As shown by Lemma 2, (A1) is satisfied if π is the unique invariant probability measure of P and $\pi(C) > 0$.

Under (A1), the probability measure

$$\pi_C : \mathcal{E} \ni A \mapsto \pi(A \cap C) / \pi(C) \tag{2.6}$$

is well defined in the light of Remark 2.3 and provides the conditional probabilities of π given C . Note that (2.3)–(2.4) express the invariant probability measure π in terms of

its restriction to the set C only. This remarkable fact is underpinning the construction of the *memoryless Kick-Kac teleportation* (KKT) process, whose evolution is described by Algorithm 1. In words, the memoryless KKT process evolves by proposing a candidate according to the Markov kernel P . If the candidate falls into the region C , it is discarded and the next state is drawn exactly from π_C , independently of the past (in this sense, the process is *memoryless*); otherwise, if the candidate falls outside C , it is accepted as the new state of the process.

Algorithm 1 The memoryless Kick-Kac teleportation process

```

1: Initialization: draw  $Y_0$ 
2: for  $k \leftarrow 1$  to  $n$  do
3:   draw  $Y_k^* \sim P(Y_{k-1}, \cdot)$ 
4:   if  $Y_k^* \notin C$  then
5:     set  $Y_k \leftarrow Y_k^*$ 
6:   else
7:     draw  $Y_k \sim \pi_C$ 
8:   end if
9: end for

```

Of course, the construction requires the set C to be chosen in such a way that exact sampling from π_C is feasible. A natural approach is based on the accept-reject algorithm. More precisely, assume that the probability measure π is absolutely continuous with respect to some σ -finite measure ν and denote by p_π a version of its Radon–Nikodym derivative with respect to ν . In addition, let $q : X \rightarrow \mathbb{R}_+$ be another probability density function with respect to ν , serving as instrumental density, and assume that $C \subset \{x \in X : p_\pi(x) \leq \epsilon q(x)\}$. Then, as π_C admits a density proportional to $\mathbb{1}_C p_\pi$ with respect to ν and $\mathbb{1}_C(x) p_\pi(x) \leq \epsilon q(x)$ for every $x \in X$, the accept-reject method can be applied to obtain i.i.d. samples from π_C , provided that the ratio $\mathbb{1}_C(x) p_\pi(x) / [\epsilon q(x)]$ is computable for all $x \in X$ such that $q(x) > 0$.

The Markov transition kernel S on $X \times \mathcal{E}$ associated with the Markov chain $(Y_n)_{n \in \mathbb{N}}$ is given by

$$Sh(y) := \int_{C^c} P(y, dy') h(y') + P(y, C) \pi_C(h), \quad (y, h) \in X \times F_+(\mathcal{E}).$$

Since π is invariant for P , it holds that

$$\pi Sh = \pi P(\mathbb{1}_{C^c} h) + \pi P(C) \pi_C(h) = \pi(\mathbb{1}_{C^c} h) + \pi(\mathbb{1}_C h) = \pi(h),$$

which means that π is also invariant for S (some kernel notation, such as Sh and πS above, that will be used in the rest of the paper is provided in Section 2.6).

We now address the π -reversibility of S . In the following, we let $\pi \otimes P$ be the probability measure in $M_1(\mathcal{E}^{\otimes 2})$ defined by $\pi \otimes P(A) := \int \mathbb{1}_A(x, y) \pi(dx) P(x, dy)$, $A \in \mathcal{E}^{\otimes 2}$. Note that with this notation, P is reversible with respect to π if and only if for every $A \in \mathcal{E}$ and $B \in \mathcal{E}$, $\pi \otimes P(A \times B) = \pi \otimes P(B \times A)$. In addition, for any given $D \in \mathcal{E}$,

2.3. The Kick-Kac teleportation process

let $\mathcal{E}_D := \{D \cap A : A \in \mathcal{E}\}$ be the trace σ -field of \mathcal{E} on D and denote, for any measure $\nu \in \mathcal{M}_1(\mathcal{E})$, by $\nu|_D$ the restriction of ν to \mathcal{E}_D .

Proposition 2.4. *Assume (A1) and that \mathcal{E}_{C^c} is countably generated. Then S is π -reversible if and only if the following two conditions are satisfied.*

(a) For every $A \in \mathcal{E}_{C^c}^{\otimes 2}$,

$$\iint \mathbb{1}_A(x, y) \pi(dx) P(x, dy) = \iint \mathbb{1}_A(y, x) \pi(dx) P(x, dy).$$

(b) There exists $\mu \in \mathcal{M}(\mathcal{E}_{C^c})$ such that $P(y, \cdot)|_{C^c} = \mu$ for π -almost all $y \in C$.

The proof of Proposition 2.4 is given in Section 2.7.1.

In the light of Proposition 2.4, we may expect S to be non-reversible in general; indeed, even if P is π -reversible (implying (a)), condition (b) will typically not be satisfied. Some classes of MCMC methods based on non-reversible kernels have been shown to exhibit favorable convergence properties compared to standard reversible methods. For instance, [34] carries through a detailed study of a particular type of *lifted Markov chains*, and shows theoretically and numerically that this class of MCMC algorithms provides better accuracy than standard reversible ones. In addition, by extending results in [129], [5] establishes that non-reversible MCMC algorithms most often lead to improved asymptotic variance in the case where a central limit theorem holds. However, whereas the non-reversibility of the memoryless KKT algorithm is encouraging, the requirement of exact sampling from π_C may limit the choice of the set C . In the next section we will show how this drawback can be circumvented by including an additional Markov transition kernel into the construction.

2.3.2 The Kick-Kac teleportation process

To avoid exact sampling from the conditional probability measure π_C defined by (2.6), we now provide an alternative to the memoryless KKT process. The process that we will construct includes another level of transitions according to some Markov kernel Q on $C \times \mathcal{E}_C$ which will be referred to as the teleportation kernel and will be supposed to satisfy the following condition.

(A2) The Markov transition kernel Q on $C \times \mathcal{E}_C$ allows π_C as invariant probability measure.

The output $(Y_k)_{k \in \mathbb{N}}$ of Algorithm 2, in which the exact sampling from π_C in Algorithm 1 is replaced by Markovian moves according to Q , will be referred to as the *Kick-Kac teleportation* (KKT) *process* in the sequel. In Algorithm 2, a candidate is proposed according to the kernel P . If it falls outside C , the candidate is accepted as the new state of the process. On the other hand, if the candidate falls into the critical region C , it is discarded, and a new state of the process is generated through a transition according to Q , starting off from the last state in C in the past history of the process. In

Algorithm 2 The Kick-Kac teleportation (KKT) process

```

1: Initialization: draw  $(Y_0, Z_0)$ 
2: for  $k \leftarrow 1$  to  $n$  do
3:   draw  $Y_k^* \sim P(Y_{k-1}, \cdot)$ 
4:   if  $Y_k^* \notin \mathbf{C}$  then
5:     set  $(Y_k, Z_k) \leftarrow (Y_k^*, Z_{k-1})$ 
6:   else
7:     draw  $Z_k \sim Q(Z_{k-1}, \cdot)$  and set  $Y_k \leftarrow Z_k$ 
8:   end if
9: end for

```

the case where \mathbf{C} is a region where the ergodic behavior of P is poor, the KKT process simply replaces transitions according to P across \mathbf{C} by transitions according to the kernel Q , which is tailor-made to target efficiently $\pi_{\mathbf{C}}$.

While $(Y_k)_{k \in \mathbb{N}}$ is not a Markov chain in general, so is $(Y_k, Z_k)_{k \in \mathbb{N}}$. It is easily seen that the Markov transition kernel of the latter bivariate process is given by

$$Rh(y, z) := \int_{\mathbf{C}^c} P(y, dy') h(y', z) + P(y, \mathbf{C}) \int_{\mathbf{C}} Q(z, dz') h(z', z'), \quad (2.7)$$

for $(y, z) \in \mathbf{X} \times \mathbf{C}$ and $h \in \mathbf{F}_+(\mathcal{E} \otimes \mathcal{E}_{\mathbf{C}})$. We now provide an expression of an invariant probability measure of R which allows π as a marginal. This constitutes a first step towards justifying theoretically that the KKT process can be used for producing approximate draws from π .

From now on, we will deal simultaneously with several Markov kernels. Thus, we will specify the notation \mathbb{P}_{μ} to avoid ambiguities. More precisely, for a given measurable space $(\mathbf{F}, \mathcal{F})$, Markov kernel K on $\mathbf{F} \times \mathcal{F}$ and initial distribution $\mu \in \mathbf{M}_1(\mathcal{F})$ we denote by \mathbb{P}_{μ}^K the probability measure on $(\mathbf{F}^{\mathbb{N}}, \mathcal{F}^{\otimes \mathbb{N}})$ induced by K and μ . The associated expectation operator is denoted \mathbb{E}_{μ}^K . If $x \in \mathbf{F}$ and $\mu = \delta_x$, we write \mathbb{P}_x^K and \mathbb{E}_x^K instead of $\mathbb{P}_{\delta_x}^K$ and $\mathbb{E}_{\delta_x}^K$, respectively. Now, define the probability measure

$$\check{\pi}(h) := \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathbf{C}}-1} h(X_k, x) \right], \quad h \in \mathbf{F}_+(\mathcal{E} \otimes \mathcal{E}_{\mathbf{C}}). \quad (2.8)$$

The following key result justifies theoretically the use of the KKT process for targeting π .

Proposition 2.5. *Assume (A1). Then, π is the marginal of $\check{\pi}$ with respect to the first component. Moreover, if also (A2) holds, then $\check{\pi}$ is an invariant probability measure for R .*

Proof. Under (A1), Theorem 2.1 applies to π . Thus, for every $\mathbf{A} \in \mathcal{E}$,

$$\check{\pi}(\mathbf{A} \times \mathbf{C}) = \int_{\mathbf{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathbf{C}}-1} \mathbb{1}_{\mathbf{A}}(X_k) \right] = \pi(\mathbf{A}),$$

2.3. The Kick-Kac teleportation process

which establishes the first claim of the proposition.

We turn to the second claim. Under (A1), Lemma 1 shows that $\mathbb{P}_x^P(\sigma_{\mathcal{C}} < +\infty) = 1$ for π -almost all $x \in \mathbf{X}$. This fundamental property will be used in several parts of the proof. By the very definitions of $\tilde{\pi}$ and R , for every $h \in F_+(\mathcal{E} \otimes \mathcal{E}_{\mathcal{C}})$,

$$\tilde{\pi}Rh = \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathcal{C}}-1} Rh(X_k, x) \right] = A + B, \quad (2.9)$$

where

$$A := \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathcal{C}}-1} \int_{\mathcal{C}^c} P(X_k, dy) h(y, x) \right],$$

$$B := \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathcal{C}}-1} P(X_k, \mathcal{C}) \right] \int_{\mathcal{C}} Q(x, dz) h(z, z).$$

Using the Markov property (in combination with Tonelli's theorem), we may write

$$\begin{aligned} A &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\mathbb{E}_x^P [h(X_1, x) \mathbb{1}_{\mathcal{C}^c}(X_1)] + \sum_{k=1}^{\infty} \mathbb{1}_{\mathcal{C}^c}(X_k) \mathbb{E}_{X_k}^P [h(X_1, x) \mathbb{1}_{\mathcal{C}^c}(X_1)] \right] \\ &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[h(X_1, x) \mathbb{1}_{\mathcal{C}^c}(X_1) + \sum_{k=1}^{\infty} \mathbb{E}_x^P [h(X_{k+1}, x) \mathbb{1}_{\mathcal{C}^c}(X_k) \mathbb{1}_{\mathcal{C}^c}(X_{k+1}) \mid X_k] \right] \\ &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[h(X_1, x) \mathbb{1}_{\mathcal{C}^c}(X_1) + \sum_{k=1}^{\sigma_{\mathcal{C}}} h(X_k, x) \mathbb{1}_{\mathcal{C}^c}(X_k) \right] \\ &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=1}^{\sigma_{\mathcal{C}}} \mathbb{1}_{\mathcal{C}^c}(X_k) h(X_k, x) \right] \\ &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=1}^{\sigma_{\mathcal{C}}-1} h(X_k, x) \right], \end{aligned}$$

where the last equality follows as $\mathbb{1}_{\mathcal{C}^c}(X_k) = 1$ for all $k \in \{1, \dots, \sigma_{\mathcal{C}} - 1\}$ and $\mathbb{1}_{\mathcal{C}^c}(X_{\sigma_{\mathcal{C}}}) = 0$ on the set $\{\sigma_{\mathcal{C}} < +\infty\}$ (having \mathbb{P}_x^P -probability one for π -almost all $x \in \mathbf{X}$). Similarly, using again the Markov property,

$$\begin{aligned} B &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathcal{C}}-1} \mathbb{1}_{\mathcal{C}}(X_{k+1}) \right] \int_{\mathcal{C}} Q(x, dz) h(z, z) \\ &= \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=1}^{\sigma_{\mathcal{C}}} \mathbb{1}_{\mathcal{C}}(X_k) \right] \int_{\mathcal{C}} Q(x, dz) h(z, z) \\ &= \int_{\mathcal{C}} \pi(dx) \int_{\mathcal{C}} Q(x, dz) h(z, z), \end{aligned}$$

where the last equality follows since $\mathbb{1}_{\mathcal{C}}(X_k) = 0$ for all $k \in \{1, \dots, \sigma_{\mathcal{C}} - 1\}$ and $\mathbb{1}_{\mathcal{C}}(X_{\sigma_{\mathcal{C}}}) = 1$ on $\{\sigma_{\mathcal{C}} < +\infty\}$. Therefore, (A2) implies that $B = \int_{\mathcal{C}} \pi(dz) h(z, z)$. Plugging the obtained expressions of A and B into (2.9) yields

$$\tilde{\pi} R h = \int_{\mathcal{C}} \pi(dx) \mathbb{E}_x^P \left[\sum_{k=0}^{\sigma_{\mathcal{C}}-1} h(X_k, x) \right] = \tilde{\pi}(h),$$

showing that $\tilde{\pi}$ is invariant for R . \square

2.3.3 A law of large numbers for the KKT process

In order to establish that $(Y_k, Z_k)_{k \in \mathbb{N}}$, evolving according to the Markov kernel R , satisfies a law of large numbers, we will first find conditions under which R admits a unique invariant probability measure. In that case, [37, Theorem 5.2.6] guarantees that the associated dynamical system is ergodic and that Birkhoff's theorem applies.

Proposition 2.6. *Assume (A1) and (A2). In addition, suppose that $\pi_{\mathcal{C}}$ and π are the unique invariant probability measures for Q and P , respectively. Then $\tilde{\pi}$ is the unique invariant probability measure for R .*

Proof. By Proposition 2.5, $\tilde{\pi}$ is an invariant probability measure for R . We show by contradiction that it is the unique invariant probability measure for R . First, assume that if μ is an other invariant probability measure for R , then there exists $c > 0$ such that for every $A \in \mathcal{E}$,

$$\tilde{\pi}(A) \leq c\mu(A). \quad (2.10)$$

Now, if R does not admit a unique probability measure, then there exist, by [37, Theorem 1.4.6(ii)], at least two distinct probability measures μ_1 and μ_2 that are mutually singular, *i.e.*, there exists $A \in \mathcal{E} \otimes \mathcal{E}_{\mathcal{C}}$ such that $\mu_1(A) = \mu_2(A^c) = 0$. Since μ_1 and μ_2 satisfy (2.10) for some constants c_1 and c_2 , we obtain that $\tilde{\pi}(A) = \tilde{\pi}(A^c) = 0$, which, finally, implies that $\tilde{\pi}(X \times \mathcal{C}) = \tilde{\pi}(A) + \tilde{\pi}(A^c) = 0$. This is contradictory since $\tilde{\pi}$ is a probability measure.

Now, we let $\mu \in \mathbf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathcal{C}})$ be any invariant probability measure for R and show (2.10) holds for some $c > 0$. Since for every $B \in \mathcal{E}_{\mathcal{C}}$,

$$\mu(X \times B) = \mu R(X \times B) = \int_{X \times B} \mu(d(y, z)) P(y, \mathcal{C}^c) + \int_{X \times \mathcal{C}} \mu(d(y, z)) P(y, \mathcal{C}) Q(z, B),$$

it holds that

$$\int_{X \times B} \mu(d(y, z)) P(y, \mathcal{C}) = \int_{\mathcal{E} \times \mathcal{C}} \mu(d(y, z)) P(y, \mathcal{C}) Q(z, B).$$

In other words, the finite nonnegative measure $\mathcal{E}_{\mathcal{C}} \ni B \mapsto \int_{X \times B} \mu(d(y, z)) P(y, \mathcal{C})$ is invariant for Q , and since $\pi_{\mathcal{C}}$ is supposed to be the unique invariant probability measure of Q , there exists a constant $\zeta \geq 0$ such that for every $B \in \mathcal{E}_{\mathcal{C}}$,

$$\int_{X \times B} \mu(d(y, z)) P(y, \mathcal{C}) = \zeta \pi_{\mathcal{C}}(B). \quad (2.11)$$

2.3. The Kick-Kac teleportation process

We show by contradiction that $\zeta > 0$. Indeed, if $\zeta = 0$, then setting $\mathbf{B} = \mathbf{C}$ in (2.11) yields

$$\int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}) = 0. \quad (2.12)$$

Consequently, for every $\mathbf{A} \in \mathcal{E}$,

$$\begin{aligned} \mu(\mathbf{A} \times \mathbf{C}) &= \mu R(\mathbf{A} \times \mathbf{C}) \\ &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}^c \cap \mathbf{A}) + \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}) Q(z, \mathbf{C} \cap \mathbf{A}) \\ &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}^c \cap \mathbf{A}) + 0 \\ &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}^c \cap \mathbf{A}) + \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C} \cap \mathbf{A}) \\ &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{A}). \end{aligned}$$

Therefore, the probability measure $\mathcal{E} \ni \mathbf{A} \mapsto \mu(\mathbf{A} \times \mathbf{C})$ is invariant for P , and since π is the unique P -invariant probability measure, it holds that $\mu(\cdot \times \mathbf{C}) = \pi$. This allows us to rewrite (2.12) according to

$$0 = \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}) = \int_{\mathbf{E} \times \mathbf{C}} \pi(dy) P(y, \mathbf{C}) = \pi(\mathbf{C}),$$

which contradicts Remark 2.3. Thus, (2.11) holds with $\zeta > 0$. Consequently, for every $h \in \mathbf{F}_+(\mathcal{E} \otimes \mathcal{E}_{\mathbf{C}})$,

$$\begin{aligned} \int_{\mathbf{C} \times \mathbf{C}} \mu(d(y, z)) h(y, z) &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) R(h \mathbb{1}_{\mathbf{C} \times \mathbf{C}})(y, z) \\ &= \int_{\mathbf{E} \times \mathbf{C}} \mu(d(y, z)) P(y, \mathbf{C}) \int_{\mathbf{C}} Q(z, dz') h(z', z) = \zeta \int \pi_{\mathbf{C}}(dz) h(z, z). \end{aligned} \quad (2.13)$$

The probability measure μ is invariant for R , but since $\mathbf{C} \times \mathbf{C}$ is not necessarily μ -accessible we cannot use Kac's theorem directly. Instead, Remark 2.2 shows that for every bounded function $h \in \mathbf{F}_+(\mathcal{E} \otimes \mathcal{E}_{\mathbf{C}})$,

$$\begin{aligned} \mu(h) &\geq \int_{\mathbf{C} \times \mathbf{C}} \mu(d(y, z)) \mathbb{E}_{(y, z)}^R \left[\sum_{k=0}^{\sigma_{\mathbf{C} \times \mathbf{C}} - 1} h(Y_k, Z_k) \right] \\ &= \zeta \int \pi_{\mathbf{C}}(dz) \mathbb{E}_z^P \left[\sum_{k=0}^{\sigma_{\mathbf{C}} - 1} h(X_k, z) \right] = \frac{\zeta \check{\pi}(h)}{\pi(\mathbf{C})}, \end{aligned}$$

where the penultimate and last equalities follow from (2.13) and (2.8), respectively. Since $\zeta > 0$, we have shown (2.10). The proof is completed. \square

The fact that $\check{\pi}$ is a unique invariant probability measure for R allows us to apply [37, Theorem 5.2.6] and Birkhoff's theorem to establish that for every measurable function $f : \mathsf{X} \times \mathsf{C} \rightarrow \mathbb{R}$ such that $\check{\pi}(|f|) < +\infty$,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \check{\pi}(f), \quad \mathbb{P}_{\check{\pi}}^R\text{-a.s.} \quad (2.14)$$

However, this property is not fully satisfying since it only holds under stationarity, *i.e.*, under $\mathbb{P}_{\check{\pi}}^R$. From a simulation perspective, it is crucial that the same result holds irrespective of how the chain is initialized. Our next result addresses this question and establishes that (2.14) holds for an arbitrary initial distribution, *i.e.*, when $\mathbb{P}_{\check{\pi}}^R$ is replaced by \mathbb{P}_{ξ}^R , where $\xi \in \mathsf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathsf{C}})$ is arbitrary. Interestingly, the assumptions under which we operate when deriving this result also turn out to be necessary and, therefore, impossible to weaken.

Theorem 2.7. *Assume (A1) and (A2). In addition, suppose that the Markov kernel R admits a unique invariant probability measure $\check{\pi}$. Then the two conditions*

- (a) *for every $x \in \mathsf{X}$, $\mathbb{P}_x^P(\sigma_{\mathsf{C}} < +\infty) = 1$ and*
- (b) *for every $x \in \mathsf{C}$ and bounded measurable function $h : \mathsf{C} \rightarrow \mathbb{R}$,*

$$\lim_{\ell \rightarrow +\infty} \ell^{-1} \sum_{k=0}^{\ell-1} h(X_k) = \pi_{\mathsf{C}}(h), \quad \mathbb{P}_x^Q\text{-a.s.},$$

hold if and only if for every $\xi \in \mathsf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathsf{C}})$ and measurable function $f : \mathsf{X} \times \mathsf{C} \rightarrow \mathbb{R}$ such that $\check{\pi}(|f|) < +\infty$,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} f(Y_k, Z_k) = \check{\pi}(f), \quad \mathbb{P}_{\xi}^R\text{-a.s.} \quad (2.15)$$

It is worthwhile to note that Theorem 2.7 does not presuppose a law of large numbers for the Markov kernel P . The proof of Theorem 2.7 relies on the following necessary and sufficient conditions—based on properties of *harmonic functions* for P (*i.e.* measurable functions h for which $P|h| < \infty$ and $Ph = h$)—for a Markov chain to satisfy a law of large numbers.

Proposition 2.8. *Let P be a Markov kernel on $\mathsf{X} \times \mathcal{E}$ admitting an invariant probability measure π . Then the two following conditions are equivalent.*

- (i) *For every $\xi \in \mathsf{M}_1(\mathcal{E})$ and measurable function $g : \mathsf{X} \rightarrow \mathbb{R}$ such that $\pi(|g|) < +\infty$,*

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} g(X_k) = \pi(g), \quad \mathbb{P}_{\xi}^P\text{-a.s.}$$

- (ii) *Every bounded harmonic function $h : \mathsf{X} \rightarrow \mathbb{R}$ for P is constant.*

2.3. The Kick-Kac teleportation process

Since the equivalence provided by Proposition 2.8 has not, as far as we know, been established in the literature before, we provide a complete proof of this result in Section 2.7.2. We now have all the tools needed for proving Theorem 2.7.

Proof of Theorem 2.7. We aim to establish the equivalence between the conditions (a) and (b) and the law of large numbers (2.15). We first assume that (a) and (b) hold. Since $\tilde{\pi}$ is an invariant probability measure for R , Proposition 2.8 applies to R . Therefore, given an arbitrary harmonic function h for R , it is sufficient to establish that h is constant. With this aim in mind, we prove that for every $(y, z) \in \mathsf{X} \times \mathsf{C}$,

$$h(y, z) = \int_{\mathsf{C}} Q(z, dz') h(z', z'). \quad (2.16)$$

Indeed, if (2.16) is true, then $h(y, z)$ does not depend on y (since the right-hand side of (2.16) does not depend on y) and then $g(z) := h(z, z)$ is a bounded harmonic function for Q since (2.16) yields $g = Qg$. Finally, combining (b) with Proposition 2.8 applied to Q shows that g is a constant function. Thus, also h is constant.

We now turn to the proof of (2.16). Set $\mathsf{A} := \{\sigma_{\bar{\mathsf{C}}} < +\infty\}$ where $\bar{\mathsf{C}} = \mathsf{C} \times \mathsf{C}$. Since $Y_{\sigma_{\bar{\mathsf{C}}}} = Z_{\sigma_{\bar{\mathsf{C}}}}$ on A , combining Lemma 3(ii) with (a) implies

$$\begin{aligned} \mathbb{E}_{(y,z)}^R \left[h(Y_{\sigma_{\bar{\mathsf{C}}}}, Z_{\sigma_{\bar{\mathsf{C}}}}) \mathbb{1}_{\mathsf{A}} \right] &= \mathbb{E}_{(y,z)}^R \left[h(Z_{\sigma_{\bar{\mathsf{C}}}}, Z_{\sigma_{\bar{\mathsf{C}}}}) \mathbb{1}_{\mathsf{A}} \right] \\ &= \mathbb{P}_y^P(\sigma_{\mathsf{C}} < \infty) \int_{\mathsf{C}} Q(z, dz') h(z', z') = \int_{\mathsf{C}} Q(z, dz') h(z', z'). \end{aligned} \quad (2.17)$$

On the other hand, by [37, Proposition 5.2.2(ii)] there exists a random variable W that is invariant for the shift operator θ and such that $\mathbb{E}_{(y,z)}^R[W] = h(y, z)$ for all $(y, z) \in \mathsf{X} \times \mathsf{C}$. Thus, by the strong Markov property, since $W \circ \theta^{\sigma_{\bar{\mathsf{C}}}} = W$ on A ,

$$\begin{aligned} \mathbb{E}_{(y,z)}^R \left[h(Y_{\sigma_{\bar{\mathsf{C}}}}, Z_{\sigma_{\bar{\mathsf{C}}}}) \mathbb{1}_{\mathsf{A}} \right] &= \mathbb{E}_{(y,z)}^R \left[\mathbb{E}_{(Y_{\sigma_{\bar{\mathsf{C}}}}, Z_{\sigma_{\bar{\mathsf{C}}}})}^R [W] \mathbb{1}_{\mathsf{A}} \right] = \mathbb{E}_{(y,z)}^R [W \circ \theta^{\sigma_{\bar{\mathsf{C}}}} \mathbb{1}_{\mathsf{A}}] \\ &= \mathbb{E}_{(y,z)}^R [W] = h(y, z), \end{aligned}$$

where we used Lemma 3(i) in the penultimate equality. Combining this equality with (2.17) yields (2.16), and we have finally established (2.15).

Conversely, assuming (2.15), we prove that (a) and (b) hold. Let $(y_0, z_0) \in \mathsf{X} \times \mathsf{C}$; then, applying (2.15) to the function $f : \mathsf{X} \times \mathsf{C} \ni (y, z) \mapsto \mathbb{1}_{\mathsf{C}}(y) = \mathbb{1}_{\bar{\mathsf{C}}}(y, z)$ and the measure $\xi = \delta_{(y_0, z_0)}$ yields, $\mathbb{P}_{(y_0, z_0)}^R$ -a.s.,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} \mathbb{1}_{\mathsf{C}}(Y_k) = \lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} \mathbb{1}_{\bar{\mathsf{C}}}(Y_k, Z_k) = \tilde{\pi}(\bar{\mathsf{C}}) = \pi(\mathsf{C}) > 0, \quad (2.18)$$

where the positivity of $\pi(\mathsf{C})$ follows from Remark 2.3 under (A1). Therefore, for all $(y_0, z_0) \in \mathsf{X} \times \mathsf{C}$, $1 = \mathbb{P}_{(y_0, z_0)}^R(\sigma_{\bar{\mathsf{C}}} < \infty) = \mathbb{P}_{y_0}^P(\sigma_{\mathsf{C}} < \infty)$, where the last equality follows from Lemma 3 by letting $v \equiv 1$ and $n \rightarrow \infty$ in (2.47). This shows (a).

We now turn to (b). Pick arbitrarily $h \in F_b(\mathcal{E}_C)$ and $x \in C$. Applying (2.15) to the function $f : X \times C \ni (y, z) \mapsto \mathbb{1}_C(y)h(y)$ and the measure $\xi = \delta_{(x,x)}$ yields

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} \mathbb{1}_C(Y_k)h(Y_k) = \pi(\mathbb{1}_C h), \quad \mathbb{P}_{(x,x)}^R\text{-a.s.}$$

Combining this limit with (2.18) provides

$$\lim_{n \rightarrow +\infty} \frac{\sum_{k=0}^{n-1} \mathbb{1}_C(Y_k)h(Y_k)}{\sum_{k=0}^{n-1} \mathbb{1}_C(Y_k)} = \frac{\pi(\mathbb{1}_C h)}{\pi(C)} = \pi_C(h), \quad \mathbb{P}_{(x,x)}^R\text{-a.s.}$$

Therefore, $\lim_{\ell \rightarrow +\infty} \ell^{-1} \sum_{k=1}^{\ell} h(Y_{\sigma_C^k}) = \pi_C(h)$, $\mathbb{P}_{(x,x)}^R$ -a.s. However, by Lemma 3(iv), the embedded process $(Y_{\sigma_C^k})_{k \in \mathbb{N}^*}$ is, under $\mathbb{P}_{(x,x)}^R$, a Markov chain with transition kernel Q and initial distribution δ_x . Thus,

$$1 = \mathbb{P}_{(x,x)}^R \left(\lim_{\ell \rightarrow +\infty} \ell^{-1} \sum_{k=1}^{\ell} h(Y_{\sigma_C^k}) = \pi_C(h) \right) = \mathbb{P}_x^Q \left(\lim_{\ell \rightarrow +\infty} \ell^{-1} \sum_{k=0}^{\ell-1} h(X_k) = \pi_C(h) \right),$$

which establishes (b). This completes the proof. \square

2.3.4 Geometric ergodicity of the KKT process

In this section, we introduce the following additional conditions under which we will establish that R is geometrically ergodic in total variation.

(A3) There exist $(\lambda, b) \in (0, 1) \times \mathbb{R}_+^*$ and a measurable function $V_P : X \rightarrow [1, +\infty)$ such that

$$PV_P \leq \lambda V_P + b \mathbb{1}_C \quad \text{and} \quad \sup_{y \in C} V_P(y) < +\infty. \quad (2.19)$$

Assumption (A3), which is a typical geometric drift condition, implies that C is accessible for P ; indeed, under (A3), by [37, Proposition 4.3.3], $\mathbb{P}_x^P(\sigma_C < +\infty) = 1$ for every $x \in X$.

Let $\epsilon > 0$ and $\nu \in M_1(\mathcal{E}_C)$. We say that $D \in \mathcal{E}_C$ is a $(1, \epsilon\nu)$ -small set for Q if for every $x \in D$ and $A \in \mathcal{E}_C$, $Q(x, A) \geq \epsilon\nu(A)$.

(A4)

- (i) There exists an accessible $(1, \epsilon\nu)$ -small set $D \in \mathcal{E}_C$ for Q such that $\nu(D) > 0$.
- (ii) It holds that $\inf_{z \in D} P(z, C) > 0$.
- (iii) There exist constants $(\lambda, b) \in (0, 1) \times \mathbb{R}_+^*$ and a measurable function $V_Q : C \rightarrow [1, +\infty)$ such that

$$QV_Q \leq \lambda V_Q + b \mathbb{1}_D \quad \text{and} \quad \sup_{z \in D} V_Q(z) < \infty. \quad (2.20)$$

2.3. The Kick-Kac teleportation process

Since we will assume (A3) and (A4) jointly, we can always, for simplicity, assume that (2.19) and (2.20) hold for the *same* constants (λ, b) (for instance, the largest ones). Note that (A4)(i) implies (see [37, Theorem 9.2.2]) that Q is irreducible, and therefore $\pi_{\mathcal{C}}$ is the unique invariant probability measure of Q under (A2). Moreover, (A4)(ii) holds true in the case where X is a topological space (equipped with its Borel σ -field \mathcal{E}) and there exists a kernel T on X and a compact set D such that $P(y, \mathsf{C}) \geq T(y, \mathsf{C}) > 0$ for all $y \in \mathsf{D}$ and the mapping $y \mapsto T(y, \mathsf{C})$ is lower semi-continuous. This setting covers, for instance, the case where P is the Markov kernel associated with the Metropolis–Hastings algorithm with positive proposal transition density.

Theorem 2.9 below states that the KKT process is geometrically ergodic provided that Q is geometrically ergodic on C and P pushes toward C according to the geometric drift condition (A3). Therefore, remarkably, no assumption concerning smallness of C with respect to P is made, and the set C may hence be potentially *large*. Note also that no assumption on geometric ergodicity of the Markov kernel P is needed. Theorem 2.9 conveys the idea that if a Markov chain evolving according to P exhibits poor ergodic behavior on C , then the action of changing its dynamics in the same region to one governed by Q , where Q is geometrically ergodic on C , may result in a geometrically ergodic KKT process.

Theorem 2.9. *Assume (A1)–(A4). Then there exist constants $C > 0$ and $\rho \in (0, 1)$ such that for every $\mu \in \mathsf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathsf{C}})$ and $n \in \mathbb{N}$,*

$$\|\mu R^n - \tilde{\pi}\|_{\text{TV}} \leq C \rho^n \int_{\mathsf{E} \times \mathsf{C}} V_P(y) V_Q(z) \mu(\mathrm{d}(y, z)). \quad (2.21)$$

Our approach for establishing the geometric ergodicity in Theorem 2.9 is non-standard. Indeed, even though it can be easily seen that $\mathsf{D} \times \mathsf{D}$ is a small set for R , it is not obvious (and this remains an open question) how to construct, under (A3)–(A4), a drift function for R outside $\mathsf{D} \times \mathsf{D}$ by combining the drift functions V_P and V_Q associated with P and Q , respectively. Instead, the geometrically ergodic bound (2.21) is obtained through a delicate application of [37, Theorem 11.4.2].

Proof of Theorem 2.9. Set $\bar{\mathsf{C}} := \mathsf{C} \times \mathsf{C}$ and $\bar{\mathsf{D}} := \mathsf{D} \times \mathsf{D}$. Under (A3), [37, Proposition 4.3.3(ii)] allows us to write, for every $y \in \mathsf{X}$,

$$\mathbb{E}_y^P[\lambda^{-\sigma_{\mathsf{C}}}] \leq V_P(y) + b/\lambda < +\infty. \quad (2.22)$$

Therefore, $\mathbb{P}_y^P(\sigma_{\mathsf{C}} < +\infty) = 1$ for every $y \in \mathsf{X}$, and by Lemma 3(i), for every $\ell \in \mathbb{N}^*$ and $(y, z) \in \mathsf{X} \times \mathsf{C}$,

$$\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathsf{C}}}^{\ell} < +\infty) = 1. \quad (2.23)$$

Moreover, by Lemma 4 and (2.22) again, $\bar{\mathsf{D}} = \mathsf{D} \times \mathsf{D}$ is an accessible $(1, \delta\epsilon\bar{\nu})$ -small set for R , where $\delta := \inf_{z \in \mathsf{D}} P(z, \mathsf{C})$ and $\bar{\nu}$ is the probability measure $\bar{\nu} : \mathcal{E} \otimes \mathcal{E}_{\mathsf{C}} \ni \mathsf{A} \mapsto \int_{\mathsf{C}} \nu(\mathrm{d}z) \mathbb{1}_{\mathsf{A}}(z, z)$ which satisfies $\bar{\nu}(\bar{\mathsf{D}}) > 0$. Thus, in order to apply [37, Theorem 11.4.2], we only need to show that there exists some $\gamma > 1$ such that

$$\sup_{(y,z) \in \bar{\mathsf{D}}} \mathbb{E}_{(y,z)}^R[\gamma^{\sigma_{\bar{\mathsf{D}}}}] < +\infty. \quad (2.24)$$

In order to establish (2.24) we first show that for every $(y, z) \in \mathbf{X} \times \mathbf{C}$,

$$\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathbf{D}}} < +\infty) = 1. \quad (2.25)$$

For this purpose, using (2.23) we can write $\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathbf{D}}} < +\infty) = \sum_{k=1}^{+\infty} \mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathbf{D}}} = \sigma_{\bar{\mathbf{C}}}^k)$, where

$$\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathbf{D}}} = \sigma_{\bar{\mathbf{C}}}^k) = \mathbb{E}_{(y,z)}^R \left[\mathbb{1}_{\mathbf{D}}(Z_{\sigma_{\bar{\mathbf{C}}}^k}) \prod_{\ell=1}^{k-1} \mathbb{1}_{\mathbf{D}^c}(Z_{\sigma_{\bar{\mathbf{C}}}^{\ell}}) \right] = \mathbb{P}_z^Q(\sigma_{\mathbf{D}} = k);$$

here the last equality is obtained by applying Lemma 3(iii) for $g_{\ell} \equiv \mathbf{1}$ for all $\ell \leq k$, $h_{\ell} \equiv \mathbb{1}_{\mathbf{D}^c}$ for all $\ell < k$, and $h_k \equiv \mathbb{1}_{\mathbf{D}}$. Thus, since [37, Proposition 4.3.3(ii)] implies that

$$\mathbb{E}_z^Q[\lambda^{-\sigma_{\mathbf{D}}}] \leq V(z) + b/\lambda < +\infty, \quad z \in \mathbf{C}, \quad (2.26)$$

it holds that $\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathbf{D}}} < +\infty) = \mathbb{P}_z^Q(\sigma_{\mathbf{D}} < +\infty) = 1$, which completes the proof (2.25).

We now turn to the proof of (2.24). Set $\eta := \sup_{y \in \mathbf{C}} V_P(y) + b/\lambda$ and pick $\gamma \in (1, \exp(\log^2 \lambda / \log \eta))$. Using (2.22), we have

$$1 \leq \lambda^{-1} \leq \sup_{y \in \mathbf{C}} \mathbb{E}_y^P[\lambda^{-\sigma_{\mathbf{C}}}] \leq \eta. \quad (2.27)$$

Therefore, we get

$$\gamma \leq \exp(\log^2 \lambda / \log \eta) = (\lambda^{-1})^{-\log \lambda / \log \eta} \leq \lambda^{-1}. \quad (2.28)$$

By Jensen's inequality and (2.27) it holds, for every $x \in \mathbf{C}$,

$$\begin{aligned} \mathbb{E}_x^P[\gamma^{\sigma_{\mathbf{C}}}] &= \mathbb{E}_x^P[\lambda^{\sigma_{\mathbf{C}} \log \gamma / \log \lambda}] \leq \mathbb{E}_x^P[\lambda^{-\sigma_{\mathbf{C}}}^{-\log \gamma / \log \lambda}] \\ &\leq \eta^{-\log \gamma / \log \lambda} = \exp\left(-\frac{\log \eta \log \gamma}{\log \lambda}\right) \leq \lambda^{-1}. \end{aligned} \quad (2.29)$$

We will soon make use of the bound (2.29); however, first, let $(y, z) \in \mathbf{X} \times \mathbf{C}$ and write, using (2.25) and (2.23),

$$\mathbb{E}_{(y,z)}^R[\gamma^{\sigma_{\bar{\mathbf{D}}}}] = \mathbb{E}_{(y,z)}^R \left[\gamma^{\sigma_{\bar{\mathbf{D}}}} \mathbb{1}_{\{\sigma_{\bar{\mathbf{D}}} < +\infty\}} \right] = \sum_{k=1}^{+\infty} r_k, \quad \text{with } r_k := \mathbb{E}_{(y,z)}^R \left[\gamma^{\sigma_{\bar{\mathbf{C}}}^k} \mathbb{1}_{\{\sigma_{\bar{\mathbf{D}}} = \sigma_{\bar{\mathbf{C}}}^k\}} \right]. \quad (2.30)$$

Applying Lemma 3(iii) with $g_{\ell}(t) \equiv \gamma^t$ for $\ell \in \{1, \dots, k\}$ and $t \in \mathbb{N}$, $h_{\ell} \equiv \mathbb{1}_{\mathbf{D}^c}$ for $\ell < k$, and $h_k \equiv \mathbb{1}_{\mathbf{D}}$ (and using the convention $\sigma_{\bar{\mathbf{C}}}^0 = 0$) yields

$$\begin{aligned} r_k &= \mathbb{E}_{(y,z)}^R \left[\gamma^{\sigma_{\bar{\mathbf{C}}}^k} \mathbb{1}_{\mathbf{D}}(Z_{\sigma_{\bar{\mathbf{C}}}^k}) \prod_{\ell=1}^{k-1} \mathbb{1}_{\mathbf{D}^c}(Z_{\sigma_{\bar{\mathbf{C}}}^{\ell}}) \right] \\ &= \mathbb{E}_{(y,z)}^R \left[\left(\prod_{\ell=1}^k \gamma^{\sigma_{\bar{\mathbf{C}}}^{\ell} - \sigma_{\bar{\mathbf{C}}}^{\ell-1}} \right) \mathbb{1}_{\mathbf{D}}(Z_{\sigma_{\bar{\mathbf{C}}}^k}) \prod_{\ell=1}^{k-1} \mathbb{1}_{\mathbf{D}^c}(Z_{\sigma_{\bar{\mathbf{C}}}^{\ell}}) \right] \\ &= \mathbb{E}_y^P[\gamma^{\sigma_{\mathbf{C}}}] \mathbb{E}_z^Q \left[\mathbb{1}_{\mathbf{D}}(X_k) \prod_{\ell=1}^{k-1} \mathbb{1}_{\mathbf{D}^c}(X_{\ell}) \mathbb{E}_{X_{\ell}}^P[\gamma^{\sigma_{\mathbf{C}}}] \right] \\ &\leq \left(\sup_{x \in \mathbf{C}} \mathbb{E}_x^P[\gamma^{\sigma_{\mathbf{C}}}] \right)^{k-1} \mathbb{E}_y^P[\gamma^{\sigma_{\mathbf{C}}}] \mathbb{P}_z^Q(\sigma_{\mathbf{D}} = k). \end{aligned}$$

2.4. A general Kick-Kac teleportation process

Now, combining this bound with (2.28) and (2.29) yields for any $(y, z) \in \mathbf{E} \times \mathbf{C}$,

$$r_k \leq \lambda^{-k+1} \mathbb{E}_y^P[\lambda^{-\sigma_C}] \mathbb{P}_z^Q(\sigma_D = k).$$

Plugging the previous inequality into (2.30) and using (2.22) and (2.26) we obtain

$$\mathbb{E}_{(y,z)}^R[\gamma^{\sigma_D}] \leq \lambda \mathbb{E}_y^P[\lambda^{-\sigma_C}] \mathbb{E}_z^Q[\lambda^{-\sigma_D}] \leq \lambda(V_P(y) + b/\lambda)(V_Q(z) + b/\lambda). \quad (2.31)$$

Thus, using (A3) and (A4)(iii), we conclude that

$$\sup_{(y,z) \in \bar{\mathbf{D}}} \mathbb{E}_{(y,z)}^R[\gamma^{\sigma_D}] \leq \lambda(\sup_{y \in \bar{\mathbf{D}}} V_P(y) + b/\lambda)(\sup_{z \in \mathbf{D}} V_Q(z) + b/\lambda) < \infty,$$

which completes the proof of (2.24).

Finally, all the assumptions of [37, Theorem 11.4.2] are satisfied and there exist $c > 0$ and $\rho \in (0, 1)$ such that for every $\mu \in \mathbf{M}_1(\mathcal{E} \otimes \mathcal{E}_C)$ and $n \in \mathbb{N}$,

$$\|\mu R^n - \tilde{\pi}\|_{\text{TV}} \leq c \rho^n \mathbb{E}_\mu^R[\gamma^{\sigma_D}] \leq c \lambda \rho^n \int \mu(d(y, z))(V_P(y) + b/\lambda)(V_Q(z) + b/\lambda),$$

where we used (2.31) in the last step. Combining this with the bound $(V_P(y) + b/\lambda)(V_Q(z) + b/\lambda) \leq (1 + b/\lambda)^2 V_P(y)V_Q(z)$ and letting $C = c\lambda(1 + b/\lambda)^2$ completes the proof of (2.21). \square

2.4 A general Kick-Kac teleportation process

2.4.1 A generalization of Kac's theorem

Recall that in order to construct the Markov teleportation process, we combined the kernel P , having invariant probability measure π , with an auxiliary kernel Q , having π_C as invariant probability measure. We now generalize this construction by replacing π_C by a more general probability measure $\tilde{\pi} \in \mathbf{M}_1(\mathcal{E})$ which is assumed to admit a bounded density with respect to π ; in other words, by assumption there exists $M \geq 1$ such that for every $x \in \mathbf{E}$,

$$\frac{d\tilde{\pi}}{d\pi}(x) \leq M. \quad (2.32)$$

Under this assumption we define the $[0, 1]$ -valued measurable function

$$\alpha(x) := \frac{1}{M} \frac{d\tilde{\pi}}{d\pi}(x), \quad x \in \mathbf{X}. \quad (2.33)$$

Of course, $\tilde{\pi} = \pi_C$ satisfies (2.32) with $M = 1/\pi(C)$, and in that case, $\alpha = \mathbb{1}_C$.

Before introducing the general Markov teleportation process, we first provide a generalization of Kac's theorem. Let $\bar{\mathbf{X}} := \mathbf{X} \times [0, 1]$ and $\bar{\mathcal{E}} := \mathcal{E} \otimes \mathcal{B}([0, 1])$ and define the Markov kernel \bar{P} on $\bar{\mathbf{X}} \times \bar{\mathcal{E}}$ by

$$\bar{P}(\bar{x}, \mathbf{A}) := \int P(x, dx') \mathbb{1}_{[0,1]}(u') \mathbb{1}_{\mathbf{A}}(x', u') du', \quad \bar{x} = (x, u) \in \bar{\mathbf{X}}, \mathbf{A} \in \bar{\mathcal{E}}. \quad (2.34)$$

Note that a transition according to \bar{P} can be decomposed into two independent moves, where the first component is updated according to the Markov kernel P and the second component is drawn from the uniform distribution on $[0, 1]$ independently of the past. For a given $\mu \in \mathbf{M}_1(\bar{\mathcal{E}})$, denote by $\bar{\mathbb{P}}_\mu$ (and $\bar{\mathbb{E}}_\mu$) the unique probability distribution (and associated expectation) on the canonical space $(\bar{\mathcal{X}}^{\mathbb{N}}, \bar{\mathcal{E}}^{\otimes \mathbb{N}})$ induced by the Markov kernel \bar{P} and the initial distribution μ . By abuse of notation, if $\mu = \delta_{(x,u)}$ for some $(x, u) \in \mathcal{X} \times [0, 1]$, we simply write $\bar{\mathbb{P}}_{x,u}$ and $\bar{\mathbb{E}}_{x,u}$ instead of $\mathbb{P}_{\delta_{(x,u)}}$ and $\mathbb{E}_{\delta_{(x,u)}}$, respectively. Moreover, when φ is a nonnegative or bounded measurable function on $(\mathcal{X}^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ such that $\bar{\mathbb{E}}_{x,u}[\varphi]$ does not depend on $u \in [0, 1]$, we indicate this by writing $\bar{\mathbb{E}}_{x,*}[\varphi]$. Finally, set $\bar{\mathcal{C}} := \{(x, u) \in \bar{\mathcal{X}} : u \leq \alpha(x)\}$ and define

$$\sigma_{\bar{\mathcal{C}}} := \inf \left\{ k \geq 1 : (X_k, U_k) \in \bar{\mathcal{C}} \right\}, \quad (2.35)$$

where $(X_k, U_k)_{k \in \mathbb{N}}$ is the canonical process on $(\bar{\mathcal{X}}^{\mathbb{N}}, \bar{\mathcal{E}}^{\otimes \mathbb{N}})$. In addition, define $\bar{\pi} := \pi \otimes \mathbf{Unif}([0, 1])$ and note that $\bar{\pi}$ is an invariant probability measure for \bar{P} . By showing that $\bar{\mathcal{C}}$ is $\bar{\pi}$ -accessible and applying Kac's theorem to \bar{P} we obtain the following generalization of Kac's theorem.

Proposition 2.10. *Let P be a Markov kernel on $\mathcal{X} \times \mathcal{E}$ with invariant probability measure π . Let $\alpha : \mathcal{X} \rightarrow [0, 1]$ be a measurable function such that $\{x \in \mathcal{X} : \alpha(x) > 0\}$ is π -accessible for P . Then, $\bar{\mathcal{C}}$ is $\bar{\pi}$ -accessible. Moreover,*

$$\pi = \pi_\alpha^0 = \pi_\alpha^1, \quad (2.36)$$

where for every measurable nonnegative or bounded function f on $(\mathcal{X}, \mathcal{E})$,

$$\pi_\alpha^0(f) = \int \pi(dx) \alpha(x) \bar{\mathbb{E}}_{x,*} \left[\sum_{k=0}^{\sigma_{\bar{\mathcal{C}}}-1} f(X_k) \right], \quad \pi_\alpha^1(f) = \int \pi(dx) \alpha(x) \bar{\mathbb{E}}_{x,*} \left[\sum_{k=1}^{\sigma_{\bar{\mathcal{C}}}} f(X_k) \right].$$

Proof. To apply Kac's theorem to the set $\bar{\mathcal{C}}$ for the kernel \bar{P} we need to check that $\bar{\mathcal{C}}$ is $\bar{\pi}$ -accessible for \bar{P} . Since, by assumption, $\{\alpha > 0\}$ is π -accessible for P , we conclude that $\sum_{k=1}^{\infty} \mathbb{P}_x^P(\alpha(X_k) > 0) > 0$ for π -almost all $x \in \mathcal{X}$. Note that since $\mathbb{P}_x^P(\alpha(X_k) > 0) > 0$ if and only if $\mathbb{E}_x^P[\alpha(X_k)] > 0$, it holds that

$$\sum_{k=1}^{\infty} \bar{P}^k((x, u), \bar{\mathcal{C}}) = \sum_{k=1}^{\infty} \bar{\mathbb{E}}_{x,u} \left[\mathbb{1}_{\{U_k \leq \alpha(X_k)\}} \right] = \sum_{k=1}^{\infty} \mathbb{E}_x^P[\alpha(X_k)] > 0$$

for $\bar{\pi}$ -almost all $(x, u) \in \bar{\mathcal{X}}$. Hence, $\bar{\mathcal{C}}$ is $\bar{\pi}$ -accessible for \bar{P} , and Kac's theorem (Theorem 2.1) applies. Thus, for every $h \in \mathbf{F}_+(\bar{\mathcal{E}})$,

$$\begin{aligned} \bar{\pi}(h) &= \int_{\bar{\mathcal{C}}} \bar{\pi}(d(x, u)) \bar{\mathbb{E}}_{x,u} \left[\sum_{k=0}^{\sigma_{\bar{\mathcal{C}}}-1} h(X_k, U_k) \right] \\ &= \int \pi(dx) \int_{[0,1]} \mathbb{1}_{\{u \leq \alpha(x)\}} \bar{\mathbb{E}}_{x,u} \left[\sum_{k=0}^{\sigma_{\bar{\mathcal{C}}}-1} h(X_k, U_k) \right] du. \end{aligned}$$

2.4. A general Kick-Kac teleportation process

Now, if $f \in F_+(\mathcal{E})$, then setting $h(x, u) = f(x)$ yields

$$\begin{aligned} \pi(f) &= \bar{\pi}(h) = \int \pi(dx) \int_{[0,1]} \mathbb{1}_{\{u \leq \alpha(x)\}} \bar{\mathbb{E}}_{x,u} \left[\sum_{k=0}^{\sigma_{\bar{c}}-1} f(X_k) \right] du \\ &= \int \pi(dx) \alpha(x) \bar{\mathbb{E}}_{x,*} \left[\sum_{k=0}^{\sigma_{\bar{c}}-1} f(X_k) \right], \end{aligned}$$

where the last equality stems from the fact that $\bar{\mathbb{E}}_{x,u}[\sum_{k=0}^{\sigma_{\bar{c}}-1} f(X_k)]$ does not depend on u . Hence, $\pi = \pi_{\alpha}^0$. The proof of $\pi = \pi_{\alpha}^1$ follows the same lines and is omitted for brevity. \square

Remark 2.11. *Similarly to Corollary 1, the generalized Kac's theorem (2.36) also holds if P admits a unique invariant probability measure π such that $\pi(\alpha) > 0$. Indeed, in that case, $\pi(\alpha > 0) > 0$, and by Lemma 2, $\{\alpha > 0\}$ is π -accessible for P . Therefore, Proposition 2.10 applies, implying (2.36).*

If $\{x \in \mathsf{X} : \alpha(x) > 0\}$ is π -accessible for P , we can, on the basis of Proposition 2.10 and Proposition 2.5, define a KKT sampler using the extended Markov kernel \bar{P} as basis kernel and a Markov kernel \bar{Q} leaving $\bar{\pi}_{\bar{c}}$ invariant as teleportation kernel. Moreover, note that since $d\tilde{\pi}/d\pi \propto \alpha$, it is easy to verify that a Markov kernel \bar{Q} leaving $\bar{\pi}_{\bar{c}}$ invariant can be formed by picking some $\tilde{\pi}$ -invariant Markov kernel Q and setting $\bar{Q} := Q \otimes \mathbf{Unif}(0, 1)|_{\bar{c}}$, i.e., \bar{Q} is the Markov kernel on $\bar{C} \times \bar{\mathcal{E}}_{\bar{c}}$ given by

$$\bar{Q}(\bar{x}, \mathsf{A}) = \iint Q(x, dx') \frac{\mathbb{1}_{[0, \alpha(x')]}(u')}{\alpha(x')} \mathbb{1}_{\mathsf{A}}(x', u') du', \quad \bar{x} = (x, u) \in \bar{C}, \mathsf{A} \in \bar{\mathcal{E}}_{\bar{c}}.$$

Using Proposition 2.10, we may now construct the *general Kick-Kac teleportation (GKKT) process* described in Algorithm 3, where Q is, as above, some Markov kernel leaving $\tilde{\pi}$ invariant. Finally, it is worth mentioning that all the results that we have derived for the KKT process can be obtained also for the GKKT process by replacing the conditions on P and Q by similar conditions on \bar{P} and \bar{Q} .

Remark 2.12. *Assume that the target distribution π is known only up to multiplicative constant, i.e., $\pi = \pi_u/\pi_u(\mathsf{X})$, where π_u is a known, unnormalized finite measure on $(\mathsf{X}, \mathcal{E})$. Moreover, assume that there exists another unnormalized measure $\tilde{\pi}_u$ on $(\mathsf{X}, \mathcal{E})$, dominated by π_u , and a finite constant M_u such that for every $x \in \mathsf{X}$,*

$$\frac{d\tilde{\pi}_u}{d\pi_u}(x) \leq M_u.$$

Then it is straightforward to check that (2.32) is satisfied with $\tilde{\pi} = \tilde{\pi}_u/\tilde{\pi}_u(\mathsf{X})$ and $M = M_u\pi_u(\mathsf{X})/\tilde{\pi}_u(\mathsf{X})$. Furthermore, as the function α defined by (2.33) can be written as $\alpha(x) = M_u^{-1}(d\tilde{\pi}_u/d\pi_u)(x)$, Algorithm 3 can be implemented even if the target is known only up to a multiplicative constant.

Algorithm 3 The general Kick-Kac teleportation (GKKT) process

```

1: Initialization Draw  $(Y_0, Z_0)$ .
2: for  $k \leftarrow 1$  to  $n$  do
3:   draw  $(Y_k^*, U_k) \sim P(Y_{k-1}, \cdot) \otimes \mathbf{Unif}([0, 1])$ 
4:   if  $U_k \geq \alpha(Y_k^*)$  then
5:     set  $(Y_k, Z_k) \leftarrow (Y_k^*, Z_{k-1})$ 
6:   else
7:     draw  $Z_k \sim Q(Z_{k-1}, \cdot)$ 
8:     set  $Y_k \leftarrow Z_k$ 
9:   end if
10: end for

```

Remark 2.13. As explained above, Algorithm 3 can be viewed as a special instance of Algorithm 2, parameterized by the base kernel \bar{P} , the critical region \bar{C} and the teleportation kernel \bar{Q} , where the generation of the uniformly distributed component of the teleportative move governed by \bar{Q} is omitted (as this draw is never used later in the algorithm, since neither $\bar{Q}((x, u), \cdot)$ nor $\bar{P}((x, u), \cdot)$ de facto depends on u). Recalling the definitions of \bar{P} and \bar{C} in (2.34) and (2.35), respectively, Lines 3–4 in Algorithm 3 are equivalent to

3': draw $(Y_k^*, U_k) \sim \bar{P}((Y_{k-1}, U_{k-1}), \cdot)$
4': **if** $(Y_k^*, U_k) \notin \bar{C}$ **then**

which exactly correspond to Lines 3–4 in Algorithm 2 with Y_k^* replaced by (Y_k^*, U_k) , P by \bar{P} , and C by \bar{C} . Moreover, alternatively, Lines 3–4 in Algorithm 3 can be expressed as

3'': draw $Y_k^* \sim P(Y_{k-1}, \cdot)$ and, conditionally on Y_k^* , $B_k \sim \text{Bernoulli}(\alpha(Y_k^*))$
4'': **if** $B_k = 0$ **then**

2.4.2 Revisiting the hybrid kernel of [20]

In this section, we compare our methodology to the one proposed in [20]. We first recall the construction of the hybrid kernel introduced in that work and then show that this can be seen as a particular instance of the GKKT process. Let P be a Markov kernel on $X \times \mathcal{E}$ leaving π invariant and let $\phi \in \mathbf{M}_1(\mathcal{E})$ be a so-called *re-entry proposal distribution*. We assume that π and ϕ have densities with respect to the same dominating measure ν , and denote, by abuse of notation, these densities by the same symbols π and ϕ , respectively. The construction of [20] follows two steps: using P and ϕ , we first construct a process $(Y'_k)_{k \in \mathbb{N}}$ on $X \cup \{\mathbf{a}\}$, where \mathbf{a} is an ‘artificial’ atom; then we obtain $(Y_k)_{k \in \mathbb{N}}$ by removing every occurrence of the state \mathbf{a} from $(Y'_k)_{k \in \mathbb{N}}$. Letting $c > 0$, the transitions of $(Y'_k)_{k \in \mathbb{N}}$ can be described as follows.

2.4. A general Kick-Kac teleportation process

Case 1 : $Y'_{k-1} = y \in \mathbf{X}$	Case 2 : $Y'_{k-1} = \mathbf{a}$
Draw $(Y_k^*, U_k) \sim P(Y_k, \cdot) \otimes \mathbf{Unif}([0, 1])$; if $U_k \geq \left(1 \wedge \frac{c\phi(Y_k^*)}{\pi(Y_k^*)}\right)$, then set $Y'_k = Y_k^*$; otherwise, set $Y'_k = \mathbf{a}$.	Draw $(Y_k^*, V_k) \sim \phi \otimes \mathbf{Unif}([0, 1])$; if $V_k \leq \left(1 \wedge \frac{\pi(Y_k^*)}{c\phi(Y_k^*)}\right)$, then set $Y'_k = Y_k^*$; otherwise, set $Y'_k = \mathbf{a}$.

It is straightforwardly seen from **Case 2** that the algorithm attempts to escape from \mathbf{a} through rejection sampling with proposal distribution ϕ and acceptance probability $1 \wedge \frac{\pi}{c\phi}$. The density (with respect to ν) of the accepted candidate is therefore $\tilde{\pi}(x) := M([c\phi(x)] \wedge \pi(x))$, $x \in \mathbf{X}$, where $M := 1 / \int [c\phi(x)] \wedge \pi(x) \nu(dx)$. Note that $\tilde{\pi}(x) \leq M\pi(x)$ for every $x \in \mathbf{X}$, showing that (2.32) is satisfied. Moreover, it can be seen from **Case 1** that if the previous state is $y \in \mathbf{X}$, a candidate drawn from $P(y, \cdot)$ is accepted with probability $1 - 1 \wedge \frac{c\phi}{\pi} = 1 - \alpha$, where α is defined in (2.33). Pruning every occurrence of \mathbf{a} in $(Y'_k)_{k \in \mathbb{N}}$, we finally obtain the same transitions as in Algorithm 3, except that the general Markov kernel Q on Line 7 is replaced by the particular distribution $\tilde{\pi}$. Thus, the approach of [20] corresponds to a memoryless version of the GKKT process and falls into the class of regeneration-type algorithms. In contrast, our GKKT, which allows the flexibility of a general kernel Q targeting the distribution $\tilde{\pi}$, does not require exact sampling from $\tilde{\pi}$.

2.4.3 Revisiting the Metropolis–Hastings algorithm

Here, we use Remark 2.13 for casting the Metropolis–Hastings (MH) algorithm into the framework of the GKKT process. Consider the MH algorithm targeting some distribution π on $(\mathbf{X}, \mathcal{E})$ by means of some proposal kernel R on $\mathbf{X} \times \mathcal{E}$. We here suppose that π and R have a density and a transition density, denoted by π and r , respectively, with respect to some σ -finite measure $\nu \in \mathbf{M}(\mathcal{E})$. Then, the associated acceptance probability is given by

$$\alpha^{\text{MH}}(x, y) := \left(\frac{\pi(y)r(y, x)}{\pi(x)r(x, y)} \right) \wedge 1, \quad (x, y) \in \mathbf{X} \times \mathbf{X}. \quad (2.37)$$

The algorithm is initialized in $X_0 = x_0 \in \mathbf{X}$. In order to show that the MCMC chain generated by this algorithm is just a special instance of the GKKT process, define the probability measure

$$\tilde{\pi}(h) := \frac{\int \pi(dx) \alpha(x) h(x)}{\int \pi(dz) \alpha(z)}, \quad h \in \mathbf{F}_+(\mathcal{E}),$$

where $\alpha(x) := \int R(x, dy) \alpha^{\text{MH}}(x, y)$, $x \in \mathbf{X}$. Note that since $0 \leq \alpha \leq 1$, $\tilde{\pi}$ satisfies (2.32) with $M := 1 / \int \pi(dz) \alpha(z)$. Consider now a GKKT process with

- (i) the π -invariant Markov kernel P is given by $P(x, \cdot) = \delta_x$, $x \in \mathbf{X}$, and
- (ii) the $\tilde{\pi}$ -invariant Markov kernel Q_α is given by

$$Q_\alpha(x, \mathbf{A}) = \frac{\int_{\mathbf{A}} R(x, dy) \alpha^{\text{MH}}(x, y)}{\int R(x, dz) \alpha^{\text{MH}}(x, z)}, \quad (x, \mathbf{A}) \in \mathbf{X} \times \mathcal{E}.$$

With this particular choice of P and Q_α , Algorithm 3, with Lines 3–4 replaced by Lines 3'–4' according to Remark 2.13, turns into Algorithm 4 below.

Now, let us examine the dynamics of Algorithm 4 more closely. Assume that for some $k \in \mathbb{N}^*$, $B_k = 1$; then $Y_k = Z_k$, and we assign the value of Y_k to the next values $(Y_\ell)_{\ell > k}$ until the Bernoulli variable B_ℓ , with success probability $\alpha(Y_{\ell-1}) = \dots = \alpha(Y_k)$, takes on the value 1. In that case, Y_ℓ is drawn from the distribution $Q_\alpha(Z_k, \cdot) = Q_\alpha(Y_k, \cdot)$. Equivalently, we may draw a geometrically distributed random variable T with success probability $\alpha(Y_k)$ and, independently, a random variable $\tilde{Y} \sim Q_\alpha(Y_k, \cdot)$ and then set $Y_k = Y_{k+1} = \dots = Y_{k+T-1}$ and $Y_{k+T} = \tilde{Y}$. However, this is exactly the course of action of the MH algorithm, in which, starting from Y_k , a draw from $R(Y_k, \cdot)$ is accepted with a probability given by the function $\alpha^{\text{MH}}(Y_k, \cdot)$. If the move is not accepted, the value of Y_k is repeated. Note that this can be equivalently described as a rejection-sampling algorithm, where samples from $Q_\alpha(Y_k, \cdot)$ are generated by drawing candidates from $R(Y_k, \cdot)$ and accepting the same with probabilities given by the function $\alpha^{\text{MH}}(Y_k, \cdot)$. It is well known that for rejection sampling, the number T of trials before acceptance is geometrically distributed with success probability $\int R(Y_k, dy) \alpha^{\text{MH}}(Y_k, y) = \alpha(Y_k)$. Moreover, T is independent of the accepted variable \tilde{Y} , the latter having the desired distribution $Q_\alpha(Y_k, \cdot)$ conditionally on acceptance.

This finally shows that the MH algorithm is just a particular GKKT process, where the base kernel is degenerated, $P(x, \cdot) = \delta_x$, and the teleportation kernel Q_α is simulated using rejection sampling.

Algorithm 4 The MH algorithm as a GKKT process

```

1: Initialization Set  $(Y_0, Z_0) \leftarrow (x_0, x_0)$ .
2: for  $k \leftarrow 1$  to  $n$  do
3:   set  $Y_k^* = Y_{k-1}$  and, conditionally to  $Y_k^*$ , draw  $B_k \sim \text{Bernoulli}(\alpha(Y_k^*))$ 
4:   if  $B_k = 0$  then
5:     set  $(Y_k, Z_k) \leftarrow (Y_k^*, Z_{k-1})$ 
6:   else
7:     draw  $Z_k \sim Q_\alpha(Z_{k-1}, \cdot)$ 
8:     set  $Y_k \leftarrow Z_k$ 
9:   end if
10: end for

```

2.5 Numerical illustrations

In the following we benchmark numerically the KKT sampler against the *Metropolis-adjusted Langevin algorithm* (MALA) [121] and the HMC algorithm [12, 104]. In all our experiments, the target distribution π is assumed to have a positive density on \mathbb{R}^d with respect to the Lebesgue measure; for ease of notation, this density will be denoted by the same symbol, π . In addition, $\log \pi$ is supposed to be continuously differentiable, and we denote its gradient by $\nabla \log \pi$.

2.5. Numerical illustrations

We first recall the expression of the transition kernel R^{MH} of the MH algorithm based on some proposal transition density $r : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$ with respect to the Lebesgue measure: for every $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ (the Borel σ -field on \mathbb{R}^d),

$$R^{\text{MH}}(x, A) = \int_A \alpha^{\text{MH}}(x, y) r(x, y) dy + \delta_x(A) \int_{\mathbb{R}^d} \{1 - \alpha^{\text{MH}}(x, y)\} r(x, y) dy,$$

where the acceptance probability function α^{MH} is given in (2.37). We will consider two different parameterizations of the MH algorithm, namely the MALA and the *random-walk Metropolis (RWM) algorithm*. In RWM, proposals are generated by means of a random walk with standard deviation $\sigma > 0$, providing an MH algorithm with proposal transition density

$$r_\sigma(x, y) = (2\sigma^2\pi)^{-d/2} \exp\left(-\frac{1}{2\sigma^2} \|y - x\|^2\right).$$

On the other hand, in the MALA, proposals are generated according to the Euler–Maruyama discretization of the overdamped Langevin diffusion, *i.e.*,

$$r_\gamma(x, y) = (4\gamma\pi)^{-d/2} \exp\left(-\frac{1}{4\gamma} \|y - x - \gamma\nabla \log \pi(x)\|^2\right),$$

where $\gamma > 0$ is a fixed stepsize. The Markov kernels associated with the RWM algorithm and the MALA will be denoted by R_σ^{RWM} and R_γ^{MALA} , respectively.

Finally, we consider the HMC algorithm, which is briefly reviewed in the following (we refer to [17, 104] for details). HMC is based on the Hamiltonian function $H(x, v) := -\log \pi(x) + \|v\|^2/2$, $(x, v) \in X \times \mathbb{R}$, associated with the potential $-\log \pi$ of the target distribution and the corresponding ordinary differential equation (ODE) and dynamics. Indeed, the latter preserves the extended target distribution $\pi \otimes N(0, I_d)$, where $N(0, I_d)$ is the standard d -dimensional Gaussian distribution. However, in most cases, integrating exactly the Hamiltonian ODE is not an option, and therefore numerical integrators are typically used instead. A very popular choice is the *Verlet integrator* which, given some starting point $(x_0, v_0) \in \mathbb{R}^{2d}$, time step size $\Delta_t > 0$ and number $N_{\text{HMC}} \in \mathbb{N}^*$ of iterations, consists in the recursion $(x_{k+1}, v_{k+1}) = \Psi_{\Delta_t}(x_k, v_k)$, for $k \in \{0, \dots, N_{\text{HMC}} - 1\}$, where, letting $v_{k+1/2} = v_k + 2^{-1}\Delta_t \log \pi(x_k)$,

$$\Psi_{\Delta_t}(x_k, v_k) := (v_{k+1/2} + 2^{-1}\Delta_t \nabla \log \pi(x_k + \Delta_t v_{k+1/2}), x_k + \Delta_t v_{k+1/2}). \quad (2.38)$$

For $k \in \{1, \dots, N_{\text{HMC}} - 1\}$, $(x_k, v_k) = \Psi_{\Delta_t}^{\circ k}(x_0, v_0)$, where $\Psi_{\Delta_t}^{\circ 1} = \Psi_{\Delta_t}$ and $\Psi_{\Delta_t}^{\circ(k+1)} = \Psi_{\Delta_t}^{\circ k} \circ \Psi_{\Delta_t}$ for $k \in \mathbb{N}^*$. It follows that $\Psi_{\Delta_t}^{\circ k}(x_0, v_0)$ is an approximation of the Hamiltonian dynamics at time $k\Delta_t$. However, failing to leave the Hamiltonian function constant, numerical integration of the Hamiltonian ODE does not preserve the extended target in general. Still, like the continuous Hamiltonian dynamics, symplecticness and reversibility still hold. These properties allow $(x_{N_{\text{HMC}}}, v_{N_{\text{HMC}}}) = \Psi_{\Delta_t}^{\circ N_{\text{HMC}}}(x_0, v_0)$ to be used as a *deterministic* proposal inside an MH algorithm with corresponding acceptance probability function

$$\alpha_{\text{HMC}}(x_0, v_0) = 1 \wedge \exp\left(H(x_0, v_0) - H \circ \Psi_{\Delta_t}^{\circ N_{\text{HMC}}}(x_0, v_0)\right). \quad (2.39)$$

Finally, to ensure that the resulting Markov kernel is irreducible and ergodic (see [46]), the starting point of the additional variable v_0 is refreshed, *i.e.*, sampled from its stationary distribution $\mathcal{N}(0, \mathbf{I}_d)$ independently of the past, at each iteration. This yields the Markov kernel

$$R_{\Delta_t, N_{\text{HMC}}}^{\text{HMC}}(x, \mathbf{A}) := (2\pi)^{-2d} \int_{\mathbb{R}^d} \mathbb{1}_{\mathbf{A}} \left(\text{proj}_x \circ \Phi_{\Delta_t}^{\circ(N_{\text{HMC}})}(x, v) \right) \alpha_{\text{HMC}}(x, v) \exp(\|v\|^2) dv \\ + \delta_x(\mathbf{A}) (2\pi)^{-2d} \int_{\mathbb{R}^d} (1 - \alpha_{\text{HMC}}(x, v)) \exp(\|v\|^2) dv, \quad x \in \mathbb{R}^d, \mathbf{A} \in \mathcal{B}(\mathbb{R}^d),$$

where Ψ_{Δ_t} and α_{HMC} are defined in (2.38) and (2.39), respectively, and $\text{proj}_x : \mathbb{R}^{2d} \ni (x, v) \mapsto x$.

2.5.1 Two multi-modal distributions

In a first example, we consider the target distribution

$$\pi(x) = (4\pi)^{-1} \left(\exp\left(-\frac{1}{2}\|x - \boldsymbol{\mu}\|^2\right) + \exp\left(-\frac{1}{2}\|x + \boldsymbol{\mu}\|^2\right) \right), \quad x \in \mathbb{R}^2, \quad (2.40)$$

corresponding to a mixture of two bivariate Gaussian distributions with identity covariance matrix and means $\boldsymbol{\mu} = (10, 0)^\top$ and $-\boldsymbol{\mu}$. In this setting, we compare the MALA, with step size $\gamma = 0.1$, with the memoryless KKT sampler. The latter is parameterized by the same base kernel $P = R_\gamma^{\text{MALA}}$, again with $\gamma = 0.1$, and the set

$$\mathbf{C} = \{x \in \mathbf{D} : \pi(x) \leq cq(x)\},$$

where $c = 1.3/\pi$, $\mathbf{D} = [-15, 15]^2$ and $q(x) = \mathbb{1}_{\mathbf{D}}(x)/30^2$ for $x \in \mathbb{R}^2$. Moreover, we draw from $\pi_{\mathbf{C}}$ using rejection sampling with instrumental density q , *i.e.*, the uniform distribution on \mathbf{D} . With this choice of q , the KKT sampler will explore the modes of π using the MALA and teleport across the low-probability regions far from the modes through independent sampling from $\pi_{\mathbf{C}}$. In addition, by the definition of \mathbf{C} , the rejection-sampling algorithm performs relatively well; the average number of rejections equals 70. Note that the KKT sampler is supposed to operate using the MALA also outside \mathbf{D} ; however, due to the large span of \mathbf{D} , this never occurred in our simulation. Figure 2.1 provides the resulting histograms of 10^6 samples. As clear from this plot, the MALA gets stuck in one of the modes, and adjusting the step size of the proposal does not help circumventing this problem. On the other hand, the KKT process—moving easily across the low-probability barrier between the modes—explores efficiently the full distribution.

In a second experiment we consider again a two-dimensional multi-modal distribution, this time given by

$$\pi(x) = 14.8^{-1} \left[0.8 \left(\int \exp(-s^4) ds \right)^{-2} \exp\left(-(x_1 + 7)^4 - (x_2 + 6.5)^4\right) \right. \\ \left. + (2\pi)^{-1} \sum_{i=1}^{14} \exp\left(-\frac{1}{2}\|x - \boldsymbol{\mu}_i\|^2\right) \right], \quad x = (x_1, x_2) \in \mathbb{R}^2, \quad (2.41)$$

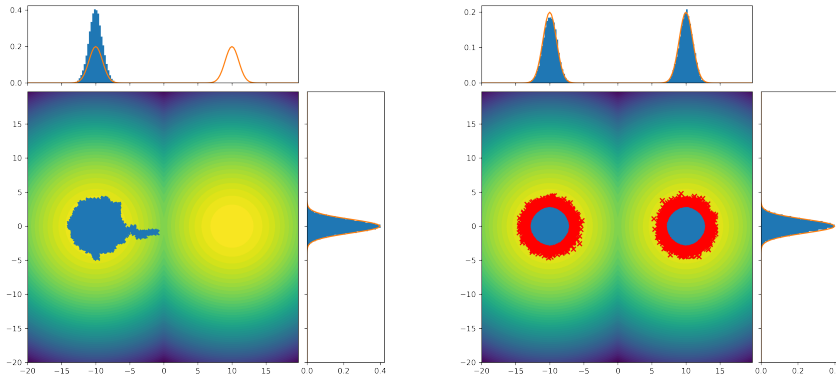


Figure 2.1: Comparison between MALA (left panel) and the memoryless KKT sampler (right panel) on the multi-modal distribution (2.40). For the memoryless KKT sampler, red and blue points correspond to samples generated by rejection sampling from π_C and MALA base kernel, respectively.

where $(\mu_i)_{i=1}^{14}$ are points in \mathbb{R}^2 , marked with black bullets in Figure 2.2. Note that one of the components of the mixture distribution (2.41) has lighter tails than the other ones, of which all are Gaussian.

For this model, we compared the MALA, with step size $\gamma = 0.8$, with the KKT sampler parameterized by the same base kernel $P = R_\gamma^{\text{MALA}}$ (again with $\gamma = 0.8$) and the set

$$C = \{x \in \mathbb{R}^2 : -\log(14.8\pi(x)) > 2\}.$$

In this case, we let the teleportative moves of the KKT process be governed by the RWM kernel $Q = R_\sigma^{\text{RWM}}$ with $\sigma = 0.8$. The algorithmic parameters γ , σ and C are tuned to obtain the best results for both methods. Figure 2.2 displays the resulting samples after 10^6 iterations. As evident from the plot, MALA fails to explore the mode of the target corresponding to the component of the mixture with the lightest tails, even though this mode is less isolated than in our previous example. The KKT sampler, on the other hand, has no difficulty at all in transiting across any low-probability barrier.

2.5.2 Stochastic volatility model

We now illustrate the proposed KKT sampler in the context of Bayesian inference in a stochastic volatility model. More precisely, following [88] we assume that we have access to a given record $(y_k)_{k=0}^N \in \mathbb{R}^{N+1}$, $N \in \mathbb{N}^*$, of observations such that for every $k \in \{0, \dots, N\}$, y_k is a draw from a one-dimensional, zero-mean Gaussian distribution with variance $\exp(2x_k)/\tau$, where the sequence $(x_k)_{k=0}^N$ is governed by the recursion

$$x_{k+1} = \rho x_k + z_{k+1},$$

with the model parameters $\rho \in (0, 1)$ and $\tau > 0$ and the noise sequence $(z_k)_{k=0}^N \in \mathbb{R}^{N+1}$ being unknown. In order to infer the parameters and the noise variables we

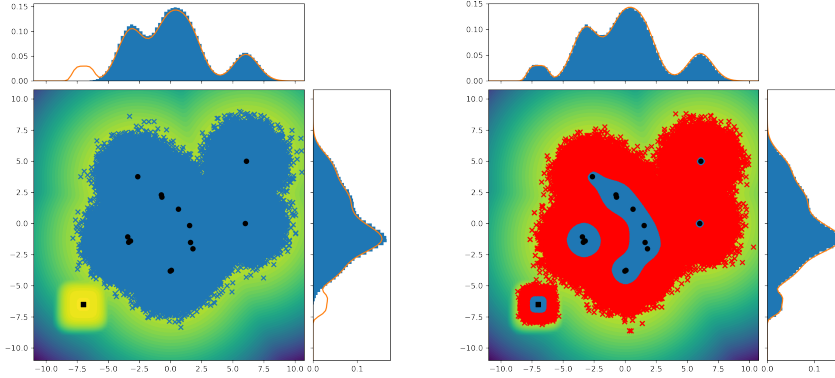


Figure 2.2: Comparison between MALA (left panel) and the KKT sampler (right panel) on the multi-modal distribution (2.41). For the KKT sampler, red and blue points correspond to samples generated by the RWM teleportation kernel and MALA base kernel, respectively.

take a Bayesian approach and assign τ a gamma prior with hyperparameter $(21, 5)$ and $(1 + \rho)/2$ a beta prior with hyperparameter $(20, 2)$. Moreover, the noise variables $(z_k)_{k=0}^N$ are assumed to be *a priori* independent and standard normally distributed and $x_0 = z_0/\sqrt{1 - \rho^2}$. In our experiment, $N + 1 = 100$ synthetic data points $(y_k)_{k=0}^N$ were generated by simulation under model parameters and noise variables drawn from the corresponding priors. In addition, the reparameterizations

$$\alpha = -\log(\tau)/2, \quad \beta = (\log(1 + \rho) - \log(1 - \rho))/2.$$

of τ and ρ provide parameters that are well supported on \mathbb{R} . We then obtain a target density on \mathbb{R}^{N+3} given by

$$\begin{aligned} -\log \pi(\alpha, \beta, (z_k)_{k=0}^N \mid (y_k)_{k=0}^N) &= 42\alpha + 5e^{-2\alpha} + 22 \log(1 + e^{-2\beta}) + 4\beta + (N + 1)\alpha \\ &\quad + 2^{-1} \sum_{k=0}^N x_k + 2^{-1} \sum_{k=0}^N (w_k + z_k^2) + \text{Cst}_{\text{SV}}, \end{aligned}$$

where $w_k := \exp(-x_k - 2\alpha)y_k^2$ and Cst_{SV} is some additive constant independent of the parameters of interest; see [88, Section G.3] for details. On this model, we compare the KKT sampler to HMC. For HMC, we consider $N_{\text{HMC}} = 35$ iterations for the Verlet integrator and adjust the time step Δ_t to obtain an acceptance rate of about 0.7 on the average. The very same HMC kernel is used as base kernel P in the KKT sampler. As teleportation kernel we use $Q = R_{\sigma}^{\text{RWM}}$, with σ selected to obtain an acceptance rate of about 0.25. Finally, the region \mathcal{C} is given by

$$\mathcal{C} = \left\{ (\alpha, \beta, (z_k)_{k=0}^N) \in \mathbb{R}^{N+3} : -\log \pi(\alpha, \beta, (z_k)_{k=0}^N \mid (y_k)_{k=0}^N) - \text{Cst}_{\text{SV}} > 75 \right\}.$$

In this setting, we run both MCMC algorithms for 10^5 iterations after discarding a burn-in period comprising 10^5 steps. For the KKT sampler, the percentage of samples

2.5. Numerical illustrations

Algorithm	ESS(α)	ESS(β)	$(\text{ESS}(z_k))_{k=0}^N$			
			mean	variance	min	max
KKT sampler	4.84	1.57	0.86	0.10	0.33	1.89
HMC	0.23	0.17	0.22	0.004	0.15	0.46

Table 2.1: Summary statistics of ESSs calculated on the basis of MCMC paths generated by the MALA and the KKT sampler for the stochastic volatility model in Section 2.5.2. The values are based on 10^5 iterations of each algorithm.

generated by the teleportation kernel is 63%. Table 2.1 reports the resulting *effective sample sizes* (ESSs) (see [119, Section 12.3.5]) per evaluation of $\log \pi$ and $\nabla \log \pi$ at each iteration for each component of the resulting MCMC trajectories. As clear from Table 2.1, the KKT sampler outperforms significantly HMC in terms of this measure.

2.5.3 The Ginzburg–Landau model

In our final numerical experiment, we consider the *Ginzburg–Landau model* used for describing phase transitions in condensed matter physics; see [87, Section 6.2]. Let $p \in \mathbb{N}$ and $d = p^3$ and define π on a three-dimensional lattice by, for $x = (x_{ijk})_{(i,j,k) \in \{1, \dots, p\}^3} \in \mathbb{R}^d$,

$$-\log \pi(x) = 2^{-1} \sum_{i,j,k=1}^p \left((1 - \tau)x_{ijk}^2 + \tau\alpha \|\tilde{\nabla} x_{ijk}\|^2 + \tau\lambda x_{ijk}^4/2 \right) + \text{Cst}_{\text{GL}},$$

where Cst_{GL} is the logarithm of the normalizing constant and τ , λ and α are all positive model parameters. In addition, $\tilde{\nabla} x_{ijk} := (x_{i_+jk} - x_{ijk}, x_{ijj_+k} - x_{ijk}, x_{ijk_+} - x_{ijk})$, where $i_+ := i + 1 \pmod p$ (and similarly for j_+ and k_+). In our simulations, the target is parameterized by $p = 5$, $\tau = 2$, $\lambda = 0.5$ and $\alpha = 0.1$. In this example, we compare the MALA, operating with the step size $\gamma = 10^{-3}$, with the KKT sampler using the base kernel $P = R_\gamma^{\text{MALA}}$ with $\gamma = 0.1$, the set

$$\mathbf{C} = \left\{ x \in \mathbb{R}^d : -\log(\pi(x)) - \text{Cst}_{\text{GL}} > 100 \right\} \quad (2.42)$$

and the teleportation kernel $Q = R_\sigma^{\text{RWM}}$ with $\sigma = 0.1$. Note that the step size γ used in the MALA is set to be quite small in comparison to that used in the KKT sampler; otherwise, the acceptance probability would degenerate to zero, resulting in a stuck algorithm. This is mainly due to the fact that $\nabla \log \pi$ is not Lipschitz in this case. In contrast, the teleportation algorithm does not suffer from this issue if the set \mathbf{C} is tuned appropriately. Indeed, with the choice (2.42) of \mathbf{C} , the MALA is used only in the neighborhoods of the modes of $\log \pi$, where $\nabla \log \pi$ is Lipschitz with a relatively small Lipschitz constant, while instead RWM—incorporating no gradient information—is used outside these regions.

In Table 2.2, we display the ESSs of the trajectories produced by the two methods, divided by the number of evaluations of either $\log \pi$ or $\nabla \log \pi$ per iteration. The values

Algorithm	mean	variance	min	max
KKT sampler	908	5438	727	1091
MALA	34	83	12	57

Table 2.2: Summary statistics of ESSs calculated on the basis of MCMC paths generated by the MALA and the KKT sampler for the Ginzburg–Landau model in Section 2.5.3. The values are based on 10^5 iterations of each algorithm.

are calculated on the basis of 10^5 iterations (after discarding burn-in periods comprising 10^5 iterations). Evidently, the KKT sampler outperforms clearly MALA, with an improvement of at least an order of magnitude in terms of average ESS.

2.6 Kernel notation

The following kernel notation will be used at several places in the paper. Let (X_1, \mathcal{E}_1) and (X_2, \mathcal{E}_2) be general measurable spaces. A possibly unnormalized transition kernel K on $X_1 \times \mathcal{E}_2$ induces the following three operations, one on $F_b(\mathcal{E}_2)$ and two on $M(\mathcal{E}_1)$:

- For $h \in F_b(\mathcal{E}_2)$ we define the function $Kh : X_1 \ni x \mapsto \int h(y) K(x, dy)$ (whenever the integral is well defined).
- For $\mu \in M(\mathcal{E}_1)$ we define the measure $\mu K : \mathcal{E}_2 \ni A \mapsto \int \mu(dx) K(x, A)$.
- For $\mu \in M(\mathcal{E}_1)$ we define the measure $\mu \otimes K : \mathcal{E}_1 \otimes \mathcal{E}_2 \ni A \mapsto \int \int_A \mu(dx) K(x, dy)$.

Now let (X_3, \mathcal{E}_3) be a third measurable space and L a transition kernel on $X_2 \times \mathcal{E}_3$. Then we define the following products between K and L , the first resulting in a kernel on $X_1 \times \mathcal{E}_3$ and the second in a kernel on $X_1 \times (\mathcal{E}_2 \otimes \mathcal{E}_3)$:

- $KL : X_1 \times \mathcal{E}_3 \ni (x, A) \mapsto \int K(x, dy) L(y, A)$.
- $K \otimes L : X_1 \times (\mathcal{E}_2 \otimes \mathcal{E}_3) \ni (x, A) \mapsto \int \int_A K(x, dy) L(y, dz)$.

2.7 Proofs and technical lemmas

2.7.1 Proof of Proposition 2.4

First, note that for all A_0 and A_1 in \mathcal{E} , by the definition of S ,

$$\begin{aligned}
 & [\pi \otimes S](A_0 \times A_1) \\
 &= \int_{A_0} \pi(dy) P(y, C^c \cap A_1) + \int_{A_0} \pi(dy) P(y, C) \pi_C(A_1) \\
 &= [\pi \otimes P]((C^c \cap A_0) \times (C^c \cap A_1)) + [\pi \otimes P]((C \cap A_0) \times (C^c \cap A_1)) \\
 &\quad + [\pi \otimes P]((C^c \cap A_0) \times C) \pi_C(A_1) + [\pi \otimes P]((C \cap A_0) \times C) \pi_C(A_1). \quad (2.43)
 \end{aligned}$$

2.7. Proofs and technical lemmas

We first prove that (a) and (b) jointly imply that S is π -reversible. To do so, we rewrite each term on the right-hand side of (2.43) to show that $[\pi \otimes S](A_0 \times A_1) = [\pi \otimes S](A_1 \times A_0)$. We start with the second and the last terms. By (b),

$$[\pi \otimes P]((C \cap A_0) \times (C^c \cap A_1)) = \pi_C(A_0)\pi(C)\mu(C^c \cap A_1),$$

and

$$[\pi \otimes P]((C \cap A_0) \times C) = \int_{C \cap A_0} \pi(dy) (1 - P(y, C^c)) = (1 - \mu(C^c))\pi(C)\pi_C(A_0).$$

We turn to the third term on the right-hand side of (2.43). Under (A1), (a) and (b) imply

$$\begin{aligned} [\pi \otimes P]((C^c \cap A_0) \times C) &= \pi(C^c \cap A_0) - [\pi \otimes P]((C^c \cap A_0) \times C^c) \\ &= [\pi \otimes P](X \times (C^c \cap A_0)) - [\pi \otimes P](C^c \times (C^c \cap A_0)) \\ &= [\pi \otimes P](C \times (C^c \cap A_0)) \\ &= \pi(C)\mu(C^c \cap A_0). \end{aligned} \tag{2.44}$$

Combining (2.43)–(2.44) yields

$$\begin{aligned} [\pi \otimes S](A_0 \times A_1) &= [\pi \otimes P]((C^c \cap A_0) \times (C^c \cap A_1)) + \pi_C(A_0)\pi(C)\mu(C^c \cap A_1) \\ &\quad + \pi_C(A_1)\pi(C)\mu(C^c \cap A_0) + (1 - \mu(C^c))\pi(C)\pi_C(A_0)\pi_C(A_1), \end{aligned}$$

and by (a) this expression is symmetric in A_0 and A_1 . Thus, $[\pi \otimes S](A_0 \times A_1) = [\pi \otimes S](A_1 \times A_0)$, which implies that S is π -reversible.

We now establish the converse. If S is π -reversible, then (a) holds trivially true (as $\mathcal{E}_{C^c}^{\otimes 2} \subset \mathcal{E}^{\otimes 2}$). We show that (b) holds as well. By (2.43), and using that S is π -reversible, we have for every $(A, B) \in \mathcal{E}_C \times \mathcal{E}_{C^c}$,

$$\int_A \pi(dy) P(y, B) = [\pi \otimes S](A \times B) = [\pi \otimes S](B \times A) = \pi_C(A) \int_B \pi(dy) P(y, C).$$

Therefore,

$$\int_A \pi(dy) (P(y, B) - \mu(B)) = 0, \tag{2.45}$$

where we have defined the measure $\mu : \mathcal{E}_{C^c} \ni B' \mapsto \int_B \pi(dy) P(y, C)/\pi(C)$. Since \mathcal{E}_{C^c} is countably generated there exists a π -system $(B_k)_{k \in \mathbb{N}}$, with $B_k \in \mathcal{E}_{C^c}$ for all k , such that $\mathcal{E}_{C^c} = \sigma((B_k)_{k \in \mathbb{N}})$ and $B_0 = C^c$. Then, by using, for every $k \in \mathbb{N}$, (2.45) with $B = B_k$ and $A = \{P(\cdot, B_k) > \mu(B_k)\}$ as well as $A = \{P(\cdot, B_k) < \mu(B_k)\}$, we conclude that there exists $X_k \in \mathcal{E}_C$ such that $\pi(X_k^c) = 0$ and for every $y \in X_k$, $P(y, B_k) = \mu(B_k)$. Now, let $X := \bigcap_{k \in \mathbb{N}} X_k$; then $\pi(X^c) = 0$ and for every $y \in X$ and $k \in \mathbb{N}$, $P(y, B_k) = \mu(B_k)$. Therefore, by Dynkin's π - λ theorem, $P(y, \cdot)|_{C^c} = \mu$ for π -almost all $y \in C$, which completes the proof.

2.7.2 Proof of Proposition 2.8

First, assume that (ii) holds. Then by [37, Theorem 5.1.8], for every measurable function $g : \mathsf{X} \rightarrow \mathbb{R}$ such that $\pi(|g|) < +\infty$,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} g(X_k) = \mathbb{E}_\pi^P[g(X_0) \mid \mathcal{I}], \quad \mathbb{P}_\pi^P\text{-a.s.}, \quad (2.46)$$

where $\mathcal{I} := \left\{ \mathbf{A} \in \mathcal{E}^{\otimes \mathbb{N}} : \mathbb{1}_{\mathbf{A}} = \mathbb{1}_{\mathbf{A}} \circ \theta \right\}$ is the σ -field of invariant sets. Under (ii), [37, Corollary 5.2.4] implies that the invariant random variable $\mathbb{E}_\pi^P[g(X_0) \mid \mathcal{I}]$ is \mathbb{P}_π^P -a.s. constant. As a consequence, \mathbb{P}_π^P -a.s.,

$$\mathbb{E}_\pi^P[g(X_0) \mid \mathcal{I}] = \mathbb{E}_\pi^P \left[\mathbb{E}_\pi^P[g(X_0) \mid \mathcal{I}] \right] = \mathbb{E}_\pi^P[g(X_0)] = \pi(g).$$

Plugging this into (2.46) yields $\mathbb{P}_\pi^P(\mathbf{A}) = 1$, where

$$\mathbf{A} := \left\{ \lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} g(X_k) = \pi(g) \right\}.$$

Since \mathbf{A} is invariant, *i.e.*, $\mathbb{1}_{\mathbf{A}} = \mathbb{1}_{\mathbf{A}} \circ \theta$, the function $h : \mathsf{X} \ni x \mapsto \mathbb{P}_x^P(\mathbf{A})$ is bounded and harmonic, and by (ii) it is hence equal to a constant ζ . Then, $1 = \mathbb{P}_\pi^P(\mathbf{A}) = \int \pi(dx) \mathbb{P}_x^P(\mathbf{A}) = \zeta$ and we conclude that $\mathbb{P}_\xi^P(\mathbf{A}) = \int \xi(dx) \mathbb{P}_x^P(\mathbf{A}) = \zeta = 1$. Thus, (i) holds true.

Conversely, assume (i) and let h be a bounded and harmonic function. Applying (i) with $g = h$ yields that for every $x \in \mathsf{X}$,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=0}^{n-1} h(X_k) = \pi(h), \quad \mathbb{P}_x^P\text{-a.s.}$$

In addition, since by [37, Proposition 5.2.2(ii)], for any $x \in \mathsf{E}$, $(h(X_k))_{k \in \mathbb{N}}$ converges \mathbb{P}_x^P -a.s. to some random variable $Y = \limsup_{k \rightarrow +\infty} h(X_k)$ such that $h(x) = \mathbb{E}_x^P[Y]$, we finally conclude that $Y = \pi(h)$, \mathbb{P}_x^P -a.s., and $h(x) = \mathbb{E}_x^P[Y] = \pi(h)$ for any $x \in \mathsf{X}$. This shows (ii).

2.7.3 Technical results for proving Theorem 2.7 and Theorem 2.9

Lemma 3. *Set $\bar{\mathsf{C}} = \mathsf{C} \times \mathsf{C}$. Then for every $(y, z) \in \mathsf{X} \times \mathsf{C}$, $n \in \mathbb{N}$ and $v \in \mathsf{F}_+(\mathcal{E})$,*

$$\mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathsf{C}}} > n\}} v(Y_n)] = \mathbb{E}_y^P[\mathbb{1}_{\{\sigma_{\mathsf{C}} > n\}} v(X_n)]. \quad (2.47)$$

If, in addition, $\inf_{y \in \mathsf{E}} \mathbb{P}_y^P(\sigma_{\mathsf{C}} < +\infty) = 1$, then the following properties hold true.

- (i) *For every $(y, z) \in \mathsf{X} \times \mathsf{C}$ and $\ell \in \mathbb{N}^*$, $\mathbb{P}_{(y,z)}^R(\sigma_{\bar{\mathsf{C}}}^\ell < +\infty) = 1$.*

2.7. Proofs and technical lemmas

(ii) For every $(g, h) \in F_+(\mathbb{N}) \times F_+(\mathcal{E}_C)$ and $\mu \in M_1(\mathcal{E} \otimes \mathcal{E}_C)$,

$$\mathbb{E}_\mu^R[g(\sigma_{\bar{c}})h(Z_{\sigma_{\bar{c}}})] = \int \mu(d(y, z)) \mathbb{E}_y^P[g(\sigma_C)]Qh(z) \quad (2.48)$$

and for every $\ell \in \mathbb{N}^*$,

$$\mathbb{E}_\mu^R \left[g(\sigma_{\bar{c}} \circ \theta_{\sigma_{\bar{c}}^{\ell-1}})h(Z_{\sigma_{\bar{c}}^\ell}) \mid \mathcal{F}_{\sigma_{\bar{c}}^{\ell-1}} \right] = \mathbb{E}_{Y_{\sigma_{\bar{c}}^{\ell-1}}}^P[g(\sigma_C)]Qh(Z_{\sigma_{\bar{c}}^{\ell-1}}), \quad \mathbb{P}_\mu^R\text{-a.s.} \quad (2.49)$$

(iii) For all $k \in \mathbb{N}^*$, $(g_i)_{i=1}^k$ in $F_+(\mathbb{N})$, $(h_i)_{i=1}^k$ in $F_+(\mathcal{E}_C)$ and $\mu \in M_1(\mathcal{E} \otimes \mathcal{E}_C)$,

$$\begin{aligned} & \mathbb{E}_\mu^R \left[\prod_{\ell=1}^k g_\ell(\sigma_{\bar{c}}^\ell - \sigma_{\bar{c}}^{\ell-1})h_\ell(Z_{\sigma_{\bar{c}}^\ell}) \right] \\ &= \int \mu(d(y, z)) \mathbb{E}_y^P[g_1(\sigma_C)]\mathbb{E}_z^Q \left[h_k(X_k) \prod_{\ell=1}^{k-1} h_\ell(X_\ell) \mathbb{E}_{X_\ell}^P[g_{\ell+1}(\sigma_C)] \right], \quad (2.50) \end{aligned}$$

where, by convention, $\sigma_{\bar{c}}^0 := 0$.

(iv) Let $(y, z) \in X \times C$; then under $\mathbb{P}_{(y,z)}^R$, $(Y_{\sigma_{\bar{c}}^k})_{k \in \mathbb{N}^*}$ is a Markov chain with transition kernel Q and initial distribution δ_z .

Proof. We establish (2.47) by induction over n . The base case $n = 0$ holds trivially true. Assuming that (2.47) holds true for some $n \in \mathbb{N}$, the Markov property implies that for every $(y, z) \in E \times C$,

$$\begin{aligned} \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{c}} > n+1\}}v(Y_{n+1})] &= \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{c}} > n\}}\mathbb{1}_{\{Y_{n+1} \notin C\}}v(Y_{n+1})] \\ &= \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{c}} > n\}}P(\mathbb{1}_{C^c}v)(Y_n)] \\ &= \mathbb{E}_y^P[\mathbb{1}_{\{\sigma_C > n\}}P(\mathbb{1}_{C^c}v)(X_n)] \\ &= \mathbb{E}_y^P[\mathbb{1}_{\{\sigma_C > n+1\}}v(X_{n+1})], \end{aligned}$$

which establishes (2.47) with n replaced by $n + 1$ and hence completes the induction step.

We now establish (i)–(iv) in turn.

(i) By setting $v \equiv 1$ in (2.47) and letting $n \rightarrow \infty$, we get that for every $(y, z) \in X \times C$, $\mathbb{P}_{(y,z)}^R(\sigma_{\bar{c}} = +\infty) = \mathbb{P}_y^P(\sigma_C = +\infty)$. Therefore, by assumption, for every $(y, z) \in E \times C$,

$$\mathbb{P}_{(y,z)}^R(\sigma_{\bar{c}} < +\infty) = \mathbb{P}_y^P(\sigma_C < +\infty) = 1,$$

from which it follows that $\mathbb{P}_{(y,z)}^R(\sigma_{\bar{c}}^\ell < +\infty) = 1$ for every $\ell \in \mathbb{N}^*$.

- (ii) We first show (2.48). Using the Markov property and the definition (2.7) of R , for every $(y, z) \in \mathbf{X} \times \mathbf{C}$,

$$\begin{aligned} \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathbf{C}}}=n\}}h(Z_n)] &= \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathbf{C}}}>n-1\}}\mathbb{1}_{\{Y_n \in \mathbf{C}\}}h(Z_n)] \\ &= \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathbf{C}}}>n-1\}}P(Y_{n-1}, \mathbf{C})Qh(Z_{n-1})] \\ &= \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathbf{C}}}>n-1\}}P(Y_{n-1}, \mathbf{C})]Qh(z) \\ &= \mathbb{E}_y^P[\mathbb{1}_{\{\sigma_{\mathbf{C}}>n-1\}}P(X_{n-1}, \mathbf{C})]Qh(z) \\ &= \mathbb{P}_y^P(\sigma_{\mathbf{C}} = n)Qh(z), \end{aligned}$$

where the penultimate equality follows from (2.47). Then, using (i) and the previous identity,

$$\begin{aligned} \mathbb{E}_{(y,z)}^R[g(\sigma_{\bar{\mathbf{C}}})h(Z_{\sigma_{\bar{\mathbf{C}}}})] &= \sum_{n=1}^+ \infty g(n) \mathbb{E}_{(y,z)}^R[\mathbb{1}_{\{\sigma_{\bar{\mathbf{C}}}=n\}}h(Z_n)] \\ &= \sum_{n=1}^+ \infty g(n) \mathbb{P}_y^P(\sigma_{\mathbf{C}} = n)Qh(z) \\ &= \mathbb{E}_y^P[g(\sigma_{\mathbf{C}})]Qh(z). \end{aligned}$$

Integrating with respect to μ establishes (2.48).

We turn to (2.49). Since $Z_{\sigma_{\bar{\mathbf{C}}}^\ell} = Z_{\sigma_{\bar{\mathbf{C}}}} \circ \theta_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}}$ on $\{\sigma_{\bar{\mathbf{C}}}^{\ell-1} < +\infty\}$, combining the strong Markov property, (i) and (2.48) yields, \mathbb{P}_μ^R -a.s.,

$$\begin{aligned} \mathbb{E}_\mu^R \left[g(\sigma_{\bar{\mathbf{C}}} \circ \theta_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}})h(Z_{\sigma_{\bar{\mathbf{C}}}^\ell}) \mid \mathcal{F}_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}} \right] &= \mathbb{E}_\mu^R \left[g(\sigma_{\bar{\mathbf{C}}} \circ \theta_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}})h(Z_{\sigma_{\bar{\mathbf{C}}} \circ \theta_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}}}) \mid \mathcal{F}_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}} \right] \\ &= \mathbb{E}_{(Y_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}}, Z_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}})}^R [g(\sigma_{\bar{\mathbf{C}}})h(Z_{\sigma_{\bar{\mathbf{C}}}})] \\ &= \mathbb{E}_{Y_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}}}^P [g(\sigma_{\mathbf{C}})]Qh(Z_{\sigma_{\bar{\mathbf{C}}}^{\ell-1}}), \end{aligned}$$

where the last equality follows from (2.48). This completes the proof (2.49).

- (iii) We proceed by induction. The base case $k = 1$ holds by (2.48). In order to carry through the induction step, assume that (2.50) holds for some $k \in \mathbb{N}^*$. Then, for all $(g_i)_{i=1}^{k+1}$ in $F_+(\mathbb{N})$, $(h_i)_{i=1}^{k+1}$ in $F_+(\mathcal{E}_{\mathbf{C}})$ and $\mu \in \mathbf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathbf{C}})$, using the tower property, (2.49) and the fact that $Y_{\sigma_{\bar{\mathbf{C}}}^\ell} = Z_{\sigma_{\bar{\mathbf{C}}}^\ell}$, \mathbb{P}_μ^R -a.s., for all $\ell \in \mathbb{N}^*$, we obtain

$$\begin{aligned} &\mathbb{E}_\mu^R \left[\prod_{\ell=1}^{k+1} g_\ell(\sigma_{\bar{\mathbf{C}}}^\ell - \sigma_{\bar{\mathbf{C}}}^{\ell-1})h_\ell(Z_{\sigma_{\bar{\mathbf{C}}}^\ell}) \right] \\ &= \mathbb{E}_\mu^R \left[\mathbb{E}_\mu^R [g_{k+1}(\sigma_{\bar{\mathbf{C}}} \circ \theta_{\sigma_{\bar{\mathbf{C}}}^k})h_{k+1}(Z_{\sigma_{\bar{\mathbf{C}}}^{k+1}}) \mid \mathcal{F}_{\sigma_{\bar{\mathbf{C}}}^k}] \prod_{\ell=1}^k g_\ell(\sigma_{\bar{\mathbf{C}}}^\ell - \sigma_{\bar{\mathbf{C}}}^{\ell-1})h_\ell(Z_{\sigma_{\bar{\mathbf{C}}}^\ell}) \right] \\ &= \mathbb{E}_\mu^R \left[\mathbb{E}_{Z_{\sigma_{\bar{\mathbf{C}}}^k}}^P [g_{k+1}(\sigma_{\mathbf{C}})]Qh_{k+1}(Z_{\sigma_{\bar{\mathbf{C}}}^k}) \prod_{\ell=1}^k g_\ell(\sigma_{\bar{\mathbf{C}}}^\ell - \sigma_{\bar{\mathbf{C}}}^{\ell-1})h_\ell(Z_{\sigma_{\bar{\mathbf{C}}}^\ell}) \right]. \end{aligned}$$

2.7. Proofs and technical lemmas

The expectation on the right-hand side is over a product of nonnegative functions of $Z_{\sigma_{\mathbb{C}}^{\ell}}$ and $\sigma_{\mathbb{C}}^{\ell} - \sigma_{\mathbb{C}}^{\ell-1}$ for $\ell \in \{1, \dots, k\}$. Therefore, using the induction hypothesis,

$$\begin{aligned} & \mathbb{E}_{\mu}^R \left[\prod_{\ell=1}^{k+1} g_{\ell}(\sigma_{\mathbb{C}}^{\ell} - \sigma_{\mathbb{C}}^{\ell-1}) h_{\ell}(Z_{\sigma_{\mathbb{C}}^{\ell}}) \right] \\ &= \int \mu(d(y, z)) \mathbb{E}_y^P [g_1(\sigma_{\mathbb{C}})] \\ & \quad \times \mathbb{E}_z^Q \left[\mathbb{E}_{X_k}^P [g_{k+1}(\sigma_{\mathbb{C}})] Q h_{k+1}(X_k) h_k(X_k) \prod_{\ell=1}^{k-1} h_{\ell}(X_{\ell}) \mathbb{E}_{X_{\ell}}^P [g_{\ell+1}(\sigma_{\mathbb{C}})] \right] \\ &= \int \mu(d(y, z)) \mathbb{E}_y^P [g_1(\sigma_{\mathbb{C}})] \mathbb{E}_z^Q \left[h_{k+1}(X_{k+1}) \prod_{\ell=1}^k h_{\ell}(X_{\ell}) \mathbb{E}_{X_{\ell}}^P [g_{\ell+1}(\sigma_{\mathbb{C}})] \right], \end{aligned}$$

which means that (2.50) holds true with k replaced by $k + 1$. Thus, the claim follows by induction.

- (iv) By applying (2.50) with $g_1 \equiv \dots \equiv g_k \equiv \mathbf{1}$, we conclude that for all $k \in \mathbb{N}^*$, $(h_i)_{i=1}^{k+1}$ in $\mathbf{F}_+(\mathcal{E}_{\mathbb{C}})$ and $\mu \in \mathbf{M}_1(\mathcal{E} \otimes \mathcal{E}_{\mathbb{C}})$,

$$\mathbb{E}_{\mu}^R \left[\prod_{\ell=1}^k h_{\ell}(Z_{\sigma_{\mathbb{C}}^{\ell}}) \right] = \int \mu(d(y, z)) \mathbb{E}_z^Q \left[\prod_{\ell=1}^k h_{\ell}(X_{\ell}) \right].$$

This shows that for every $(y, z) \in \mathbf{E} \times \mathbf{C}$, under $\mathbb{P}_{(y,z)}^R$, $(Z_{\sigma_{\mathbb{C}}^{\ell}})_{\ell \in \mathbb{N}^*}$ is a Markov chain with transition kernel Q and initial distribution $\delta_{(y,z)}$. We may now complete the proof by noting that $Y_{\sigma_{\mathbb{C}}^{\ell}} = Z_{\sigma_{\mathbb{C}}^{\ell}}$, \mathbb{P}_{μ}^R -a.s., for all $\ell \in \mathbb{N}^*$.

□

Lemma 4. Assume (A4)(i)–(ii) and that for every $y \in \mathbf{X}$, $\mathbb{P}_y^P(\sigma_{\mathbb{C}} < +\infty) = 1$. Set $\delta := \inf_{z \in \mathbf{D}} P(z, \mathbf{C})$ and $\bar{\mathbf{D}} := \mathbf{D} \times \mathbf{D}$ and define the probability measure $\bar{\nu} : \mathcal{E} \otimes \mathcal{E}_{\mathbb{C}} \ni \mathbf{A} \mapsto \int_{\mathbf{C}} \nu(dz) \mathbb{1}_{\mathbf{A}}(z, z)$. Then the set $\bar{\mathbf{D}} = \mathbf{D} \times \mathbf{D}$ is an accessible $(1, \delta\epsilon\bar{\nu})$ -small set for R such that $\bar{\nu}(\bar{\mathbf{D}}) > 0$.

Proof. Under (A4)(i)–(ii) it holds that for every $(y, z) \in \bar{\mathbf{D}}$ and $\mathbf{A} \in \mathcal{E} \otimes \mathcal{E}_{\mathbb{C}}$,

$$R((y, z), \mathbf{A}) \geq P(y, \mathbf{C}) \int Q(z, dz') \mathbb{1}_{\mathbf{A}}(z', z') \geq \delta\epsilon\bar{\nu}(\mathbf{A}),$$

showing that $\bar{\mathbf{D}}$ is a $(1, \delta\epsilon\bar{\nu})$ -small set for R . In addition, $\bar{\nu}(\bar{\mathbf{D}}) = \nu(\mathbf{D}) > 0$. To complete the proof, it remains to show that $\bar{\mathbf{D}}$ is accessible for R . By assumption, Q admits an accessible small set and is hence irreducible. Furthermore, [37, Theorem 9.2.4] implies that there exists a maximal irreducibility measure $\psi \in \mathbf{M}(\mathcal{E}_{\mathbb{C}})$ for Q . We define the measure $\bar{\psi} : \mathcal{E} \otimes \mathcal{E}_{\mathbb{C}} \ni \mathbf{A} \mapsto \int_{\mathbf{C}} \psi(dz) \mathbb{1}_{\mathbf{A}}(z, z)$ and complete the proof by showing that R is $\bar{\psi}$ -irreducible. Indeed, since ν is, by (A4)(i), an irreducibility measure for Q such that

$\nu(\mathbf{D}) > 0$, it holds that $\bar{\psi}(\bar{\mathbf{D}}) = \psi(\mathbf{D}) > 0$; thus, by [37, Lemma 3.5.2], $\bar{\mathbf{D}}$ is accessible for R if R is $\bar{\psi}$ -irreducible.

To establish $\bar{\psi}$ -irreducibility, consider a set $\mathbf{A} \in \mathcal{E} \otimes \mathcal{E}_{\mathbf{C}}$ such that $\bar{\psi}(\mathbf{A}) > 0$. Setting $\mathbf{A}_1 := \{z \in \mathbf{C} : (z, z) \in \mathbf{A}\}$, it holds that $\bar{\psi}(\mathbf{A}) = \int \mathbb{1}_{\mathbf{A}}(z, z) \psi(dz) = \psi(\mathbf{A}_1) > 0$, and since Q is ψ -irreducible, it follows that $\sum_{k=1}^{+\infty} Q^k(z, \mathbf{A}_1) > 0$ for all $z \in \mathbf{C}$. Now for all $(y, z) \in \mathbf{X} \times \mathbf{C}$, $Y_k = Z_k$, $\mathbb{P}_{(y,z)}^R$ -a.s. on $\{Y_k \in \mathbf{C}\}$, and therefore

$$\begin{aligned} \sum_{n=1}^{+\infty} R^n((y, z), \mathbf{A}) &\geq \sum_{n=1}^{+\infty} \mathbb{P}_{(y,z)}^R \left((Y_n, Z_n) \in \mathbf{A} \cap \bar{\mathbf{C}}, Y_n = Z_n \right) \\ &= \sum_{\ell=1}^{+\infty} \mathbb{P}_{(y,z)}^R \left(Z_{\sigma_{\bar{\mathbf{C}}}^{\ell}} \in \mathbf{A}_1 \right) = \sum_{\ell=1}^{+\infty} Q^{\ell}(z, \mathbf{A}_1) > 0, \end{aligned}$$

where the first and last equalities follow from Lemma 3(i) and Lemma 3(iii), respectively. The Markov kernel R is therefore $\bar{\psi}$ -irreducible, and the proof is finalized. \square

CHAPTER 3

Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics

ALAIN DURMUS¹, AURÉLIEN ENFROY², ÉRIC MOULINES³ AND GABRIEL STOLTZ⁴

We study the convergence in total variation and V -norm of discretization schemes of the underdamped Langevin dynamics. Such algorithms are very popular and commonly used in molecular dynamics and computational statistics to approximatively sample from a target distribution of interest. We show first that, for a very large class of schemes, a minorization condition uniform in the stepsize holds. This class encompasses popular methods such as the Euler-Maruyama scheme and the schemes based on splitting strategies. Second, we provide mild conditions ensuring that the class of schemes that we consider satisfies a geometric Foster–Lyapunov drift condition, again uniform in the stepsize. This allows us to derive geometric convergence bounds, with a convergence rate scaling linearly with the stepsize. This kind of result is of prime interest to obtain estimates on norms of solutions to Poisson equations associated with a given numerical method.

¹Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190 Gif-sur-Yvette, France. Email: alain.durmus@ens-paris-saclay.fr

²Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, Palaiseau, France. Email: aurelien.enfroy@ens-paris-saclay.fr

³CMAP - École polytechnique, France. Email: eric.moulines@polytechnique.edu

⁴CERMICS, Ecole des Ponts, Marne-la-Vallée, France & MATERIALS team-project, Inria Paris, France. Email: gabriel.stoltz@enpc.fr

3.1 Introduction

Langevin dynamics are nowadays one of the default dynamics to sample configurations of molecular systems in computational statistical physics; see for instance [2, 58, 90] for reference textbooks on molecular dynamics, as well as the more mathematically oriented works [82, 86, 108]. They are also gaining increasing popularity in Bayesian statistics and machine learning [10, 118] to obtain approximate samples from the a posteriori distribution of a statistical model [26, 31]. In this paper, we are interested in the Langevin dynamics, sometimes coined underdamped or kinetic, which describes the evolution of the position $(X_t)_{t \geq 0}$ and the velocity $(V_t)_{t \geq 0}$ of a system by the $2d$ -dimensional stochastic differential equation (SDE):

$$\begin{aligned} dX_t &= V_t dt, \\ dV_t &= [b(X_t) - \kappa V_t] dt + \sigma dB_t. \end{aligned} \tag{3.1}$$

Here, $\kappa, \sigma > 0$ are some friction and diffusion coefficients respectively, and $(B_t)_{t \geq 0}$ is a standard d -dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geq 0})$ satisfying the usual conditions. Note that, for notational simplicity, the mass matrix of the system is set to be the identity matrix and the friction coefficient is a scalar, independent of the position. It would nonetheless be possible to consider symmetric definite positive mass matrices, and position dependent friction matrices. A typical choice for the vector field $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ is $b = -\nabla U$ for some potential energy U , in which case the unique invariant probability measure of (3.1) is the Boltzmann–Gibbs probability measure, whose density with respect to the Lebesgue measure is proportional to $(x, v) \mapsto \exp(-\kappa[2U(x) + \|v\|^2]/\sigma^2)$. However, we are also interested in situations where the drift does not arise from a gradient, as in nonequilibrium molecular dynamics simulations [29, 55] (see for instance [86, Section 5] for a mathematical introduction to this field).

There are various techniques to prove the convergence of the continuous dynamics (3.1), for instance hypoocoercivity techniques [126, 130], Lyapunov [95, 117, 137] and coupling [49] methods. Moreover, although the dynamics is degenerate, *i.e.* the covariance matrix associated with (3.1) is not invertible, it can be shown using a combination of controllability arguments and hypoellipticity [117], that, for any $t_0 > 0$ and initial condition $(X_0, V_0) = (x_0, v_0) \in \mathbb{R}^{2d}$, the random variable (X_{t_0}, V_{t_0}) has a distribution with a positive density $(x_{t_0}, v_{t_0}) \mapsto p_{t_0}((x_0, v_0), (x_{t_0}, v_{t_0}))$ with respect to the Lebesgue measure on \mathbb{R}^{2d} , implying that the Markov semigroup associated with (3.1) is irreducible. Finally, quantitative bounds on p_{t_0} for any $t_0 > 0$ can be established by various techniques, such as Malliavin calculus [8], representation formulas for pinned diffusions together with comparison principles [111, 112], or through Gaussians bounds obtained with the so-called parametrix method [76, 85].

Obtaining quantitative convergence rates at the discrete level for discretization scheme associated with (3.1) is more difficult. In particular, it is of prime concern to establish convergence rates ρ_γ for discretization schemes with timestep $\gamma > 0$ which are similar to their continuous counterpart, *i.e.* such that $\log(\rho_\gamma)$ scales linearly in the

timestep. More precisely, if we denote by R_γ the Markov kernel associated with a given discretization scheme, it is sensible to expect that, under appropriate conditions and for any $\gamma > 0$ small enough, this kernel admits an invariant distribution π_γ and $\mathbf{d}(\mu_0 R_\gamma^k, \pi_\gamma) \leq C \bar{\rho}^{\gamma k}$ for an initial distribution μ_0 , where $C \geq 0$ and $\bar{\rho} \in [0, 1)$, while \mathbf{d} is some distance on the set of probability measures on \mathbb{R}^{2d} . Currently, one of the main options to obtain such a convergence is to prove Lyapunov estimates and minorization conditions which are uniform in the timestep, *i.e.* that such conditions holds for $R_\gamma^{\lceil t_0/\gamma \rceil}$ for some $t_0 > 0$ for constants which do not depend on γ . This strategy was used for overdamped Langevin dynamics in [16] and [33]. While Lyapunov conditions are based on direct algebraic computations, and may require to consider implicit schemes [77, 95], showing a minorization condition uniform in the timestep is the main bottleneck of this approach. For non-degenerate stochastic dynamics, this type of results can be established relying on Malliavin calculus [9], but the resulting proof is rather involved. It may also be possible to write a direct proof as in [16] from the Girsanov theorem, although this is however possible only for dynamics with additive noise. Finally, the coupling approach developed in [33, 44] cannot be applied to discretizations of the degenerate SDE (3.1).

In this paper, we consider another approach to obtaining minorization conditions uniform in the timestep: the idea is to consider the numerical scheme over small physical times as a perturbation of a given Gaussian process. This approach was first advocated by one of the authors to easily present the main results and rationale from [33, 44] for discretization schemes of the overdamped Langevin dynamics; see Section 3.3. It turns out that this approach can be extended to various discretization schemes for the underdamped Langevin dynamics, in particular the splittings schemes proposed in [18, 84], which are becoming increasingly popular in molecular dynamics (see [2, Section 12.2]). This contribution allows to amend and correct the proof of [83, Lemma 2.8] and extend this result to unbounded spaces.

Outline of the work. The present document is organized as follows. We present the numerical schemes we consider and state the main results we obtain in Section 3.2; see in particular Theorem 3.1 for the minorization condition, and Theorem 3.4 for the Lyapunov condition. Both results are stated so that the dependence on the timestep is explicit. The remaining sections are devoted to the proofs of these results. The proof of the minorization condition is written in Section 3.4, with, for pedagogical purposes, a sketch of the proof for nondegenerate dynamics in Section 3.3; while the proof of the Lyapunov condition can be read in Section 3.5. For completeness, some proofs and derivations are deferred to the appendix.

Notation. In order to present more concisely our results, we use the following notation throughout this work. For $m, n, p, q \in \mathbb{N}^*$, the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ of a $m \times n$ matrix $\mathbf{A} = (a_{i,j})_{(i,j) \in \{1, \dots, m\} \times \{1, \dots, n\}}$ and a $p \times q$ matrix $\mathbf{B} = (b_{i,j})_{(i,j) \in \{1, \dots, p\} \times \{1, \dots, q\}}$ is the $pm \times qn$ dimensional matrix with entries

$$\forall (i, j) \in \{1, \dots, pm\} \times \{1, \dots, qn\}, \quad (\mathbf{A} \otimes \mathbf{B})_{i,j} = a_{\lceil i/p \rceil, \lceil j/q \rceil} b_{i - \lfloor (i-1)/p \rfloor p, j - \lfloor (j-1)/q \rfloor q}.$$

Equivalently, and more explicitly,

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}.$$

For two symmetric matrices \mathbf{A} and \mathbf{B} , we say that $\mathbf{A} \succeq \mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is positive semi-definite. We denote by $\mathbf{0}_d$ and $\mathbf{1}_d$ the d -dimensional vectors with all components equal to 0 and 1 respectively.

The set $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ -field of \mathbb{R}^d . The Euclidean scalar product of vectors x and y in \mathbb{R}^d is denoted by $\langle x, y \rangle = x^\top y$, the Euclidean norm of x being $\|x\|$. For any $n \in \mathbb{N}^*$ and for any matrix \mathbf{A} of size $n \times n$, the notation $\|\mathbf{A}\|_{\text{op}}$ stands for the induced norm defined by $\|\mathbf{A}\|_{\text{op}} = \sup\{\|\mathbf{A}x\| : x \in \mathbb{R}^n \text{ with } \|x\| = 1\}$. For $k, n, m \in \mathbb{N}^*$, the set of k -times continuously differentiable functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is denoted by $C^k(\mathbb{R}^n, \mathbb{R}^m)$. For $f \in C^1(\mathbb{R}^d, \mathbb{R})$, ∇f is the gradient of f , and for $f \in C^2(\mathbb{R}^d, \mathbb{R})$ we denote by Δf the Laplacian of f . When $f \in C^1(\mathbb{R}^{2d}, \mathbb{R})$, $\nabla_x f$ the gradient of f restricted to the first d components and $\nabla_v f$ the gradient of f restricted to the last d components. When $f \in C^2(\mathbb{R}^{2d}, \mathbb{R})$, $\Delta_x f$ is the Laplacian of f restricted to the first d components and $\Delta_v f$ the Laplacian of f restricted to the last d components. The closed ball centered at $\tilde{x}_0 \in \mathbb{R}^d$ (with $d \geq 1$) with radius $M \geq 0$ is $\bar{B}_d(x_0, M) = \{x \in \mathbb{R}^d : \|x - x_0\| \leq M\}$. Finally, the d -dimensional standard normal distribution is denoted φ and by abuse of notation its density with respect to the Lebesgue measure is also denoted by $z \mapsto \varphi(z)$. For some measurable functions $\mathcal{V} : \mathbb{R}^n \rightarrow [1, \infty)$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}$, we define $\|g\|_{\mathcal{V}} = \sup_{x \in \mathbb{R}^n} \{|g(x)|/\mathcal{V}(x)\} < \infty$. The \mathcal{V} -norm of a signed measure ξ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is defined as $\|\xi\|_{\mathcal{V}} = \int_{\mathbb{R}^n} \mathcal{V}(x) d|\xi|(x)$, where $|\xi|$ is the absolute value of ξ . In the case $\mathcal{V} \equiv 1$, the \mathcal{V} -norm is the total variation norm and it is denoted by $\|\cdot\|_{\text{TV}}$. Equivalently (see [37, Theorem D.3.2] for details), $\|\xi\|_{\mathcal{V}}$ can be defined as $\|\xi\|_{\mathcal{V}} = \sup\{\xi(g) : \|g\|_{\mathcal{V}} \leq 1\}$.

3.2 Setting and main results

We first discuss in Section 3.2.1 the general structure of the discretization schemes we consider, and illustrate the relevance of the structural assumptions we make by various examples in Section 3.2.2. We then state the main results of this work in Section 3.2.3, namely minorization and drift conditions uniform in the discretization timestep, from which we immediately deduce a geometric convergence with a rate uniform in the timestep as well.

3.2.1 Structural assumptions on the numerical schemes

Discretization schemes for (3.1) are obtained in practice by introducing a positive timestep $\gamma > 0$. They correspond to a Markov chain $\{(X_k, V_k)\}_{k \in \mathbb{N}}$, where (X_k, V_k) approximates $(X_{k\gamma}, V_{k\gamma})$, the solution of (3.1) at time $k\gamma$. More precisely, we consider

3.2. Setting and main results

the following general structure on the induction defining the numerical schemes: for $k \in \mathbb{N}$,

$$\begin{aligned} X_{k+1} &= X_k + \gamma V_k + \gamma f_\gamma \left(X_k, \gamma^\delta V_k, \gamma^{\delta+1/2} \sigma_\gamma Z_{k+1}, W_{k+1} \right) + \gamma^{\delta+1/2} \sigma_\gamma \mathbf{D}_\gamma Z_{k+1}, \\ V_{k+1} &= \tau_\gamma V_k + \gamma g_\gamma \left(X_k, \gamma^\delta V_k, \gamma^{\delta+1/2} \sigma_\gamma Z_{k+1}, W_{k+1} \right) + \sqrt{\gamma} \sigma_\gamma Z_{k+1}, \end{aligned} \quad (3.2)$$

where $\delta > 0$ is a positive parameter, the family $(W_{k+1})_{k \in \mathbb{N}}$ is a sequence of independent and identically distributed (i.i.d.) random variables with common distribution μ_w on a measurable space $(\mathbf{W}, \mathcal{W})$, independent of the family $(Z_{k+1})_{k \in \mathbb{N}}$ of i.i.d. d -dimensional standard Gaussian random variables. In many cases of interest (see Section 3.2.2 below), $(W_{k+1})_{k \in \mathbb{N}}$ is a family of i.i.d. standard Gaussian random variables. The actual numerical schemes under consideration are encoded by the measurable functions $f_\gamma, g_\gamma : \mathbb{R}^{3d} \times \mathbf{W} \rightarrow \mathbb{R}^d$, as well as by $\sigma_\gamma, \tau_\gamma > 0$ and $\mathbf{D}_\gamma \in \mathbb{R}^{d \times d}$. We illustrate the choice of the form for the recursion (3.2) by several discretization schemes for (3.1) in Section 3.2.2. In particular, the parameter δ is 1 in all these examples, but the result we present can be stated for any positive value of δ . Let us also emphasize that some arguments in the functions f_γ, g_γ are scaled by powers of γ . This is motivated by the stability analysis in our proof of the minorization condition, and allows to state Lipschitz conditions independent of the timestep in A2 below. Finally, we always assume that $\gamma \in (0, \bar{\gamma}]$, for some fixed $\bar{\gamma} > 0$. Typically, $\bar{\gamma}$ represents a threshold which ensures that the scheme under consideration is stable. Here, for ease of presentation, we assume in A1 below that $\bar{\gamma}$ is even smaller than a specific constant.

We consider the following assumptions on the coefficients and functions entering (3.2). The first assumption A1 expresses some form of consistency of the coefficients $\sigma_\gamma, \tau_\gamma$ in (3.2), which are related to the coefficients σ, κ in (3.1). In addition, we also impose some upper bound on $\bar{\gamma}$ to simplify the derivation of our main results but it could be easily relaxed.

A1. 1) *There exists $C_\kappa \geq 0$ such that $\bar{\gamma} \leq (\kappa + 2C_\kappa/\kappa)^{-1}$, and for any $\gamma \in (0, \bar{\gamma}]$, it holds $|\tau_\gamma - e^{-\kappa\gamma}| \leq C_\kappa \gamma^2$. In addition, for any $\gamma \in (0, \bar{\gamma}]$, $\tau_\gamma \in (0, 1)$.*

2) *There exist $\bar{\sigma}, \mathcal{D} \in \mathbb{R}_+$ such that $\sup_{\gamma \in (0, \bar{\gamma}]} \sigma_\gamma \leq \bar{\sigma}$ and $\sup_{\gamma \in (0, \bar{\gamma}]} \|\mathbf{D}_\gamma\|_{\text{op}} \leq \mathcal{D}$. Finally,*

$$\lim_{\gamma \downarrow 0} \sigma_\gamma = \sigma.$$

The conditions in A2 express some form of Lipschitz stability with respect to scaled variables, and quantifies the fact that perturbations arising from f_γ, g_γ can be of order γ with respect to positions, while they are restricted to be of order $\gamma^{1+\delta}$ with respect to momenta.

A2. *For all $w \in \mathbf{W}$, the functions $(x, v, z) \mapsto (f_\gamma(x, v, z, w), g_\gamma(x, v, z, w))$ are C^1 . In addition, there exists $L \geq 0$ such that, for any $\gamma \in (0, \bar{\gamma}]$, $w \in \mathbf{W}$ and $(x, v, z), (x', v', z') \in \mathbb{R}^{3d}$,*

$$\begin{aligned} \|f_\gamma(x, v, z, w) - f_\gamma(x', v', z', w)\| &\leq L [\|(x, v) - (x', v')\| + \|z - z'\|], \\ \|g_\gamma(x, v, z, w) - g_\gamma(x', v', z', w)\| &\leq L [\|(x, v) - (x', v')\| + \|z - z'\|]. \end{aligned}$$

Note that we could assume that **A2** only holds for μ_w -almost all $w \in W$, but strengthen this condition for ease of presentation.

Euler–Maruyama scheme. We first illustrate our choice of the framework specified by (3.2) by the Euler–Maruyama discretization of (3.1), which reads

$$X_{k+1} = X_k + \gamma V_k, \quad V_{k+1} = (1 - \kappa\gamma)V_k + \gamma b(X_k) + \sqrt{\gamma}\sigma Z_k. \quad (3.3)$$

This numerical scheme fits into the framework (3.2) upon taking

$$\tau_\gamma = 1 - \kappa\gamma, \quad \sigma_\gamma = \sigma, \quad \mathbf{D}_\gamma = 0, \quad f_\gamma(x, v, z, w) = 0, \quad g_\gamma(x, v, z, w) = b(x).$$

The parameter δ is irrelevant. Note that **A1** and **A2** hold under the condition that b is Lipschitz.

3.2.2 Examples of numerical schemes of the form (3.2) satisfying **A1–A2**

In what follows, we show that most popular discretization schemes associated to (3.1) can be cast into the framework that we consider in (3.2) and satisfy **A1** and **A2**. All these numerical schemes can be obtained by decomposing the generator \mathcal{A} of (3.1) as

$$\mathcal{A} = \mathcal{A} + \mathcal{B} + \mathcal{C}, \quad (3.4)$$

where, for any $g \in C^2(\mathbb{R}^{2d})$ and $(x, v) \in \mathbb{R}^{2d}$, the elementary operators $\mathcal{A}, \mathcal{B}, \mathcal{C}$ act as

$$\begin{aligned} \mathcal{A}g(x, v) &= v^\top \nabla_x g(x, v), \\ \mathcal{B}g(x, v) &= b(x)^\top \nabla_v g(x, v), \\ \mathcal{C}g(x, v) &= -\kappa v^\top \nabla_v g(x, v) + \frac{\sigma^2}{2} \Delta_v g(x, v). \end{aligned} \quad (3.5)$$

Note that the dynamics associated with \mathcal{A} and \mathcal{B} simply correspond to the deterministic flows $t \mapsto (x + vt, v)$ and $t \mapsto (x, v + b(x)t)$, while the dynamics associated with \mathcal{C} is the Ornstein–Uhlenbeck process $(x, e^{-\kappa t}v + \sigma \int_0^t e^{-\kappa(t-s)} dB_s)_{t \geq 0}$. Note that, for any $t > 0$, the random variable $e^{-\kappa t}v + \sigma \int_0^t e^{-\kappa(t-s)} dB_s$ is Gaussian with mean $e^{-\kappa t}v$ and covariance matrix $t\tilde{\sigma}_t^2 I_d$, where

$$\tilde{\sigma}_t^2 = \sigma^2 \frac{1 - e^{-2\kappa t}}{2\kappa t}. \quad (3.6)$$

3.2.2.1 Stochastic exponential Euler scheme

The elementary stochastic dynamics with generator $\mathcal{A} + \mathcal{C}$ is also analytically integrable and corresponds to an Ornstein–Uhlenbeck process (see Lemma 21). One obtains the following numerical scheme by this analytic integration, upon fixing the drift to the current value $b(X_k)$:

$$\begin{aligned} X_{k+1} &= X_k + \frac{1 - e^{-\kappa\gamma}}{\kappa} V_k + \frac{\kappa\gamma + e^{-\kappa\gamma} - 1}{\kappa^2} b(X_k) + \eta_{k+1}, \\ V_{k+1} &= e^{-\kappa\gamma} V_k + \frac{1 - e^{-\kappa\gamma}}{\kappa} b(X_k) + \xi_{k+1}, \end{aligned} \quad (3.7)$$

3.2. Setting and main results

with

$$\xi_{k+1} = \sigma \int_0^\gamma e^{-\kappa(\gamma-s)} dB_{k\gamma+s}, \quad \eta_{k+1} = \sigma \int_0^\gamma \frac{1 - e^{-\kappa(\gamma-s)}}{\kappa} dB_{k\gamma+s}. \quad (3.8)$$

The scheme (3.7) corresponds to a stochastic exponential Euler integrator, see [28, 75] and references therein in a general framework. It has also been recently considered and studied in the machine learning community, starting with [26]. Similar schemes were in fact developed in the molecular dynamics communities in the late 70s based on the analytical solution of Langevin dynamics for $b = 0$ provided in [23], see for instance [53].

Note that the random variable $(\eta_{k+1}, \xi_{k+1})_{k \in \mathbb{N}}$, given in (3.8), is a family of i.i.d. Gaussian random vectors with covariance matrix

$$\Sigma^{(\gamma)} \otimes \mathbf{I}_d = \begin{pmatrix} \Sigma_1^{(\gamma)} & \Sigma_2^{(\gamma)} \\ \Sigma_2^{(\gamma)} & \Sigma_3^{(\gamma)} \end{pmatrix} \otimes \mathbf{I}_d, \quad (3.9)$$

where $\Sigma^{(\gamma)}$ has entries

$$\Sigma_1^{(\gamma)} = \sigma^2 \int_0^\gamma \left(\frac{1 - e^{-\kappa(\gamma-s)}}{\kappa} \right)^2 ds = \frac{\sigma^2}{2\kappa^2} \left[2\gamma - \frac{3 - 4e^{-\kappa\gamma} + e^{-2\kappa\gamma}}{\kappa} \right], \quad (3.10)$$

$$\Sigma_2^{(\gamma)} = \sigma^2 \int_0^\gamma e^{-\kappa(\gamma-s)} \frac{1 - e^{-\kappa(\gamma-s)}}{\kappa} ds = \frac{\sigma^2(1 - e^{-\kappa\gamma})^2}{2\kappa^2},$$

$$\Sigma_3^{(\gamma)} = \sigma^2 \int_0^\gamma e^{-2\kappa(\gamma-s)} ds = \frac{\sigma^2(1 - e^{-2\kappa\gamma})}{2\kappa}. \quad (3.11)$$

Note that $\Sigma_3^{(\gamma)}/\gamma = \tilde{\sigma}_\gamma^2$ (recall (3.6)). Introduce also $Z_{k+1} = \xi_{k+1}/\sqrt{\Sigma_3^{(\gamma)}}$, and W_{k+1} satisfying

$$\eta_{k+1} = \sqrt{\Sigma_1^{(\gamma)}} \left(\alpha_\gamma Z_{k+1} + \sqrt{1 - \alpha_\gamma^2} W_{k+1} \right), \quad \alpha_\gamma = \Sigma_2^{(\gamma)} / \sqrt{\Sigma_1^{(\gamma)} \Sigma_3^{(\gamma)}}.$$

It can be verified (see Lemma 20) that $\Sigma^{(\gamma)}$ is invertible and therefore $\alpha_\gamma < 1$. Then, an easy computation shows that $(Z_{k+1})_{k \in \mathbb{N}}$ and $(W_{k+1})_{k \in \mathbb{N}}$ are independent families of i.i.d. Gaussian random vectors with identity covariance matrix. With this notation, the numerical scheme (3.7) can be rewritten as

$$X_{k+1} = X_k + \frac{1 - e^{-\kappa\gamma}}{\kappa} V_k + \frac{\kappa\gamma + e^{-\kappa\gamma} - 1}{\kappa^2} b(X_k) + \sqrt{\Sigma_1^{(\gamma)}(1 - \alpha_\gamma^2)} W_{k+1} + \gamma^{3/2} \tilde{\sigma}_\gamma \mathbf{D}_\gamma Z_{k+1},$$

$$V_{k+1} = e^{-\kappa\gamma} V_k + \frac{1 - e^{-\kappa\gamma}}{\kappa} b(X_k) + \sqrt{\gamma} \tilde{\sigma}_\gamma Z_{k+1},$$

where $\mathbf{D}_\gamma = \Sigma_2^{(\gamma)} / \left[\tilde{\sigma}_\gamma \sqrt{\gamma^3 \Sigma_3^{(\gamma)}} \right]$. This fits into the framework (3.2) upon setting $\delta = 1$,

$$\tau_\gamma = e^{-\kappa\gamma}, \quad \sigma_\gamma = \tilde{\sigma}_\gamma,$$

$$g_\gamma(x, v, z, w) = \frac{1 - e^{-\kappa\gamma}}{\kappa\gamma} b(x),$$

$$f_\gamma(x, v, z, w) = \frac{1 - \kappa\gamma - e^{-\kappa\gamma}}{\kappa\gamma^2} \left(v - \frac{\gamma}{\kappa} b(x) \right) + \sqrt{\Sigma_1^{(\gamma)}(1 - \alpha_\gamma^2)} w.$$

The conditions in **A1** are easily seen to be satisfied since $\tilde{\sigma}_\gamma/\sigma \rightarrow 1$ as $\gamma \rightarrow 0$, while $\mathbf{D}_\gamma \rightarrow 1/2$. Finally, the conditions in **A2** hold true when b is Lipschitz.

3.2.2.2 Splitting schemes

A systematic way of constructing numerical schemes for Langevin dynamics is to rely on splitting procedures based on the decomposition (3.4), as systematically studied in [82, 83, 84] for kinetic Langevin dynamics. The interest of these schemes is that they reduce to symplectic integrators of the Hamiltonian dynamics as $\kappa \rightarrow 0$ with $\sigma^2 = O(\kappa)$. We consider here first and second order numerical discretizations based on Lie-Trotter or Strang splittings built upon the operators $\mathcal{A}, \mathcal{B}, \mathcal{C}$ introduced in (3.4). These operators are the generators of elementary SDEs which can be analytically integrated, as discussed after (3.5).

First order schemes. Schemes of weak order 1 are obtained, up to cyclic permutations, by composing the elementary dynamics associated with the operators in the order $\mathcal{A}, \mathcal{B}, \mathcal{C}$ or $\mathcal{A}, \mathcal{C}, \mathcal{B}$. There are therefore 6 possible first order splitting schemes, which can all be cast in the framework (3.2). Some general comments can be formulated on the 6 splittings schemes.

- There is no dependence on w for the functions f_γ, g_γ , and no dependence on z for g_γ .
- Functions f_γ which genuinely depend on z are obtained when \mathcal{C} appears before \mathcal{A} . This corresponds for example to the scheme $\mathcal{C}\mathcal{A}\mathcal{B}$.
- Functions g_γ which genuinely depend on v are obtained when \mathcal{A} appears before \mathcal{B} . This corresponds for example to the scheme $\mathcal{C}\mathcal{A}\mathcal{B}$.
- The functions f_γ, g_γ are quite similar for schemes where consecutive applications of \mathcal{B}, \mathcal{C} are exchanged (as for the evolutions $\mathcal{A}\mathcal{B}\mathcal{C}$ and $\mathcal{A}\mathcal{C}\mathcal{B}$; or for $\mathcal{B}\mathcal{C}\mathcal{A}$ and $\mathcal{C}\mathcal{B}\mathcal{A}$).

In view of these remarks, the simplest scheme, from a structural viewpoint, is associated with $\mathcal{B}\mathcal{A}\mathcal{C}$ (although the schemes associated with $\mathcal{A}\mathcal{B}\mathcal{C}$ and $\mathcal{A}\mathcal{C}\mathcal{B}$ are also quite simple), while the most complicated is the one associated with $\mathcal{C}\mathcal{A}\mathcal{B}$. We next write out more precisely these two schemes, as paradigmatic examples of first order splittings.

The numerical scheme associated with $\mathcal{B}\mathcal{A}\mathcal{C}$ reads

$$X_{k+1} = X_k + \gamma V_k + \gamma^2 b(X_k), \quad V_{k+1} = e^{-\kappa\gamma} V_k + \gamma e^{-\kappa\gamma} b(X_k) + \sqrt{\gamma} \tilde{\sigma}_\gamma Z_{k+1},$$

3.2. Setting and main results

with $\tilde{\sigma}_\gamma$ defined in (3.6). This corresponds to (3.2) with $\tau_\gamma = e^{-\kappa\gamma}$, $\sigma_\gamma = \tilde{\sigma}_\gamma$, $\mathbf{D}_\gamma = 0$ (δ is irrelevant), and $f_\gamma(x, v, z, w) = \gamma b(x)$, $g_\gamma(x, v, z, w) = e^{-\kappa\gamma} b(x)$. The numerical scheme associated with $\mathcal{C}\mathcal{A}\mathcal{B}$ reads

$$\begin{aligned} X_{k+1} &= X_k + \gamma e^{-\kappa\gamma} V_k + \gamma^{3/2} \tilde{\sigma}_\gamma Z_{k+1}, \\ V_{k+1} &= e^{-\kappa\gamma} V_k + \gamma b\left(X_k + \gamma e^{-\kappa\gamma} V_k + \gamma^{3/2} \tilde{\sigma}_\gamma Z_{k+1}\right) + \sqrt{\gamma} \tilde{\sigma}_\gamma Z_{k+1}, \end{aligned}$$

which corresponds to (3.2) with $\delta = 1$, $\tau_\gamma = e^{-\kappa\gamma}$, $\sigma_\gamma = \tilde{\sigma}_\gamma$, $\mathbf{D}_\gamma = \mathbf{I}_d$, and $f_\gamma(x, v, z, w) = \gamma^{-1}(e^{-\kappa\gamma} - 1)v$, $g_\gamma(x, v, z, w) = b(x + e^{-\kappa\gamma}v + z)$. For both schemes, the conditions **A1** and **A2** hold true when b is Lipschitz.

Second order schemes. Schemes of weak order 2 are obtained, up to cyclic permutations of the operators, by a Strang splitting based on the operators \mathcal{A} , \mathcal{B} , \mathcal{C} . The scheme $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$, for instance, corresponds to integrating the elementary dynamics associated with \mathcal{A} for a time $\gamma/2$, then the elementary dynamics associated with \mathcal{B} for a time $\gamma/2$, then the elementary dynamics associated with \mathcal{C} for a time γ , then again the elementary dynamics associated with \mathcal{B} for a time $\gamma/2$, and finally the elementary dynamics associated with \mathcal{A} for a time $\gamma/2$.

As for first order splitting schemes, there are 6 possible schemes, which can all be cast in the framework (3.2). Some general comments can be formulated on the 6 splittings schemes.

- There is no dependence on w for the functions f_γ, g_γ when \mathcal{C} is between instances of \mathcal{A} (as for the schemes $\mathcal{B}\mathcal{A}\mathcal{C}\mathcal{A}\mathcal{B}$, $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$ and $\mathcal{A}\mathcal{C}\mathcal{B}\mathcal{C}\mathcal{A}$). This dependence is linear for f_γ when the operator \mathcal{A} is in the central place (as for the schemes $\mathcal{C}\mathcal{B}\mathcal{A}\mathcal{B}\mathcal{C}$ and $\mathcal{B}\mathcal{C}\mathcal{A}\mathcal{C}\mathcal{B}$).
- The function g_γ does not depend on z, w for schemes with \mathcal{A} at the first and last places (as for the schemes $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$ and $\mathcal{A}\mathcal{C}\mathcal{B}\mathcal{C}\mathcal{A}$).
- The functions f_γ, g_γ are quite similar when consecutive operators \mathcal{B}, \mathcal{C} are exchanged (as for the schemes $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$ and $\mathcal{A}\mathcal{C}\mathcal{B}\mathcal{C}\mathcal{A}$; as well as for $\mathcal{B}\mathcal{C}\mathcal{A}\mathcal{C}\mathcal{B}$ and $\mathcal{C}\mathcal{B}\mathcal{A}\mathcal{B}\mathcal{C}$).

In view of these remarks, the simplest scheme, from a structural viewpoint, is associated with $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$ (or with $\mathcal{A}\mathcal{C}\mathcal{B}\mathcal{C}\mathcal{B}$), while the most complicated is the one associated with $\mathcal{C}\mathcal{A}\mathcal{B}\mathcal{A}\mathcal{C}$. We next write out more precisely these two schemes, as paradigmatic examples of second order splittings.

The numerical scheme associated with $\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{B}\mathcal{A}$ reads

$$\begin{aligned} X_{k+1} &= X_k + \frac{\gamma(1 + e^{-\kappa\gamma})}{2} V_k + \frac{\gamma^2(1 + e^{-\kappa\gamma})}{4} b\left(X_k + \frac{\gamma}{2} V_k\right) + \frac{\gamma^{3/2}}{2} \tilde{\sigma}_\gamma Z_{k+1}, \\ V_{k+1} &= e^{-\kappa\gamma} V_k + \frac{\gamma(1 + e^{-\kappa\gamma})}{2} b\left(X_k + \frac{\gamma}{2} V_k\right) + \sqrt{\gamma} \tilde{\sigma}_\gamma Z_{k+1}, \end{aligned}$$

where $\tilde{\sigma}_\gamma$ is given by (3.6). It can indeed be written as (3.2) with $\delta = 1$, $\tau_\gamma = e^{-\kappa\gamma}$, $\sigma_\gamma = \tilde{\sigma}_\gamma$, $\mathbf{D}_\gamma = \mathbf{I}_d/2$, and

$$\begin{aligned} f_\gamma(x, v, z, w) &= \frac{e^{-\kappa\gamma} - 1}{2\gamma}v + \frac{\gamma(1 + e^{-\kappa\gamma})}{4}b\left(x + \frac{1}{2}v\right), \\ g_\gamma(x, v, z, w) &= \frac{1 + e^{-\kappa\gamma}}{2}b\left(x + \frac{1}{2}v\right). \end{aligned}$$

The numerical scheme associated with $\mathcal{C}\mathcal{A}\mathcal{B}\mathcal{A}\mathcal{C}$ reads

$$\begin{aligned} X_{k+1} &= X_k + \gamma e^{-\kappa\gamma/2}V_k + \frac{\gamma^2}{2}b\left(X_k + \frac{\gamma e^{-\kappa\gamma/2}}{2}V_k + \frac{\gamma^{3/2}}{2\sqrt{2}}\tilde{\sigma}_{\gamma/2}\xi_{k+1}^1\right) + \frac{\gamma^{3/2}}{\sqrt{2}}\tilde{\sigma}_{\gamma/2}\xi_{k+1}^1, \\ V_{k+1} &= e^{-\kappa\gamma}V_k + \gamma e^{-\kappa\gamma/2}b\left(X_k + \frac{\gamma e^{-\kappa\gamma/2}}{2}V_k + \frac{\gamma^{3/2}}{2\sqrt{2}}\tilde{\sigma}_{\gamma/2}\xi_{k+1}^1\right) \\ &\quad + \sqrt{\frac{\gamma}{2}}\tilde{\sigma}_{\gamma/2}\left(e^{-\kappa\gamma/2}\xi_{k+1}^1 + \xi_{k+1}^2\right), \end{aligned}$$

where $(\xi_{k+1}^1)_{k \geq 0}$ and $(\xi_{k+1}^2)_{k \geq 0}$ are two independent families of i.i.d. standard d -dimensional Gaussian random variables. In order to write this scheme under the form (3.2), similarly to Section 3.2.2.1, we introduce the sequence of random variables $(Z_{k+1})_{k \in \mathbb{N}}$ defined as

$$\sqrt{\frac{\gamma}{2}}\tilde{\sigma}_{\gamma/2}\left(e^{-\kappa\gamma/2}\xi_{k+1}^1 + \xi_{k+1}^2\right) = \sqrt{\gamma}\sigma_\gamma Z_{k+1}, \quad \sigma_\gamma^2 = \tilde{\sigma}_{\gamma/2}^2 \frac{1 + e^{-\kappa\gamma}}{2}.$$

By construction, $(Z_{k+1})_{k \geq 0}$ are i.i.d. standard d -dimensional Gaussian random variables. We next consider the family of i.i.d. standard d -dimensional Gaussian random variables $(W_{k+1})_{k \geq 0}$ independent of $(Z_{k+1})_{k \geq 0}$ defined through the relation

$$\xi_{k+1}^1 = \alpha_\gamma Z_{k+1} + \sqrt{1 - \alpha_\gamma^2}W_{k+1}, \quad \alpha_\gamma = \mathbb{E}\left[\xi_{k+1}^1 Z_{k+1}\right] = \frac{e^{-\kappa\gamma/2}}{\sqrt{1 + e^{-\kappa\gamma}}} \in (0, 1).$$

In terms of the random variables Z_{k+1}, W_{k+1} , the scheme $\mathcal{C}\mathcal{A}\mathcal{B}\mathcal{A}\mathcal{C}$ can be reformulated as (3.2) with $\delta = 1$, $\tau_\gamma = e^{-\kappa\gamma}$, $\mathbf{D}_\gamma = e^{-\kappa\gamma/2}/(1 + e^{-\kappa\gamma})\mathbf{I}_d$ and

$$\begin{aligned} f_\gamma(x, v, z, w) &= \frac{e^{-\kappa\gamma/2} - 1}{\gamma}v + \frac{\gamma}{2}b\left(x + \frac{e^{-\kappa\gamma/2}}{2}v + \frac{e^{-\kappa\gamma/2}}{2(1 + e^{-\kappa\gamma})}z + \sqrt{\frac{\gamma^3\tilde{\sigma}_{\gamma/2}^2}{8(1 + e^{-\kappa\gamma})}}w\right) \\ &\quad + \sqrt{\frac{\gamma^3\tilde{\sigma}_{\gamma/2}^2}{2(1 + e^{-\kappa\gamma})}}w, \\ g_\gamma(x, v, z, w) &= e^{-\kappa\gamma/2}b\left(x + \frac{e^{-\kappa\gamma/2}}{2}v + \frac{e^{-\kappa\gamma/2}}{2(1 + e^{-\kappa\gamma})}z + \sqrt{\frac{\gamma^3\tilde{\sigma}_{\gamma/2}^2}{8(1 + e^{-\kappa\gamma})}}w\right). \end{aligned} \tag{3.12}$$

The conditions **A1** and **A2** are then easily seen to hold true for the two schemes $\mathcal{C}\mathcal{A}\mathcal{B}\mathcal{A}\mathcal{C}$ and $\mathcal{C}\mathcal{A}\mathcal{B}\mathcal{A}\mathcal{C}$ when b is Lipschitz.

3.2.2.3 Numerical schemes with stochastic gradients

We consider in this section discretizations of (3.1) using stochastic approximation strategies. Such a methodology is particularly appealing in statistics and machine learning where the field b can be very expensive to evaluate or cannot even be accessed [38, 136]. In these contexts, the Langevin dynamics have been primarily considered for either performing Bayesian inference [1, 35, 91] or optimizing an objective function [11, 15, 109, 114]. In the first case, $b = \nabla \log \pi$ where $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ is the a posteriori distribution of a statistical model, which can generally be written as $-\log \pi = \sum_{k=1}^N U_i$, with N the number of observations and $U_i : \mathbb{R}^d \rightarrow \mathbb{R}$. It has been also proposed to use Langevin dynamics to find an element of $\arg \min_{\mathbb{R}^d} f$ for some function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ by setting $b = -\nabla f$ and taking σ small. In particular, we are interested in the situation where ∇f can only be estimated through estimators which can potentially be biased.

The use of stochastic approximation for b in these two settings can be formalized as follows. We suppose that there exist a probability measure μ_Y on a measurable space (Y, \mathcal{Y}) , and a measurable function $H : \mathbb{R}^d \times Y \rightarrow \mathbb{R}^d$ such that

$$\forall x \in \mathbb{R}^d, \quad \tilde{b}(x) = \int_Y H_x(y) \mu_Y(dy), \quad \sup_{\mathbb{R}^d} \|\tilde{b} - b\| < +\infty.$$

In addition, we suppose that we can generate i.i.d. samples $(Y_{k+1})_{k \in \mathbb{N}}$ from μ_Y . Then, a stochastic approximation of the discretization of (3.1) essentially consists in replacing the evaluation of b at each iteration k by $H_{X_k}(Y_{k+1})$. For example, following [25], the Euler Maruyama discretization (3.3) can be generalized as

$$X_{k+1} = X_k + \gamma V_k, \quad V_{k+1} = (1 - \kappa\gamma)V_k + \gamma H_{X_k}(Y_{k+1}) + \sqrt{\gamma}\sigma Z_k.$$

This numerical scheme fits into the framework (3.2) upon taking $W_k = Y_k$ for any $k \geq 1$,

$$\tau_\gamma = 1 - \kappa\gamma, \quad \sigma_\gamma = \sigma, \quad \mathbf{D}_\gamma = 0, \quad f_\gamma(x, v, z, w) = 0, \quad g_\gamma(x, v, z, w) = H_x(w).$$

The parameter δ is irrelevant. Note that **A1** and **A2** hold under the condition that for any $y \in Y$, the function $x \mapsto H_x(y)$ is Lipschitz with a Lipschitz constant independent of y . The stochastic exponential Euler and splitting schemes presented in Sections 3.2.2.1 and 3.2.2.2 can also be adapted to take into account a stochastic approximation of b , for instance by relying on the methodology developed in [24].

3.2.3 Minorization, drift and convergence uniform in the timestep

To state our results, we introduce the Markov kernel R_γ associated with (3.2), *i.e.*, for any $(x, v) \in \mathbb{R}^{2d}$ and $B \in \mathcal{B}(\mathbb{R}^{2d})$,

$$R_\gamma((x, v), B) = \int_{\mathbb{R}^{d+m}} \mathbb{1}_B \left(\Gamma_\gamma(x, v, (\gamma\sigma_\gamma^2)^{1/2}z, w) \right) \varphi(z) dz \mu_W(dw),$$

where $\varphi(z)$ is the density of the d -dimensional standard normal distribution, and

$$\Gamma_\gamma(x, v, z, w) = \left(x + \gamma v + \gamma f_\gamma \left(x, \gamma^\delta v, \gamma^\delta z, w \right) + \gamma^\delta \mathbf{D}_\gamma z, \tau_\gamma v + \gamma g_\gamma \left(x, \gamma^\delta v, \gamma^\delta z, w \right) + z \right). \quad (3.13)$$

We present in the following our main results which allow us to conclude to the uniform \mathcal{V} -geometric ergodicity for R_γ of the form $\|\delta_{(x,v)}R_\gamma^k - \pi_\gamma\|_{\mathcal{V}} \leq C\mathcal{V}(x,v)\rho^{\gamma k}$ for any $k \in \mathbb{N}$, $\gamma \in (0, \bar{\gamma}]$ small enough, and where $C \geq 0$ and $\rho \in [0, 1)$ are independent of γ . The proof of this result standardly follows from quantitative minorization and Lyapunov conditions.

Minorization condition uniform in the timestep. Our first main result shows that $R_\gamma^{\lfloor t_0/\gamma \rfloor}$ satisfies a minorization condition with a constant which depends only on the physical time $t_0 > 0$ (considered sufficiently small) and not on γ .

Theorem 3.1. *Assume that **A1** and **A2** hold. Then there exists $\bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$ and $M \geq 0$, there are $\varepsilon_{t_0, M} > 0$ and $\bar{\gamma}_{t_0} > 0$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$ and $(x, v), (x', v') \in \bar{B}_{2d}(\mathbf{0}_{2d}, M)$,*

$$\left\| R_\gamma^{\lfloor t_0/\gamma \rfloor + 1}((x, v), \cdot) - R_\gamma^{\lfloor t_0/\gamma \rfloor + 1}((x', v'), \cdot) \right\|_{\text{TV}} \leq 2(1 - \varepsilon_{t_0, M}). \quad (3.14)$$

In words, Theorem 3.1 ensures that for any compact set $K \subset \mathbb{R}^{2d}$, there exists $t_0, \gamma_{t_0} > 0$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$, K is 1-small for $R_\gamma^{\lfloor t_0/\gamma \rfloor}$. The proof of this result can be read in Section 3.4. The main steps are sketched out in the simple case of nondegenerate Langevin dynamics in Section 3.3.

Lyapunov condition uniform in the timestep. To ensure the existence of a unique stationary distribution μ_γ for R_γ , and obtain a rate of convergence to stationarity, we need to consider additional assumptions. We focus on conditions which allow to ensure the uniform V -geometric ergodicity of R_γ with a Lyapunov function built upon a function $U : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying the following condition.

D1. *The function $U : \mathbb{R}^d \rightarrow \mathbb{R}$ is C^1 , $U(x) \geq 0$ for any $x \in \mathbb{R}^d$, and $U(0) = 0$, $\nabla U(0) = 0$. In addition, there exists $L \geq 0$ such that for any $x, y \in \mathbb{R}^d$, $\|\nabla U(x) - \nabla U(y)\| \leq L\|x - y\|$.*

The condition $\inf_{\mathbb{R}^d} U \geq 0$ is not restrictive since any function bounded from below can be shifted to be nonnegative. The conditions $U(0) = 0$ and $\nabla U(0) = 0$ could be relaxed but allow to simplify some computations and are therefore considered for ease of presentation. When b comes from the gradient of a potential (up to some perturbation) function, as discussed around (3.18), then this potential is a natural candidate for the function U . In this case, **D1** is necessary to ensure the stability of most of the schemes.

The Lyapunov function $\bar{\mathcal{W}}_{\gamma, \varpi} : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ we consider is parametrized by a constant $\varpi > 0$, and is of the following exponential form for any $\gamma \in (0, \bar{\gamma}]$ and $x, v \in \mathbb{R}^d$:

$$\bar{\mathcal{W}}_{\gamma, \varpi}(x, v) = \exp\left(\varpi\sqrt{1 + \mathcal{W}_\gamma(x, v)}\right), \quad (3.15)$$

where $\mathcal{W}_\gamma : \mathbb{R}^{2d} \rightarrow \mathbb{R}_+$, given explicitly by (3.73), is the sum of $U(x)$ and a positive definite quadratic form in (x, v) . As in previous studies on discretization of Langevin

3.2. Setting and main results

dynamics such as [95, Equation (8.3)], the Lyapunov function $\overline{\mathcal{W}}_{\gamma, \varpi}$ depends on the stepsize γ through \mathcal{W}_γ (in fact, \mathcal{W}_γ converges as $\gamma \rightarrow 0$ to a Lyapunov function for the continuous dynamics (3.1), similarly to the family of Lyapunov functions considered in [95]; see [95, Equation (3.6)]).

By the estimates in Section 3.5 (see Lemma 16 and (3.90)), the design of \mathcal{W}_γ ensures that there exist $\underline{c}_w, \bar{c}_w \in \mathbb{R}_+$ and $\bar{\gamma}_w > 0$ such that, for any $\gamma \in (0, \bar{\gamma}_w]$,

$$\underline{c}_w \mathcal{V} \leq \mathcal{W}_\gamma \leq \bar{c}_w \mathcal{V}, \quad \mathcal{V}(x, v) = \|x\|^2 + \|v\|^2 + U(x).$$

Therefore, there exist $\varpi_1, \varpi_2 > 0$ such that for any $\varpi > 0$, $\gamma \in (0, \bar{\gamma}_w]$ and $x, v \in \mathbb{R}^d$,

$$\left[\overline{\mathcal{V}}_\varpi(x, v) \right]^{\varpi_1} \leq \overline{\mathcal{W}}_{\gamma, \varpi}(x, v) \leq \left[\overline{\mathcal{V}}_\varpi(x, v) \right]^{\varpi_2}, \quad (3.16)$$

where

$$\overline{\mathcal{V}}_\varpi(x, v) = \exp \left(\varpi \sqrt{1 + \mathcal{V}(x, v)} \right).$$

In particular, convergence bounds on $\|\delta_{(x,v)} R_\gamma^k - \mu_\gamma\|_{\overline{\mathcal{W}}_{\gamma, \varpi}}$ for $\varpi > 0$ (where μ_γ is the unique invariant probability measure associated with R_γ) thus imply convergence bounds on $\|\delta_{(x,v)} R_\gamma^k - \mu_\gamma\|_{\overline{\mathcal{V}}_\varpi^{\varpi_1}}$.

In view of the minorization condition provided by Theorem 3.1 and using for example [37, Theorem 19.4.1], it is sufficient to establish a Lyapunov condition for $R_\gamma^{\lceil t_0/\gamma \rceil}$ uniform in the stepsize $\gamma > 0$ in order to obtain exponential convergence bounds for R_γ . To this end, we first establish a Lyapunov condition for R_γ , under additional technical conditions on the family of functions f_γ, g_γ for $\gamma \in (0, \bar{\gamma}]$, in relation with the potential U considered in **D1**. To state these conditions, we introduce the function $\mathcal{F} : \mathbb{R}^{3d} \times \mathbb{R}^{m_1+m_2} \rightarrow \mathbb{R}_+$, such that for any $x, v, z \in \mathbb{R}^d$ and $w \in \mathbb{R}^{m_1+m_2}$,

$$\mathcal{F}(x, v, z, w) = \frac{\|\nabla U(x)\|^2}{L^2} + \|v\|^2 + \|z\|^2 + \|w\|^2 + \|x\|.$$

Note that the position x appears through the two terms $\|\nabla U(x)\|^2$ and $\|x\|$ (mind the fact that the latter norm is not squared).

D2. $W = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, $\mathcal{W} = \mathcal{B}(\mathbb{R}^{m_1}) \otimes \mathcal{B}(\mathbb{R}^{m_2})$ and there exist $\alpha_U, \zeta_U > 0$, $\delta_U \in (0, 1]$ and $C_U \geq 0$ for which, for any $\gamma \in (0, \bar{\gamma}]$, there is $\vartheta_\gamma \in \mathbb{R}$ with

$$\sup_{\gamma \in (0, \bar{\gamma}]} |\vartheta_\gamma| \leq \bar{\vartheta},$$

such that the following estimates hold: for any $x, v, z \in \mathbb{R}^d$ and for μ_w -almost every $w = (w_1, w_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$,

$$\begin{aligned} & \left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w \right) \right\|^2 + \left\| g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w \right) + \alpha_U \nabla U(x) \right\|^2 \\ & \leq C_U \left[1 + \gamma^{\delta_U} \mathcal{F} \left(x, v, \gamma^{1/2} \sigma_\gamma z, w \right) \right], \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} \langle x, f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w) \rangle &\leq \gamma^\delta \vartheta_\gamma \langle x, v \rangle + \gamma^{\delta_U} C_U \|x\| \|w_1\| \\ &\quad + C_U \left[1 + \gamma^{\delta_U} \mathcal{F}(x, v, \gamma^{1/2} \sigma_\gamma z, w) \right], \\ \langle x, g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w) \rangle &\leq -\zeta_U \left[\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right] \\ &\quad + C_U \gamma^{\delta_U} \left[1 + \mathcal{F}(x, v, \gamma^{1/2} \sigma_\gamma z, w) \right]. \end{aligned}$$

The condition (3.17) means that, at first order in γ , g_γ is some bounded perturbation of $\alpha_U \nabla U$, which holds in many applications. The parameter ϑ_γ is the prefactor of the term linear in v in the expression of f_γ . This explains why an unsigned term $\gamma^\delta \vartheta_\gamma \langle x, v \rangle$ appears on the right hand side of the third inequality above. Since f_γ and $g_\gamma + \alpha_U \nabla U$ corresponds to some form of second order expansion in the timestep γ for the schemes we consider, the fact that $\delta_U > 0$ in **D2** is relatively easy to verify. Finally, note that if **D2** is satisfied for $\delta_U > 1$, we can replace this parameter by $1 \wedge \delta_U$ upon modifying the constants appearing in this assumption.

Note that **D2** is satisfied by the Euler-Maruyama scheme (3.3) and the ones described in Section 3.2.2 in the case $b = -\nabla U$ with U satisfying **D1** and

$$\liminf_{\|x\| \rightarrow +\infty} \frac{\langle \nabla U(x), x \rangle}{\|x\| + \|\nabla U(x)\|^2} > 0. \quad (3.18)$$

The latter condition is satisfied by potentials which are convex or behaving at infinity as $\|x\|^a$ with $1 \leq a \leq 2$. We check in the appendix Section 3.6.7 the following result.

Proposition 3.2. *Assume that $b = -\nabla U$ with $U : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying **D1** and (3.18). Then the condition **D2** holds for the second order splitting scheme leading to (3.12).*

The other cases considered in Section 3.2.2 are easier. The condition **D2** also holds when b is an appropriate perturbation of $-\nabla U$.

In contrast to [95], we consider a weaker assumption even when b derives from a potential. More precisely, in the case $b = -\nabla U$, the condition (3.18) is strengthened in [95, Corollary 7.4] to $\liminf_{\|x\| \rightarrow +\infty} \langle \nabla U(x), x \rangle / (\|x\|^2 + \|\nabla U(x)\|^2) > 0$.

To ease the presentation, we restrict our study to the case when the extra noise satisfies the following conditions. By abuse of notation, we denote by φ the d -dimensional standard normal distribution.

D3. $\mu_w = \mu_{w_1} \otimes \mu_{w_2}$ where μ_{w_1} and μ_{w_2} are probability measures on \mathbb{R}^{m_1} and \mathbb{R}^{m_2} respectively, such that the following two conditions hold.

- 1) (a) There exists $C_S > 0$, such that, for any Lipschitz continuous function $h : \mathbb{R}^{d+m_1} \rightarrow \mathbb{R}$, with Lipschitz constant $\|h\|_{\text{Lip}}$, and any $s \in \mathbb{R}$, it holds $\int_{\mathbb{R}^{d+m_1}} |h| d(\varphi \otimes \mu_{w_1}) < +\infty$ and

$$\int_{\mathbb{R}^{d+m_1}} \exp(sh) d(\varphi \otimes \mu_{w_1}) \leq \exp\left(s \int_{\mathbb{R}^{d+m_1}} h d(\varphi \otimes \mu_{w_1}) + C_S \|h\|_{\text{Lip}}^2 \frac{s^2}{2}\right);$$

3.2. Setting and main results

(b) There exists $\tilde{L} \geq 0$ such that for any $x, v, z, z' \in \mathbb{R}^d$, $w_1, w'_1 \in \mathbb{R}^{m_1}$ and $w_2 \in \mathbb{R}^{m_2}$,

$$\begin{aligned} & \|f_\gamma(x, v, z, (w_1, w_2)) - f_\gamma(x, v, z', (w'_1, w_2))\| \\ & + \|g_\gamma(x, v, z, (w_1, w_2)) - g_\gamma(x, v, z', (w'_1, w_2))\| \leq \tilde{L} \|(z, w_1) - (z', w'_1)\|. \end{aligned}$$

2) There exists $\bar{\gamma}_W \in (0, \bar{\gamma}]$ such that $\int_{\mathbb{R}^{m_2}} e^{\bar{\gamma}_W \|w_2\|^2} \mu_{w_2}(dw_2) < +\infty$.

Remark 3.3. If μ_{w_1} admits a first moment and satisfies a log-Sobolev inequality with constant $\tilde{C}_S > 0$, i.e. for any continuously differentiable function $h : \mathbb{R}^{m_1} \rightarrow \mathbb{R}_+$ such that $\int_{\mathbb{R}^{m_1}} h(w_1) \mu_{w_1}(dw_1) = 1$,

$$\int_{\mathbb{R}^{m_1}} h(w_1) \log(h(w_1)) \mu_{w_1}(dw_1) \leq 2\tilde{C}_S \int_{\mathbb{R}^{m_1}} \|\nabla h\|^2 \mu_{w_1}(dw_1),$$

then by [7, Propositions 5.5.1 and 5.2.7], $\varphi \otimes \mu_{w_1}$ satisfies a log-Sobolev inequality with constant $\max(1, \tilde{C}_S)$. Therefore by Herbst's argument (see [7, Proposition 5.4.1] or [19, Theorem 5.5]), **D3-1)-(a)** holds. The latter case is satisfied in all our examples in Sections 3.2.2.1 and 3.2.2.2.

Note that **D3-2)** is not restrictive and covers the case where μ_{w_2} is (sub-)Gaussian, which is frequently the case for the schemes presented in Section 3.2.2.3.

Theorem 3.4. Assume that **A1**, **D1**, **D2** and **D3** hold. Then there exist $\varpi_U, \bar{\gamma}_U > 0$, $\lambda_U \in (0, 1)$ and $R_U, b_U \geq 0$ such that, for any $\gamma \in (0, \bar{\gamma}_U]$,

$$\forall x, v \in \mathbb{R}^d, \quad R_\gamma \overline{\mathcal{W}}_{\gamma, \varpi_U}(x, v) \leq \lambda_U^\gamma \overline{\mathcal{W}}_{\gamma, \varpi_U}(x, v) + \gamma b_U \mathbb{1}_{[0, R_U]}(\|x\| + \|v\|),$$

where $\overline{\mathcal{W}}_{\gamma, \varpi}$ is defined in (3.15).

The proof of this result is postponed to Section 3.5.

Exponential convergence uniform in the timestep When the statement of Theorem 3.4 holds, [43, Lemma 1] implies that, for any $k \in \mathbb{N}^*$ and $\gamma \in (0, \bar{\gamma}_U]$,

$$\forall x, v \in \mathbb{R}^d, \quad R_\gamma^k \overline{\mathcal{W}}_{\gamma, \varpi_U}(x, v) \leq \lambda_U^{k\gamma} \overline{\mathcal{W}}_{\gamma, \varpi_U}(x, v) + b_U \frac{1}{\lambda_U^{\bar{\gamma}_U} |\log \lambda_U|}.$$

Therefore, in view of [37, Theorem 19.4.1] applied to R_γ^k with $k = \lceil t_0/\gamma \rceil$ where t_0 is such that the conclusions of Theorem 3.1 hold, we obtain:

Theorem 3.5. Assume that **A1**, **A2**, **D1**, **D2** and **D3** hold. Then there exists $\bar{\gamma}_U > 0$ such that, for any $\gamma \in (0, \bar{\gamma}_U]$, the Markov chain associated with R_γ admits a unique invariant probability measure μ_γ . Moreover, there exist $\varpi_U > 0$, $A_U \geq 0$ and $\rho_U \in (0, 1)$ such that for any $\gamma \in (0, \bar{\gamma}_U]$,

$$\forall x, v \in \mathbb{R}^d, \quad \|\delta_{(x,v)} R_\gamma^k - \mu_\gamma\|_{\overline{\mathcal{W}}_{\gamma, \varpi_U}} \leq A_U \rho_U^{k\gamma} \overline{\mathcal{W}}_{\gamma, \varpi_U}(x, v). \quad (3.19)$$

Then, combining this result and (3.16) we get that there exist $\varpi_1, \varpi_2 > 0$ such that, for any $\gamma \in (0, \bar{\gamma}_U]$,

$$\forall x, v \in \mathbb{R}^d, \quad \left\| \delta_{(x,v)} R_\gamma^k - \mu_\gamma \right\|_{\overline{\mathcal{V}}^{\varpi_1}} \leq A_U \rho_U^{k\gamma} \overline{\mathcal{V}}^{\varpi_2}(x, v),$$

where $\overline{\mathcal{V}}^{\varpi}$ is defined in (3.16).

Another corollary of (3.19) is the following: there exists $K_U \geq 0$ (which can be computed explicitly in terms of A_U, ρ_U) such that, for any measurable function $\phi : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ with $\int_{\mathbb{R}^d} \phi d\pi_\gamma = 0$ and $\|\phi\|_{\overline{\mathcal{W}}_{\gamma, \varpi_U}} < +\infty$, and for any $\gamma \in (0, \bar{\gamma}_U]$, the function $\psi_\gamma = \sum_{k=0}^{+\infty} R_\gamma^k \phi$ is well-defined, satisfies $\|\psi_\gamma\|_{\overline{\mathcal{W}}_{\gamma, \varpi_U}} \leq K_U \gamma^{-1} \|\phi\|_{\overline{\mathcal{W}}_{\gamma, \varpi_U}}$ and is solution of the Poisson equation associated with R_γ and ϕ , namely $\{\text{Id} - R_\gamma\} \psi_\gamma = \phi$.

3.3 Strategy of proof of Theorem 3.1 in a simple case

We present in this section the main ideas behind the proof of Theorem 3.1. We illustrate the strategy in the simplest case, namely a Euler–Maruyama discretization of (overdamped) Langevin dynamics:

$$\overline{X}_{k+1} = \overline{X}_k + \gamma b(\overline{X}_k) + \sqrt{\gamma} Z_{k+1} = \Upsilon_\gamma(\overline{X}_k, \sqrt{\gamma} Z_{k+1}), \quad \Upsilon_\gamma(x, z) = x + \gamma b(x) + z, \quad (3.20)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $(Z_{k+1})_{k \in \mathbb{N}}$ is a sequence of i.i.d. standard d -dimensional Gaussian random variables. Note that here Υ_γ plays the same role as Γ_γ in (3.13). In this context, the transition kernel reads, for $x \in \mathbb{R}^d$ and $B \in \mathcal{B}(\mathbb{R}^d)$, $Q_\gamma(x, B) = \int_{\mathbb{R}^d} \mathbb{1}_B \{ \Upsilon_\gamma(x, \sqrt{\gamma} z) \} \varphi(z) dz$, and the counterpart of Theorem 3.1 is the following result.

Theorem 3.6. *Assume that $b \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ is globally Lipschitz. Then there exists $\bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$ and $M \geq 0$, there are $\varepsilon_{t_0, M} > 0$ and $\bar{\gamma}_{t_0} > 0$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$ and $x, x' \in \overline{B}_d(\mathbf{0}_d, M)$,*

$$\left\| Q_\gamma^{\lceil t_0/\gamma \rceil}(x, \cdot) - Q_\gamma^{\lceil t_0/\gamma \rceil}(x', \cdot) \right\|_{\text{TV}} \leq 2(1 - \varepsilon_{t_0, M}). \quad (3.21)$$

To prove this result, we need to study the distribution of \overline{X}_k defined by (3.20) for k of order t_0/γ for some $t_0 > 0$. The main idea is that the recursion (3.20) can be seen as a perturbation of the evolution with $b \equiv 0$ provided t_0 is sufficiently small. To this end, we first notice that a straightforward induction gives, for any $k \in \mathbb{N}$,

$$\overline{X}_{k+1} = \overline{X}_0 + G^{(k+1)} + \gamma \sum_{i=0}^k b(\overline{X}_i), \quad G^{(k+1)} = \sqrt{\gamma} \sum_{i=1}^{k+1} Z_i. \quad (3.22)$$

If $t_0 > 0$ is fixed, the variance of $G^{(k+1)}$ is of order t_0 for $k \approx t_0/\gamma$, from which the proof of Theorem 3.6 easily follows when $b \equiv 0$. To treat the case $b \not\equiv 0$, we rewrite (3.22) as

$$\overline{X}_{k+1} = G^{(k+1)} + \Phi_{\gamma, \overline{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k)}(G^{(k+1)}), \quad (3.23)$$

3.3. Strategy of proof of Theorem 3.1 in a simple case

where $\tilde{\mathbf{Z}}^{(k+1)} = (\tilde{Z}_1^{(k+1)}, \dots, \tilde{Z}_k^{(k+1)})$ are some i.i.d. Gaussian random variables defined below and independent of $G^{(k+1)}$, and $\Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k)} : \mathbb{R}^d \rightarrow \mathbb{R}^d$. We show in addition that $g \mapsto \Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k)}(g)$ is Lipschitz with Lipschitz constant strictly smaller than 1 for t_0 sufficiently small. As a result, the mapping $g \mapsto g + \Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k)}(g)$ is a C^1 diffeomorphism, so that, by a change a variable, the random variable \bar{X}_{k+1} admits a density with respect to the Lebesgue measure, and this density can even be bounded from below.

In order to rigorously formalize the above discussion, we first need to provide expressions for the random variables $\tilde{\mathbf{Z}}^{(k+1)}$. Relying on Cochran's theorem, natural candidates are obtained by a linear combination of the first k original Gaussian increments (Z_1, \dots, Z_k) and $G^{(k+1)}$, as

$$\tilde{Z}_i^{(k+1)} = Z_i - \frac{1}{\sqrt{\gamma(k+1)}} G^{(k+1)}, \quad i \in \{1\}k. \quad (3.24)$$

The random vector $(\tilde{\mathbf{Z}}^{(k+1)}, G^{(k+1)})$ is still a Gaussian vector, with $\tilde{\mathbf{Z}}^{(k+1)}$ independent of $G^{(k+1)}$ by construction. In addition, $\tilde{Z}_1^{(k+1)}, \dots, \tilde{Z}_k^{(k+1)}$ are i.i.d. Gaussian random variables with covariance matrix $(k/[k+1])\text{Id}$. Then,

$$\bar{X}_{k+1} = \bar{\Upsilon}_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k+1)}(G^{(k+1)}), \quad \text{with} \quad \bar{\Upsilon}_{\gamma, x, \tilde{\mathbf{z}}^{(k+1)}}^{(k+1)}(g) = x + g + \gamma \sum_{i=0}^k b(x_i), \quad (3.25)$$

for a given family $\tilde{\mathbf{z}}^{(k+1)} = \{\tilde{z}_j\}_{j=1}^k \subset \mathbb{R}^d$, and where the iterates (x_1, \dots, x_k) are recursively defined from $x_0 = x$ and $\tilde{\mathbf{z}}^{(k+1)}$ as $x_i = \Upsilon_\gamma(x_{i-1}, \sqrt{\gamma}\tilde{z}_i + g/[k+1])$ for $i \in \{1\}k$. Then, (3.23) holds with $\Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k)}(g) = \bar{\Upsilon}_{\gamma, \bar{X}_0, \tilde{\mathbf{Z}}^{(k+1)}}^{(k+1)}(g) - g$. Introducing $n_0 = \lfloor t_0/\gamma \rfloor + 1$, the iterated transition kernel can then be expressed as follows:

$$Q_\gamma^{\lfloor t_0/\gamma \rfloor + 1}(x, B) = \int_{\mathbb{R}^{(n_0+1)d}} \mathbb{1}_B \left\{ \bar{\Upsilon}_{\gamma, x, \tilde{\mathbf{z}}^{(n_0)}}^{(n_0)}(g) \right\} \varphi_{t_0, \gamma}(g) \psi_{t_0, \gamma}(\tilde{\mathbf{z}}^{(n_0)}) d\tilde{\mathbf{z}}^{(n_0)} dg, \quad (3.26)$$

where $\varphi_{t_0, \gamma}$ and $\psi_{t_0, \gamma}$ correspond to the densities of $G^{(n_0)}$ and $\tilde{\mathbf{Z}}^{(n_0)}$, respectively. Here we have assumed without loss of generality, taking for example $\gamma \leq t_0/2$, that $n_0 \geq 1$.

Now that the output of the Markov chain has been rewritten as a perturbation of the output obtained with $b = 0$, we can proceed with a quantitative analysis to obtain lower bounds on (3.26). By a reasoning similar to the one leading to Proposition 3.8 below, there exists $\bar{t}_0 > 0$ (sufficiently small) such that the function $g \mapsto \Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{z}}^{n_0}}^{(n_0)}(g)$ in (3.23) is Lipschitz with Lipschitz constant strictly smaller than 1 for any $t_0 \in (0, \bar{t}_0]$ (provided $\gamma > 0$ is sufficiently small). Therefore, $\bar{\Upsilon}_{\gamma, \bar{X}_0, \tilde{\mathbf{z}}^{(n_0)}}^{(n_0)} = \text{Id} + \Phi_{\gamma, \bar{X}_0, \tilde{\mathbf{z}}^{n_0}}^{(n_0)}$ is a perturbation of identity and hence a $C^1(\mathbb{R}^d, \mathbb{R}^d)$ -diffeomorphism. Denoting by $\bar{\Upsilon}_{\gamma, \bar{X}_0, \tilde{\mathbf{z}}^{(n_0)}}^{(\leftarrow, n_0)}$ its inverse, we obtain, by a change of variable,

$$Q_\gamma^{\lfloor t_0/\gamma \rfloor + 1}(x, B) = \int_{\mathbb{R}^{(n_0+1)d}} \mathbb{1}_B(u) \varphi_{t_0, \gamma} \left(\bar{\Upsilon}_{\gamma, x, \tilde{\mathbf{z}}^{(n_0)}}^{(\leftarrow, n_0)}(u) \right) J_{\gamma, x, \tilde{\mathbf{z}}^{(n_0)}}(u) \psi_{t_0, \gamma}(\tilde{\mathbf{z}}^{(n_0)}) d\tilde{\mathbf{z}}^{(n_0)} du, \quad (3.27)$$

where $J_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}(u)$ is the absolute value of the determinant of the Jacobian matrix associated with $\bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}$.

We consider at this stage the point $x = \mathbf{0}_d$ as some reference initial condition, and provide a lower bound on $Q_\gamma^{\lfloor t_0/\gamma \rfloor + 1}(x, B)$ based on (3.27) in terms of the following minorization probability measure:

$$\mu_{t_0,\gamma}(B) = \int_{\mathbb{R}^{(n_0+1)d}} \mathbb{1}_B \left\{ \bar{\Upsilon}_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(n_0)}(g/\sqrt{2}) \right\} \varphi_{t_0,\gamma}(g) \psi_{t_0,\gamma}(\bar{\mathbf{z}}^{(n_0)}) d\bar{\mathbf{z}}^{(n_0)} dg.$$

This expression is similar to (3.26), with $x = \mathbf{0}_d$ and g replaced by $g/\sqrt{2}$ (in order to make use of the inequality (3.28) below). Similarly to (3.27), using the change of variable $g \mapsto \bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(n_0)}(g/\sqrt{2})$, we have

$$\mu_{t_0,\gamma}(B) = 2^{\frac{d}{2}} \int_{\mathbb{R}^{(n_0+1)d}} \mathbb{1}_B(u) \varphi_{t_0,\gamma} \left(\sqrt{2} \bar{\Upsilon}_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right) J_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}(u) \psi_{t_0,\gamma}(\bar{\mathbf{z}}^{(n_0)}) d\bar{\mathbf{z}}^{(n_0)} du.$$

From this expression and (3.27), we get

$$Q_\gamma^{\lfloor t_0/\gamma \rfloor + 1}(x, B) = 2^{-d/2} \int_{\mathbb{R}^{(n_0+1)d}} \mathbb{1}_B(u) A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}(u) \mu_{t_0,\gamma}(dg d\bar{\mathbf{z}}^{(n_0)}),$$

where $A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}(u) = A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(1)}(u) A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(2)}(u)$ with

$$A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(1)}(u) = \frac{J_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}(u)}{J_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}(u)}, \quad A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(2)}(u) = \frac{\varphi_{t_0,\gamma} \left(\bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right)}{\varphi_{t_0,\gamma} \left(\sqrt{2} \bar{\Upsilon}_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right)}.$$

When $A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}$ is lower bounded by a positive quantity $a_{t_0,M}$ for $x \in \bar{B}_d(\mathbf{0}_d, M)$, we obtain $Q_\gamma^{\lfloor t_0/\gamma \rfloor + 1}(x, B) \geq a_{t_0,M} 2^{-d/2} \mu_{t_0,\gamma}(B)$, which immediately implies (3.21). The lower bound on $A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}$ is proved in two steps:

(a) by obtaining upper and lower bounds on the Jacobians $J_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}(u)$, leveraging the fact that the mapping $g \mapsto \bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(n_0)}(g)$ and its inverse are Lipschitz (as in Proposition 3.8 below) and making use of Hadamard's inequality (see Proposition 3.10);

(b) by making use of the following inequality, which motivates the factor $\sqrt{2}$ in the argument of the denominator of $A_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(2)}(u)$ and the definition of $\mu_{t_0,\gamma}$:

$$\begin{aligned} n_0 \gamma \log A_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(2)}(u) &= \left\| \bar{\Upsilon}_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right\|^2 - \frac{1}{2} \left\| \bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right\|^2 \\ &\geq - \left\| \bar{\Upsilon}_{\gamma,x,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) - \bar{\Upsilon}_{\gamma,\mathbf{0}_d,\bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow,n_0)}(u) \right\|^2. \end{aligned}$$

Here we have used that $\varphi_{t_0,\gamma}$ is the density of $G^{(n_0)}$ defined in (3.22) and the Young inequality on $\langle a, b \rangle$ for $a, b \in \mathbb{R}^d$, which implies that

$$\|a - b\|^2 \geq \|a\|^2/2 - \|b\|^2. \quad (3.28)$$

3.4. Proof of the minorization condition

When $x \mapsto \bar{\Upsilon}_{\gamma, x, \bar{\mathbf{z}}^{(n_0)}}^{(\leftarrow, n_0)}(u)$ is Lipschitz on $\bar{\mathbb{B}}_d(\mathbf{0}_d, M)$ with a constant K_{t_0} , uniformly in $\gamma \in (0, \gamma_{t_0}]$ (as in Proposition 3.9 below), one finds $A_{\gamma, x, \bar{\mathbf{z}}^{(n_0)}}^{(2)}(u) \geq e^{-K_{t_0}^2 M^2 / (n_0 \gamma)}$ for all $x \in \bar{\mathbb{B}}_d(\mathbf{0}_d, M)$.

3.4 Proof of the minorization condition

The proof is organized in several steps. We first start by rewriting the end point of the Markov chain in terms of the initial conditions and noise increments in Section 3.4.1, in a form amenable to perturbative treatments. The dominant part of the evolution is given by the outcome of discretization schemes corresponding to (3.2) in the case $b = 0$. We therefore carefully study this case in Section 3.4.2, where we condition solutions by the sum of the random increments characterizing the endpoints, and write out a decomposition into intermediate increments independent of the sum. We next quantify, by stability estimates, how numerical solutions depend on the random increments used to generate them (see Section 3.4.3). This finally allows us to prove Theorem 3.1 in Section 3.4.4 by considering the actual solutions of the numerical method as perturbations of discretizations of (3.2) in the case $b = 0$, provided the physical time t_0 is sufficiently small.

3.4.1 Algebraic structure of the equations

It is convenient, in particular for the stability estimates of Section 3.4.3, to rewrite (X_{k+1}, V_{k+1}) as a function of (X_0, V_0) and the realizations of the random variables $\{(Z_{i+1}, W_{i+1})\}_{i=0}^k$ needed to define the $(k+1)$ -th iterate. First, in view of (3.2) and (3.13), the iterates of the Markov chain can be written as $(X_{k+1}, V_{k+1}) = \Gamma_\gamma(X_k, V_k, \sqrt{\gamma}\sigma_\gamma Z_{k+1}, W_{k+1})$. This allows to write

$$(X_{k+1}, V_{k+1}) = \Gamma_\gamma^{(k+1)}\left(X_0, V_0, \{\sqrt{\gamma}\sigma_\gamma Z_i\}_{i=1}^{k+1}, \{W_i\}_{i=1}^{k+1}\right), \quad (3.29)$$

where the applications $\Gamma_\gamma^{(i)} : \mathbb{R}^{2d} \times \mathbb{R}^{i \times d} \times \mathbb{W}^i \rightarrow \mathbb{R}^{2d}$ are recursively defined as follows:

$$\begin{aligned} \Gamma_\gamma^{(0)}(x, v) &= (x, v), \\ \Gamma_\gamma^{(i)}\left(x, v, \{z_j\}_{j=1}^i, \{w_j\}_{j=1}^i\right) &= \Gamma_\gamma\left(\Gamma_\gamma^{(i-1)}\left(x, v, \{z_j\}_{j=1}^{i-1}, \{w_j\}_{j=1}^{i-1}\right), z_i, w_i\right), \quad i \geq 1. \end{aligned} \quad (3.30)$$

The next result provides a more explicit and constructive expression for $\Gamma_\gamma^{(k+1)}$. It is stated for the sequence $(x_{k+1}, v_{k+1}) = \Gamma_\gamma(x_k, v_k, z_{k+1}, w_{k+1})$, with $(x_i, v_i) = \Gamma_\gamma^{(i)}(x, v, \{z_j\}_{j=1}^i, \{w_j\}_{j=1}^i)$ for $i \in \{1\} k + 1$ and a given initial condition $(x, v) \in \mathbb{R}^{2d}$.

Lemma 5. *Define, for $k \geq 0$, the vectors $\mathbf{g}_1^{(k+1)} = (\mathbf{g}_{1,1}^{(k+1)}, \dots, \mathbf{g}_{1,k+1}^{(k+1)})$ and $\mathbf{g}_2^{(k+1)} = (\mathbf{g}_{2,1}^{(k+1)}, \dots, \mathbf{g}_{2,k+1}^{(k+1)})$ with components*

$$\mathbf{g}_{1,i}^{(k+1)} = \frac{\gamma(1 - \tau_\gamma^{k-i+1})}{1 - \tau_\gamma}, \quad \mathbf{g}_{2,i}^{(k+1)} = \tau_\gamma^{k-i+1}, \quad i \in \{1\} k + 1. \quad (3.31)$$

Then,

$$\begin{aligned} x_{k+1} &= x + \gamma \frac{1 - \tau_\gamma^{k+1}}{1 - \tau_\gamma} v + \sum_{i=0}^{k-1} \mathbf{g}_{1,i+1}^{(k+1)} \left[\gamma g_\gamma(x_i, \gamma^\delta v_i, \gamma^\delta z_{i+1}, w_{i+1}) + z_{i+1} \right] \\ &\quad + \gamma \sum_{i=0}^k f_\gamma(x_i, \gamma^\delta v_i, \gamma^\delta z_{i+1}, w_{i+1}) + \gamma^\delta \sum_{i=0}^k \mathbf{D}_\gamma z_{i+1}, \\ v_{k+1} &= \tau_\gamma^{k+1} v + \sum_{i=0}^k \mathbf{g}_{2,i+1}^{(k+1)} \left[\gamma g_\gamma(x_i, \gamma^\delta v_i, \gamma^\delta z_{i+1}, w_{i+1}) + z_{i+1} \right]. \end{aligned}$$

The proof is obtained by a simple induction and the equalities $\tau_\gamma \mathbf{g}_{2,j}^{(k)} = \mathbf{g}_{2,j}^{(k+1)}$ and $\mathbf{g}_{1,i+1}^{(k)} + \gamma \mathbf{g}_{2,i+1}^{(k)} = \mathbf{g}_{1,i+1}^{(k+1)}$ for $j \in \{0\}$ k and $i \in \{0\}$ $k - 1$. Lemma 5 allows to rewrite (3.29) as:

$$\begin{aligned} \begin{pmatrix} X_{k+1} \\ V_{k+1} \end{pmatrix} &= \mathbf{M}_\gamma^{k+1} \begin{pmatrix} X_0 \\ V_0 \end{pmatrix} + \gamma \sum_{i=0}^k \Theta^{(i)}(X_i, V_i, \sqrt{\gamma} \sigma_\gamma Z_{i+1}, W_{i+1}) \\ &\quad + \sigma_\gamma \left[\mathbf{G}^{(k+1)} + \gamma^\delta \begin{pmatrix} \mathbf{G}_3^{(k+1)} \\ \mathbf{0}_d \end{pmatrix} \right], \end{aligned} \quad (3.32)$$

where the matrix which multiplies the initial condition reads

$$\mathbf{M}_\gamma^{k+1} = \left[\begin{pmatrix} 1 & \gamma(1 - \tau_\gamma^{k+1})/(1 - \tau_\gamma) \\ 0 & \tau_\gamma^{k+1} \end{pmatrix} \otimes \mathbf{I}_d, \right] \quad (3.33)$$

the drift part of the dynamics is encoded by

$$\Theta^{(i)}(x, v, z, w) = \begin{pmatrix} \mathbf{g}_{1,i+1}^{(k+1)} \\ \mathbf{g}_{2,i+1}^{(k+1)} \end{pmatrix} \otimes g_\gamma(x, \gamma^\delta v, \gamma^\delta z, w) + \begin{pmatrix} f_\gamma(x, \gamma^\delta v, \gamma^\delta z, w) \\ \mathbf{0}_d \end{pmatrix}, \quad i \in \{0, \dots, k\},$$

while the actual noise obtained at the end of the iterations is given by $\mathbf{G}_3^{(k+1)}$ and

$$\mathbf{G}^{(k+1)} = \left(\mathbf{G}_1^{(k+1)}, \mathbf{G}_2^{(k+1)} \right), \quad (3.34)$$

with

$$\mathbf{G}_1^{(k+1)} = \sqrt{\gamma} \sum_{i=0}^{k-1} \mathbf{g}_{1,i+1}^{(k+1)} Z_{i+1}, \quad \mathbf{G}_2^{(k+1)} = \sqrt{\gamma} \sum_{i=0}^k \mathbf{g}_{2,i+1}^{(k+1)} Z_{i+1}, \quad \mathbf{G}_3^{(k+1)} = \sqrt{\gamma} \mathbf{D}_\gamma \sum_{i=0}^k Z_{i+1}. \quad (3.35)$$

When **A2** is satisfied, it holds, for any $(x, v, z), (x', v', z') \in \mathbb{R}^{3d}$, $w \in \mathbb{W}$ and $i \in \{0\}$ k ,

$$\left\| \Theta_\gamma^{(i)}(x, v, z, w) - \Theta_\gamma^{(i)}(x', v', z', w) \right\| \leq (2 + k\gamma) \mathbf{L} \left(\|x - x'\| + \gamma^\delta \|v - v'\| + \gamma^\delta \|z - z'\| \right). \quad (3.36)$$

3.4. Proof of the minorization condition

This result easily follows from the bounds

$$\sup_{i \in \{0, \dots, k\}} |\mathbf{g}_{1,i+1}^{(k+1)}| \leq k\gamma, \quad \sup_{i \in \{0, \dots, k\}} |\mathbf{g}_{2,i+1}^{(k+1)}| \leq 1. \quad (3.37)$$

These bounds are in turn a consequence of the fact that $\tau_\gamma \in (0, 1)$ and, for the first inequality, $\mathbf{g}_{1,i}^{(k+1)} = \gamma \sum_{j=0}^{k-1} \tau_\gamma^j$.

3.4.2 Structure and properties of the noise

We study in this section the structure of the Gaussian noise (3.34) in (3.32). We present in Section 3.4.2.1 some estimates on the covariance matrix $\Sigma^{(t_0)}$, for $t_0 > 0$ small enough, of the continuous process (3.1) with $b \equiv 0$. Then, we relate in Section 3.4.2.2 the statistics of the Gaussian noise $G^{(k+1)}$ in (3.32) to $\Sigma^{(t_0)}$ for $k \approx t_0/\gamma$. We finally provide a decomposition into a final effective Gaussian increment and independent intermediate increments, by an orthogonal decomposition (see Section 3.4.2.3). More precisely, we explicitly write out a linear transformation \mathbf{A} such that $(\tilde{\mathbf{Z}}^{(k+1)}, G^{(k+1)}) = \mathbf{A}\mathbf{Z}^{(k)}$ with $G^{(k+1)}$ independent of $\tilde{\mathbf{Z}}^{(k+1)}$ and $\tilde{\mathbf{Z}}^{(k+1)}$ a Gaussian vector whose components are i.i.d. d -dimensional standard Gaussian random variables.

3.4.2.1 Properties of the noise covariance of the continuous process

Recall the expression (3.9) of the covariance matrix $\Sigma^{(t)}$ of the Gaussian process corresponding to (3.1) when $b \equiv 0$. Our first technical result provides some bounds on the covariance matrix $\Sigma^{(t_0)}$ associated with the underlying reference Gaussian process over times $t_0 > 0$.

Lemma 6. *There exist $\bar{t}_0 > 0$ and positive constants $\underline{\varrho}_1, \underline{\varrho}_2, \underline{\varrho}_3, \bar{\varrho}_1, \bar{\varrho}_2, \bar{\varrho}_3$ with $\underline{\varrho}_1 \underline{\varrho}_3 - \bar{\varrho}_2^2 > 0$, such that, for any $t_0 \in (0, \bar{t}_0]$,*

$$\underline{\varrho}_1 \leq \frac{\Sigma_1^{(t_0)}}{\sigma^2 t_0^3} \leq \bar{\varrho}_1, \quad \underline{\varrho}_2 \leq \frac{\Sigma_2^{(t_0)}}{\sigma^2 t_0^2} \leq \bar{\varrho}_2, \quad \underline{\varrho}_3 \leq \frac{\Sigma_3^{(t_0)}}{\sigma^2 t_0} \leq \bar{\varrho}_3.$$

The result is an immediate consequence of the limit $\lim_{t_0 \downarrow 0} \Sigma_1^{(t_0)} t_0^{-3} \sigma^2 / 3$ (which can be seen from (3.10) by approximating the integrand in the integral appearing on the right hand side of the first equality), as well as the limits $\lim_{t_0 \downarrow 0} \Sigma_2^{(t_0)} t_0^{-2} = \sigma^2 / 2$ and $\lim_{t_0 \downarrow 0} \Sigma_3^{(t_0)} t_0^{-1} = \sigma^2$ which are obtained in a similar way. A detailed version of the proof is provided in the appendix (see Section 3.6.1).

Lemma 7. *Assume that A1 holds. Then there exist $\bar{\rho}_c, \bar{t}_0 > 0$ such that the following inequality holds in the sense of 2×2 symmetric matrices: for any $t_0 \in (0, \bar{t}_0]$,*

$$t_0 \bar{\rho}_c^{-1} \begin{pmatrix} t_0^2 & 0 \\ 0 & 1 \end{pmatrix} \preceq \Sigma^{(t_0)} \preceq t_0 \bar{\rho}_c \begin{pmatrix} t_0^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Proof. Introduce, for $t_0, \rho_c > 0$,

$$\overline{\mathbf{B}}^{(t_0, \rho_c)} = t_0 \rho_c \begin{pmatrix} t_0^2 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \Sigma_1^{(t_0)} & \Sigma_2^{(t_0)} \\ \Sigma_2^{(t_0)} & \Sigma_3^{(t_0)} \end{pmatrix}, \quad \underline{\mathbf{B}}^{(t_0, \rho_c)} = \begin{pmatrix} \Sigma_1^{(t_0)} & \Sigma_2^{(t_0)} \\ \Sigma_2^{(t_0)} & \Sigma_3^{(t_0)} \end{pmatrix} - t_0 \rho_c^{-1} \begin{pmatrix} t_0^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

The aim is to choose \bar{t}_0 and $\bar{\rho}_c$ such that $\overline{\mathbf{B}}^{(t_0, \bar{\rho}_c)}$ and $\underline{\mathbf{B}}^{(t_0, \bar{\rho}_c)}$ are both positive for any $t \in (0, \bar{t}_0]$. By Sylvester's criterion [67, Theorem 7.2.5], the result of the lemma is implied by the following statement: there exist $\rho_c, \bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$,

$$t_0 \rho_c - \Sigma_3^{(t_0)} > 0, \quad \Sigma_3^{(t_0)} - t_0 \rho_c^{-1} > 0, \quad \det(\overline{\mathbf{B}}^{(t_0, \rho_c)}) > 0, \quad \det(\underline{\mathbf{B}}^{(t_0, \rho_c)}) > 0.$$

The first two conditions are satisfied for the value of $\bar{t}_0 > 0$ given by Lemma 6 and $\rho_c \geq \rho_c^{(1)} = \sigma^2 \max(\bar{\varrho}_3, \underline{\varrho}_3^{-1})$. Moreover, by Lemma 6, there exist $\rho_c^{(2)} > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$,

$$\begin{aligned} \det(\overline{\mathbf{B}}^{(t_0, \rho_c^{(2)})}) &= (t_0^3 \rho_c^{(2)} - \Sigma_1^{(t_0)}) (t_0 \rho_c^{(2)} - \Sigma_3^{(t_0)}) - (\Sigma_2^{(t_0)})^2 \\ &\geq (t_0^3 \rho_c^{(2)} - \bar{\varrho}_1 \sigma^2 t_0^3) (t_0 \rho_c^{(2)} - \bar{\varrho}_3 \sigma^2 t_0) - \bar{\varrho}_2^2 \sigma^4 t_0^4 \\ &= t_0^4 \left[(\rho_c^{(2)} - \bar{\varrho}_1 \sigma^2) (\rho_c^{(2)} - \bar{\varrho}_3 \sigma^2) - \bar{\varrho}_2^2 \sigma^4 \right] > 0. \end{aligned}$$

Similar computations show that there exists $\rho_c^{(3)} > 0$ for which $\det(\underline{\mathbf{B}}^{(t_0, \rho_c^{(3)})}) > 0$ for any $t_0 \in (0, \bar{t}_0]$. This completes the proof upon setting $\bar{\rho}_c = \max(\rho_c^{(1)}, \rho_c^{(2)}, \rho_c^{(3)})$. \square

3.4.2.2 Properties of the final Gaussian noise $\mathbf{G}^{(k+1)}$

We specify in this section the limit of the covariance of the noise $\mathbf{G}^{(k+1)}$ defined in (3.34) as $\gamma \rightarrow 0$ and $k\gamma \approx t_0$. The following result, whose proof can be read in appendix-Section 3.6.2, gathers useful estimates which allow to compare the covariance of $\mathbf{G}^{(k+1)}$ to the covariance $\Sigma^{(t_0)}$ of the underlying continuous process with $b = 0$ (see Lemma 9 below).

Lemma 8. *Assume that A1 holds. Then, for any $\gamma \in (0, \bar{\gamma}]$ and $\ell \geq 1$, $|\tau_\gamma^\ell - e^{-\kappa\gamma^\ell}| \leq C_\kappa \ell \gamma^2$. Moreover, for any $\gamma \in (0, \bar{\gamma}]$,*

$$|\tau_\gamma - 1| \leq (\kappa + C_\kappa \gamma) \gamma, \quad \left| \frac{\gamma}{1 - \tau_\gamma} - \frac{1}{\kappa} \right| \leq \left(\frac{2C_\kappa}{\kappa^2} + 1 \right) \gamma. \quad (3.38)$$

Lemma 9. *The random variable $\mathbf{G}^{(k+1)}$ defined in (3.34) is a 2d-dimensional Gaussian random variable with mean zero and covariance matrix*

$$\mathbf{c}^{(k+1)} = \begin{pmatrix} c_1^{(k+1)} & c_2^{(k+1)} \\ c_2^{(k+1)} & c_3^{(k+1)} \end{pmatrix} \otimes \mathbf{I}_d, \quad (3.39)$$

3.4. Proof of the minorization condition

with

$$\begin{cases} c_1^{(k+1)} = \gamma \|\mathbf{g}_1^{(k+1)}\|^2 = \frac{\gamma^2}{(1-\tau_\gamma)^2} \frac{1}{1+\tau_\gamma} \left[(1+\tau_\gamma)k\gamma \right. \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \left. - \frac{\tau_\gamma(2+\tau_\gamma) - 2\tau_\gamma(1+\tau_\gamma)\tau_\gamma^k + \tau_\gamma^{2(k+1)}}{(1-\tau_\gamma)/\gamma} \right], \\ c_2^{(k+1)} = \gamma \langle \mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \rangle = \frac{\gamma^2}{(1-\tau_\gamma)^2} \frac{\tau_\gamma}{1+\tau_\gamma} \left[1 - (1+\tau_\gamma)\tau_\gamma^k + \tau_\gamma^{2k+1} \right], \\ c_3^{(k+1)} = \gamma \|\mathbf{g}_2^{(k+1)}\|^2 = \frac{\gamma}{1-\tau_\gamma} \frac{1}{1+\tau_\gamma} \left(1 - \tau_\gamma^{2(k+1)} \right). \end{cases} \quad (3.40)$$

where $\mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)}$ are given in (3.31). In addition, when **A1** holds, there exists for any $\bar{t}_0 > 0$ a constant $C_{\bar{t}_0} \geq 0$ such that, for any $i \in \{1, 2, 3\}$ and $\gamma \in (0, \bar{\gamma}]$,

$$\sup_{t_0 \in (0, \bar{t}_0]} \left| c_i^{(\lfloor t_0/\gamma \rfloor + 1)} - \frac{\Sigma_i^{(t_0)}}{\sigma^2} \right| \leq C_{\bar{t}_0} \gamma. \quad (3.41)$$

Proof. Note that since $(Z_i)_{i \in \{1, \dots, k+1\}}$ are i.i.d. zero-mean Gaussian random variables and $G_1^{(k+1)}$ and $G_2^{(k+1)}$ are linear combination of elements of this family, $G^{(k+1)}$ is a $2d$ -dimensional zero-mean Gaussian random variable with covariance matrix

$$\begin{pmatrix} \text{Cov}(G_1^{(k+1)}) & \text{Cov}(G_1^{(k+1)}, G_2^{(k+1)}) \\ \text{Cov}(G_1^{(k+1)}, G_2^{(k+1)}) & \text{Cov}(G_2^{(k+1)}) \end{pmatrix}.$$

By (3.35) and (3.31), straightforward computations give

$$\begin{aligned} \text{Cov}(G_1^{(k+1)}) &= \frac{\gamma^3}{(1-\tau_\gamma)^2} \sum_{i=0}^{k-1} (1-\tau_\gamma^{k-i})^2 \mathbf{I}_d \\ &= \frac{\gamma^3}{(1-\tau_\gamma)^2} \left[k - 2\tau_\gamma \frac{1-\tau_\gamma^k}{1-\tau_\gamma} + \tau_\gamma^2 \frac{1-\tau_\gamma^{2k}}{1-\tau_\gamma^2} \right] \mathbf{I}_d, \\ \text{Cov}(G_1^{(k+1)}, G_2^{(k+1)}) &= \frac{\gamma^2}{1-\tau_\gamma} \sum_{i=0}^{k-1} \left[(1-\tau_\gamma^{k-i}) \tau_\gamma^{k-i} \right] \mathbf{I}_d \\ &= \frac{\gamma^2}{1-\tau_\gamma} \left[\tau_\gamma \frac{1-\tau_\gamma^k}{1-\tau_\gamma} - \tau_\gamma^2 \frac{1-\tau_\gamma^{2k}}{1-\tau_\gamma^2} \right] \mathbf{I}_d, \\ \text{Cov}(G_2^{(k+1)}) &= \gamma \sum_{i=0}^k \tau_\gamma^{2(k-i)} \mathbf{I}_d = \frac{\gamma}{1-\tau_\gamma^2} (1-\tau_\gamma^{2(k+1)}) \mathbf{I}_d, \end{aligned}$$

from which we obtain the expressions of the coefficients $c_i^{(k+1)}$ for $1 \leq i \leq 3$ after some reorganization allowing to compare the resulting expressions more easily with (3.10)-(3.11) (upon replacing γ by $k\gamma$).

We finally show (3.41). By Lemma 8, we have, for any $\bar{t}_0 > 0$ and $\gamma \in (0, \bar{\gamma}]$,

$$\begin{aligned} \sup_{t_0 \in (0, \bar{t}_0]} \left| \tau_\gamma^{\lfloor t_0/\gamma \rfloor + 1} - e^{-\kappa t_0} \right| &\leq C_\kappa (\bar{t}_0 + 1) \gamma, \\ \sup_{t_0 \in (0, \bar{t}_0]} \left| \tau_\gamma^{2\lfloor t_0/\gamma \rfloor + 1} - e^{-2\kappa t_0} \right| &\leq 2C_\kappa (\bar{t}_0 + 1) \gamma. \end{aligned}$$

The inequality (3.41) is then a simple consequence of the definition (3.9) of $\Sigma^{(t_0)}$, (3.40) and the estimates (3.38) in Lemma 8. \square

3.4.2.3 Decomposition into final noise and independent increments

After analyzing the covariance of $G^{(k+1)}$ defined by (3.34) in the limit $\gamma \rightarrow 0$ and $k\gamma \approx t_0$, we aim at providing the conditional distribution of (X_{k+1}, V_{k+1}) given $G^{(k+1)}$. This will be a crucial step to use a perturbation argument similar to the one presented for overdamped Langevin dynamics in Section 3.3. To this end, we introduce in this section the linear transformation $\tilde{\mathbf{Z}}^{(k+1)}$ of $\mathbf{Z}^{(k)}$ which plays the same role as (3.24), i.e., $\tilde{\mathbf{Z}}^{(k+1)}$ is independent of $G^{(k+1)}$ and its components are i.i.d. standard Gaussian random variables. For underdamped Langevin dynamics, we introduce for $k \geq 2$ and $\gamma \in (0, \bar{\gamma}]$ the vectors $\boldsymbol{\alpha}(k+1, \gamma) = (\alpha_1(k+1, \gamma), \dots, \alpha_{k+1}(k+1, \gamma))$ and $\boldsymbol{\beta}(k+1, \gamma) = (\beta_1(k+1, \gamma), \dots, \beta_{k+1}(k+1, \gamma))$, and define, for any $i \in \{1, \dots, k-1\}$,

$$\tilde{Z}_i^{(k+1)} = Z_i - \sqrt{\gamma} \beta_i(k, \gamma) G_2^{(k+1)} - \sqrt{\gamma} \alpha_i(k, \gamma) G_1^{(k+1)}. \quad (3.42)$$

The components k and $k+1$ of the vectors $\boldsymbol{\alpha}(k+1, \gamma), \boldsymbol{\beta}(k+1, \gamma)$ are not needed at this stage, but they will turn out to be useful later on. As specified below in Lemma 12, the scaling factor $\sqrt{\gamma}$ in front of the coefficients $\alpha_i(k, \gamma), \beta_i(k, \gamma)$ ensures that the latter coefficients are of order 1 at most as $\gamma \rightarrow 0$ and $k\gamma \approx t_0$.

Remark 3.7. *In order to further motivate the chosen scalings in (3.42), note that the linear transform (3.24) for overdamped Langevin dynamics can be rewritten as*

$$\tilde{Z}_i^{(k+1)} = Z_i - \sqrt{\gamma} \frac{G^{(k+1)}}{(k+1)\gamma}, \quad i \in \{1\}k.$$

The counterpart of the coefficients $\alpha_i(k+1, \gamma), \beta_i(k+1, \gamma)$ in this context is simply $1/[(k+1)\gamma]$, which is indeed of order 1 when $k\gamma$ is of order 1.

The choice of $\boldsymbol{\alpha}(k+1, \gamma), \boldsymbol{\beta}(k+1, \gamma)$ corresponds to the orthogonal projection of the Gaussian variables $\{Z_{i+1}\}_{i=0}^{k-2}$ onto the orthogonal of the vector space spanned by $\mathbf{g}_1^{(k+1)}$ and $\mathbf{g}_2^{(k+1)}$. This will ensure that the covariance of these projected vectors and $G^{(k+1)}$ vanishes, and hence that the projected vectors $\tilde{\mathbf{Z}}^{(k+1)} = (\tilde{Z}_1^{(k+1)}, \dots, \tilde{Z}_{k-1}^{(k+1)})$ are independent of $G^{(k+1)}$.

We start by constructing the orthogonal projector onto $\text{Span}(\mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)})$. We need to restrict the discussion to iteration indices and time steps in the set

$$\mathbb{E}_c = \left\{ (k, \gamma) \in \mathbb{N}^* \times (0, +\infty) \mid c_1^{(k+1)} c_3^{(k+1)} - \left(c_2^{(k+1)} \right)^2 \neq 0 \right\}, \quad (3.43)$$

3.4. Proof of the minorization condition

where the coefficients $(c_i^{(k+1)})_{1 \leq i \leq 3}$ are defined in Lemma 9. Note that the condition to be satisfied in (3.43) is in fact that the determinant of the matrix $\mathbf{c}^{(k+1)}$ in (3.39) is positive (it is always nonnegative). This condition is not restrictive since, by Lemma 6 and (3.41), there exists $\bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$, there is $\bar{\gamma}_{t_0} > 0$ for which $(\lfloor t_0/\gamma \rfloor, \gamma) \in \mathbf{E}_c$ for any $\gamma \in (0, \bar{\gamma}_{t_0}]$.

Lemma 10. *For any $(k, \gamma) \in \mathbf{E}_c$, and $j \in \{1, \dots, k+1\}$, consider $\beta_j(k+1, \gamma), \alpha_j(k+1, \gamma) \in \mathbb{R}$ the unique solution of*

$$\begin{pmatrix} c_1^{(k+1)} & c_2^{(k+1)} \\ c_2^{(k+1)} & c_3^{(k+1)} \end{pmatrix} \begin{pmatrix} \alpha_j(k+1, \gamma) \\ \beta_j(k+1, \gamma) \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{1,j}^{(k+1)} \\ \mathbf{g}_{2,j}^{(k+1)} \end{pmatrix}, \quad (3.44)$$

and define

$$\mathbf{P}^{(k)} = \mathbf{I}_{k+1} - \gamma \boldsymbol{\beta}(k+1, \gamma) \left[\mathbf{g}_2^{(k+1)} \right]^\top - \gamma \boldsymbol{\alpha}(k+1, \gamma) \left[\mathbf{g}_1^{(k+1)} \right]^\top \in \mathbb{R}^{(k+1) \times (k+1)}. \quad (3.45)$$

Then, $\mathbf{P}^{(k)}$ is the orthogonal projection onto $\text{Span} \left(\mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \right)^\perp$.

Proof. We compute the action of $\mathbf{P}^{(k)}$ on the vector space generated by $\mathbf{g}_1^{(k+1)}$ and $\mathbf{g}_2^{(k+1)}$ and its orthogonal. Note first that the definition (3.45) ensures that $\mathbf{P}^{(k)} w = w$ for all $w \in \text{Span} \left(\mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \right)^\perp$. Moreover, in view of Lemma 9 and by the definition (3.44) of the coefficients of $\boldsymbol{\alpha}(k+1, \gamma), \boldsymbol{\beta}(k+1, \gamma)$,

$$\begin{aligned} \gamma \left\| \mathbf{g}_1^{(k+1)} \right\|^2 \boldsymbol{\alpha} + \gamma \left\langle \mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \right\rangle \boldsymbol{\beta} &= \mathbf{g}_1^{(k+1)}, \\ \gamma \left\langle \mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \right\rangle \boldsymbol{\alpha} + \gamma \left\| \mathbf{g}_2^{(k+1)} \right\|^2 \boldsymbol{\beta} &= \mathbf{g}_2^{(k+1)}, \end{aligned}$$

which implies that $\mathbf{P}^{(k)} \mathbf{g}_1^{(k+1)} = \mathbf{P}^{(k)} \mathbf{g}_2^{(k+1)} = 0$, and finally $\mathbf{P}^{(k)} w = 0$ for $w \in \text{Span} \left(\mathbf{g}_1^{(k+1)}, \mathbf{g}_2^{(k+1)} \right)$. \square

We are now in position to specify the law of $\tilde{\mathbf{Z}}^{(k+1)} = \left(\tilde{Z}_1^{(k+1)}, \dots, \tilde{Z}_{k-1}^{(k+1)} \right)$.

Lemma 11. *Let $(k, \gamma) \in \mathbf{E}_c$. Then, $\tilde{\mathbf{Z}}^{(k+1)}$ defined by (3.42) is a $d \times (k-1)$ -dimensional zero-mean Gaussian random variable with positive definite covariance matrix $\mathbf{S}^{(k)} \left[\mathbf{S}^{(k)} \right]^\top$, where*

$$\mathbf{S}^{(k)} = \left[\left(\mathbf{I}_{k-1} \quad \mathbf{0}_{k-1,2} \right) \mathbf{P}^{(k)} \right] \otimes \mathbf{I}_d \in \mathbb{R}^{(k-1)d \times (k+1)d}. \quad (3.46)$$

In addition, $\tilde{\mathbf{Z}}^{(k+1)}$ is independent of $\mathbf{G}^{(k+1)}$ defined in (3.35).

Proof. It is easy to see that $\tilde{\mathbf{G}}^{(k)} = \left(\tilde{\mathbf{Z}}^{(k+1)}, \mathbf{G}^{(k+1)} \right) = \mathbf{A} \mathbf{Z}^{(k)}$ with

$$\mathbf{A} = \left[\begin{pmatrix} \mathbf{I}_{k-1} & \mathbf{0}_{k-1,2} \\ \mathbf{0}_{2,k-1} & \mathbf{0}_{2,2} \end{pmatrix} \mathbf{P}^{(k)} + \sqrt{\gamma} \begin{pmatrix} \mathbf{0}_{k-1,k+1} \\ \left[\mathbf{g}_1^{(k+1)} \right]^\top \\ \left[\mathbf{g}_2^{(k+1)} \right]^\top \end{pmatrix} \right] \otimes \mathbf{I}_d.$$

The matrix \mathbf{A} is invertible by Gaussian elimination in view of the definition (3.45) of $\mathbf{P}^{(k)}$ and the ones (3.31) of $\mathbf{g}_1^{(k+1)}$, $\mathbf{g}_2^{(k+1)}$, and

$$\begin{aligned} \det(\mathbf{A})^{1/d} &= \det \left[\begin{pmatrix} \mathbf{I}_{k-1} & \mathbf{0}_{k-1,2} \\ \mathbf{0}_{2,k-1} & \mathbf{0}_{2,2} \end{pmatrix} \mathbf{P}^{(k)} + \sqrt{\gamma} \begin{pmatrix} \mathbf{0}_{k-1,k+1} \\ [\mathbf{g}_1^{(k+1)}]^\top \\ [\mathbf{g}_2^{(k+1)}]^\top \end{pmatrix} \right] = \det \begin{pmatrix} \mathbf{I}_{k-1} & \mathbf{0}_{k-1,2} \\ \sqrt{\gamma} [\mathbf{g}_1^{(k+1)}]^\top \\ \sqrt{\gamma} [\mathbf{g}_2^{(k+1)}]^\top \end{pmatrix} \\ &= \gamma \det \begin{pmatrix} \mathbf{g}_{1,k}^{(k+1)} & \mathbf{g}_{1,k+1}^{(k+1)} \\ \mathbf{g}_{2,k}^{(k+1)} & \mathbf{g}_{2,k+1}^{(k+1)} \end{pmatrix} > 0. \end{aligned}$$

Therefore, $\tilde{\mathbf{G}}^{(k)}$ is a zero-mean Gaussian random variable with positive definite covariance matrix $\mathbf{A}\mathbf{A}^\top$. As a result, $\tilde{\mathbf{Z}}^{(k+1)}$ defined by (3.42) is a $d \times (k-1)$ -dimensional zero-mean Gaussian random variable with positive definite covariance matrix $\mathbf{S}^{(k)}[\mathbf{S}^{(k)}]^\top$ with $\mathbf{S}^{(k)}$ given by (3.46). Finally, since $\mathbf{P}^{(k)}\mathbf{g}_1^{(k+1)} = \mathbf{P}^{(k)}\mathbf{g}_2^{(k+1)} = 0$ in view of (3.45), it follows that for $i \in \{1, \dots, k-1\}$, $\text{Cov}(\tilde{Z}_i^{(k)}, G_1^{(k)}) = 0$, $\text{Cov}(\tilde{Z}_i^{(k)}, G_2^{(k)}) = 0$ and so $\tilde{\mathbf{Z}}^{(k+1)}$ is independent of $\mathbf{G}^{(k+1)}$ because all random variables at hand are Gaussian. \square

We conclude this section by providing some bounds of the coefficients of $\alpha(k+1, \gamma)$, $\beta(k+1, \gamma)$ defined in (3.44). We denote by $\|\xi\|_\infty = \max(|\xi_1|, \dots, |\xi_k|)$ the ℓ^∞ norm of a vector $\xi = (\xi_1, \dots, \xi_k) \in \mathbb{R}^k$.

Lemma 12. *Assume that A1 holds. Then, there exists $\bar{t}_0 > 0$ and $K \in \mathbb{R}_+$ such that, for any $t_0 \in (0, \bar{t}_0]$,*

$$\limsup_{\gamma \downarrow 0} \|\alpha(\lfloor t_0/\gamma \rfloor + 1, \gamma)\|_\infty \leq K/t_0^2, \quad \limsup_{\gamma \downarrow 0} \|\beta(\lfloor t_0/\gamma \rfloor + 1, \gamma)\|_\infty \leq K/t_0.$$

Proof. We consider $\bar{t}_0 > 0$ as given by Lemmas 6 and 7, and such that, for any $t_0 \in (0, \bar{t}_0]$, there is $\bar{\gamma}_{t_0}$ for which $(\lfloor t_0/\gamma \rfloor, \gamma) \in \mathbf{E}_c$ for any $\gamma \in (0, \bar{\gamma}_{t_0}]$ (see the discussion before Lemma 10). For such values of \bar{t}_0 and γ , the system (3.44) for $k_0 = \lfloor t_0/\gamma \rfloor$ can be explicitly solved as

$$\begin{aligned} \alpha_j(k_0 + 1, \gamma) &= \frac{c_3^{(k_0+1)} \mathbf{g}_{1,j}^{(k_0+1)} - c_2^{(k_0+1)} \mathbf{g}_{2,j}^{(k_0+1)}}{c_1^{(k_0+1)} c_3^{(k_0+1)} - (c_2^{(k_0+1)})^2}, \\ \beta_j(k_0 + 1, \gamma) &= \frac{c_1^{(k_0+1)} \mathbf{g}_{2,j}^{(k_0+1)} - c_2^{(k_0+1)} \mathbf{g}_{1,j}^{(k_0+1)}}{c_1^{(k_0+1)} c_3^{(k_0+1)} - (c_2^{(k_0+1)})^2}. \end{aligned}$$

By (3.37) and Lemma 9, it follows that, for any $t_0 \in (0, \bar{t}_0]$,

$$\limsup_{\gamma \downarrow 0} |\alpha_j(k_0 + 1, \gamma)| \leq \limsup_{\gamma \downarrow 0} \frac{t_0 \left(|c_3^{(k_0+1)}| + |c_2^{(k_0+1)}| \right)}{\left| c_1^{(k_0+1)} c_3^{(k_0+1)} - (c_2^{(k_0+1)})^2 \right|} = \frac{\sigma^2 \left(t_0 \Sigma_3^{(t_0)} + \Sigma_2^{(t_0)} \right)}{\left| \Sigma_1^{(t_0)} \Sigma_3^{(t_0)} - (\Sigma_2^{(t_0)})^2 \right|},$$

3.4. Proof of the minorization condition

and

$$\limsup_{\gamma \downarrow 0} |\beta_j(k_0 + 1, \gamma)| \leq \limsup_{\gamma \downarrow 0} \frac{|c_1^{(k_0+1)}| + t_0 |c_2^{(k_0+1)}|}{\left| c_1^{(k_0+1)} c_3^{(k_0+1)} - \left(c_2^{(k_0+1)} \right)^2 \right|} = \frac{\sigma^2 \left(\Sigma_1^{(t_0)} + t_0 \Sigma_2^{(t_0)} \right)}{\left| \Sigma_1^{(t_0)} \Sigma_3^{(t_0)} - \left(\Sigma_2^{(t_0)} \right)^2 \right|}.$$

The numerators of the last term in the two previous equalities are upper bounded by Lemma 6 as $t_0 \Sigma_3^{(t_0)} + \Sigma_2^{(t_0)} \leq (\bar{\rho}_2 + \bar{\rho}_3) \sigma^2 t_0^2$ and $\Sigma_1^{(t_0)} + t_0 \Sigma_2^{(t_0)} \leq (\bar{\rho}_1 + \bar{\rho}_2) \sigma^2 t_0^3$; while the denominator is lower bounded with the Minkowski determinant theorem (see e.g. [92, Section 4.1.8]) and Lemma 7 as

$$\left| \Sigma_1^{(t_0)} \Sigma_3^{(t_0)} - \left(\Sigma_2^{(t_0)} \right)^2 \right| \geq \det \begin{bmatrix} t_0 & \left(t_0^2 & 0 \right) \\ \bar{\rho}_c & 0 & 1 \end{bmatrix} = \frac{t_0^4}{\bar{\rho}_c^2}.$$

This finally gives the claimed estimates. \square

3.4.3 Stability estimates

We provide in this section estimates on the mapping $\Gamma_\gamma^{(k+1)}$ defined in (3.30) providing the $(k+1)$ -th iterates (X_{k+1}, V_{k+1}) of the Markov chain defined in (3.2) as a function of the initial condition (x, v) and the realizations $\{z_{i+1}, w_{i+1}\}_{i=0}^k$. We first establish in Section 3.4.3.1 that this function is Lipschitz continuous with respect to the initial condition (x, v) and the realizations $\{z_{i+1}\}_{i=0}^k$, and provide upper bounds on its Lipschitz constant. Pursuing on the same argument as Section 3.3 (see (3.25)), we then show in Section 3.4.3.2 that (X_{k+1}, V_{k+1}) can be written as

$$(X_{k+1}, V_{k+1}) = \bar{\Gamma}_{\gamma, x, v, \{\tilde{Z}_j^{(k+1)}\}_{j=1}^{k-1}, \{W_j\}_{j=1}^{k+1}}^{(k+1)} \left(G_1^{(k+1)}, G_2^{(k+1)} \right),$$

with $(G_1^{(k+1)}, G_2^{(k+1)})$ and $\{\tilde{Z}_j^{(k+1)}\}_{j=1}^{k-1}$ respectively defined in (3.35) and (3.42). We establish Lipschitz properties of the function $(g_1, g_2) \mapsto \bar{\Gamma}_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g_1, g_2)$. Then, we prove in Section 3.4.3.3 that this function is in fact a C^1 -diffeomorphism. We conclude this section by studying some properties of its inverse seen as a function of (g_1, g_2) and the initial conditions (x, v) .

3.4.3.1 Stability with respect to initial conditions and all noise increments

The following result provides some Lipschitz bounds for the functions $\Gamma_\gamma^{(k+1)}$ defined in (3.30). They are stated in terms of a norm on \mathbb{R}^{2d} parametrized by some positive parameter $\lambda > 0$, namely $\|x\| + \lambda \|v\|$. The proof of Proposition 3.8 below will require λ to be chosen sufficiently small.

Lemma 13. *Assume that A2 holds and $\sup_{\gamma \in (0, \bar{\gamma})} \tau_\gamma \leq 1$ and fix $\lambda > 0$, a time step $\gamma > 0$, a maximal number of iterations $N \in \mathbb{N}^*$, and realizations $\{w_j\}_{j=1}^{N+1} \subset \mathbb{R}^m$ of the*

additional noise. For two initial conditions $(x, v), (x', v') \in \mathbb{R}^{2d}$ and realizations of the noise $\{z_j\}_{j=1}^{N+1}, \{z'_j\}_{j=1}^{N+1} \subset \mathbb{R}^d$, define for any $k \in \{0, \dots, N-1\}$ the iterates

$$\begin{aligned} (x_{k+1}, v_{k+1}) &= \Gamma_\gamma^{(k+1)} \left(x, v, \{z_j\}_{j=1}^{k+1}, \{w_j\}_{j=1}^{k+1} \right), \\ (x'_{k+1}, v'_{k+1}) &= \Gamma_\gamma^{(k+1)} \left(x', v', \{z'_j\}_{j=1}^{k+1}, \{w_j\}_{j=1}^{k+1} \right). \end{aligned}$$

Then, for any $k \in \{0, \dots, N\}$,

$$\begin{aligned} \|x_k - x'_k\| + \lambda \|v_k - v'_k\| &\leq L_{\Gamma, \gamma}(\lambda)^k \left(\|x - x'\| + \lambda \|v - v'\| \right) \\ &\quad + M_{\Gamma, \gamma}(\lambda) \sum_{i=1}^k L_{\Gamma, \gamma}(\lambda)^{k-i} \|z_i - z'_i\|, \end{aligned} \quad (3.47)$$

where

$$L_{\Gamma, \gamma}(\lambda) = 1 + \gamma \left[\frac{1}{\lambda} + (1 + \lambda) \max \left(1, \frac{\gamma^\delta}{\lambda} \right) \mathbf{L} \right], \quad M_{\Gamma, \gamma}(\lambda) = \lambda + \gamma^\delta \mathcal{D} + \gamma^{1+\delta} (1 + \lambda) \mathbf{L}. \quad (3.48)$$

Moreover, when the initial conditions coincide (namely when $(x, v) = (x', v')$),

$$\sum_{i=1}^k \|x_i - x'_i\| \leq \gamma^\delta (\mathcal{D} + \gamma \mathbf{L}) \|z_k - z'_k\| + \{k M_{\Gamma, \gamma}(\lambda) L_{\Gamma, \gamma}(\lambda)^k + \mathcal{L}_{k, \gamma, \lambda}^x\} \sum_{i=1}^{k-1} \|z_i - z'_i\|, \quad (3.49)$$

and

$$\sum_{i=1}^k \|v_i - v'_i\| \leq \left(1 + \gamma^{1+\delta} \mathbf{L} \right)^k \|z_k - z'_k\| + k \mathcal{L}_{k, \gamma, \lambda}^v \sum_{i=1}^{k-1} \|z_i - z'_i\|, \quad (3.50)$$

with

$$\begin{aligned} \mathcal{L}_{k, \gamma, \lambda}^x &= \gamma \mathbf{L} M_{\Gamma, \gamma}(\lambda) L_{\Gamma, \gamma}(\lambda)^k \\ &\quad + \gamma \left(1 + \gamma^\delta \mathbf{L} \right) \left(k \gamma M_{\Gamma, \gamma}(\lambda) \mathbf{L} \left(1 + \gamma^{1+\delta} \mathbf{L} \right)^k L_{\Gamma, \gamma}(\lambda)^k + \left(1 + \gamma^{1+\delta} \mathbf{L} \right)^k \right), \\ \mathcal{L}_{k, \gamma, \lambda}^v &= (k-1) \gamma M_{\Gamma, \gamma}(\lambda) \left(1 + \gamma^{1+\delta} \mathbf{L} \right)^k L_{\Gamma, \gamma}(\lambda)^k \mathbf{L} + \left(1 + \gamma^{1+\delta} \mathbf{L} \right)^k. \end{aligned} \quad (3.51)$$

For reasons that will appear more clearly in Lemma 14 and its proof, the difference $\|z_k - z'_k\|$ is isolated on purpose in (3.49)-(3.50) and appears with a prefactor at most 1 instead of k in contrast to the other differences $\|z_i - z'_i\|$ for $i \in \{1, \dots, k-1\}$.

Proof. Since $(x_{k+1}, v_{k+1}) = \Gamma_\gamma(x_k, v_k, z_{k+1}, w_{k+1})$ for $k \in \{0\} N-1$, we obtain by **A2**

3.4. Proof of the minorization condition

and (3.13) that

$$\begin{aligned}
& \|x_{k+1} - x'_{k+1}\| \\
& \leq \|x_k - x'_k\| + \gamma \|v_k - v'_k\| + \gamma^\delta \|\mathbf{D}_\gamma(z_{k+1} - z'_{k+1})\| \\
& \quad + \gamma \left\| f_\gamma(x_k, \gamma^\delta v_k, \gamma^\delta z_{k+1}, w_{k+1}) - f_\gamma(x'_k, \gamma^\delta v'_k, \gamma^\delta z'_{k+1}, w_{k+1}) \right\| \\
& \leq (1 + \gamma\mathbf{L}) \|x_k - x'_k\| + \gamma (1 + \gamma^\delta \mathbf{L}) \|v_k - v'_k\| + \gamma^\delta (\mathcal{D} + \gamma\mathbf{L}) \|z_{k+1} - z'_{k+1}\|,
\end{aligned} \tag{3.52}$$

and, since $\tau_\gamma \leq 1$,

$$\begin{aligned}
& \|v_{k+1} - v'_{k+1}\| \leq \tau_\gamma \|v_k - v'_k\| + \|z_{k+1} - z'_{k+1}\| \\
& \quad + \gamma \left\| g_\gamma(x_k, \gamma^\delta v_k, \gamma^\delta z_{k+1}, w_{k+1}) - g_\gamma(x'_k, \gamma^\delta v'_k, \gamma^\delta z'_{k+1}, w_{k+1}) \right\| \\
& \leq \gamma\mathbf{L} \|x_k - x'_k\| + (1 + \gamma^{1+\delta}\mathbf{L}) \|v_k - v'_k\| + (1 + \gamma^{1+\delta}\mathbf{L}) \|z_{k+1} - z'_{k+1}\|.
\end{aligned} \tag{3.53}$$

Therefore,

$$\begin{aligned}
& \|x_{k+1} - x'_{k+1}\| + \lambda \|v_{k+1} - v'_{k+1}\| \\
& \leq (1 + \gamma(1 + \lambda)\mathbf{L}) \|x_k - x'_k\| + \left(\lambda [1 + \gamma^{1+\delta}\mathbf{L}] + \gamma [1 + \gamma^\delta\mathbf{L}] \right) \|v_k - v'_k\| \\
& \quad + \mathbf{M}_{\Gamma, \gamma}(\lambda) \|z_{k+1} - z'_{k+1}\| \\
& \leq \left[1 + \gamma \max \left((1 + \lambda)\mathbf{L}, \frac{1}{\lambda} + \left(1 + \frac{1}{\lambda}\right) \gamma^\delta \mathbf{L} \right) \right] (\|x_k - x'_k\| + \lambda \|v_k - v'_k\|) \\
& \quad + \mathbf{M}_{\Gamma, \gamma}(\lambda) \|z_{k+1} - z'_{k+1}\|.
\end{aligned}$$

The bound (3.47) then follows from the inequality $1 + \gamma \max \left((1 + \lambda)\mathbf{L}, \lambda^{-1} + (1 + \lambda^{-1}) \gamma^\delta \mathbf{L} \right) \leq \mathbf{L}_{\Gamma, \gamma}(\lambda)$, and an easy induction on k .

We now prove (3.49). For the sum up to indices $k - 1$, we use (3.47) and $x_0 = x'_0, v_0 = v'_0$ to write

$$\begin{aligned}
\sum_{i=1}^{k-1} \|x_i - x'_i\| & \leq \sum_{i=1}^{k-1} \{ \|x_i - x'_i\| + \lambda \|v_i - v'_i\| \} \\
& \leq \mathbf{M}_{\Gamma, \gamma}(\lambda) \sum_{i=1}^{k-1} \sum_{j=1}^i \mathbf{L}_{\Gamma, \gamma}(\lambda)^{i-j} \|z_j - z'_j\| \\
& \leq \mathbf{M}_{\Gamma, \gamma}(\lambda) (k-1) \mathbf{L}_{\Gamma, \gamma}(\lambda)^k \sum_{i=1}^{k-1} \|z_i - z'_i\|.
\end{aligned} \tag{3.54}$$

The last term in the sum is bounded with (3.52) and (3.47) as

$$\begin{aligned}
 \|x_k - x'_k\| &\leq (1 + \gamma\mathbf{L}) \|x_{k-1} - x'_{k-1}\| + \gamma \left(1 + \gamma^\delta \mathbf{L}\right) \|v_{k-1} - v'_{k-1}\| \\
 &\quad + \gamma^\delta (\mathcal{D} + \gamma\mathbf{L}) \|z_k - z'_k\| \\
 &\leq \gamma \left(1 + \gamma^\delta \mathbf{L}\right) \|v_{k-1} - v'_{k-1}\| + \gamma^\delta (\mathcal{D} + \gamma\mathbf{L}) \|z_k - z'_k\| \\
 &\quad + (1 + \gamma\mathbf{L}) M_{\Gamma, \gamma}(\lambda) L_{\Gamma, \gamma}(\lambda)^k \sum_{i=1}^{k-1} \|z_i - z'_i\|. \tag{3.55}
 \end{aligned}$$

It remains to bound $\|v_{k-1} - v'_{k-1}\|$ on the right-hand side. By an easy induction based on (3.53) and since $x_0 = x'_0$, $v_0 = v'_0$, we obtain, for $\ell \in \{0, \dots, N-1\}$,

$$\|v_\ell - v'_\ell\| \leq \gamma\mathbf{L} \sum_{i=1}^{\ell-1} \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^{\ell-1-i} \|x_i - x'_i\| + \sum_{i=1}^{\ell} \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^{\ell-i+1} \|z_i - z'_i\|. \tag{3.56}$$

In particular, using the previous inequality for $\ell = k-1$ and (3.54), we get

$$\begin{aligned}
 \|v_{k-1} - v'_{k-1}\| &\leq \gamma\mathbf{L} \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k \sum_{i=1}^{k-1} \|x_i - x'_i\| + \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k \sum_{i=1}^{k-1} \|z_i - z'_i\| \\
 &\leq \left(k\gamma M_{\Gamma, \gamma}(\lambda) \mathbf{L} \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k L_{\Gamma, \gamma}(\lambda)^k + \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k\right) \sum_{i=1}^{k-1} \|z_i - z'_i\|,
 \end{aligned}$$

which, plugged in the estimate (3.55) for $\|x_k - x'_k\|$ and combined with (3.54), leads to (3.49).

To obtain (3.50), we use first (3.56) and isolate the k -th term in the sum to write

$$\begin{aligned}
 \sum_{i=1}^k \|v_i - v'_i\| &\leq k\gamma\mathbf{L} \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k \sum_{i=1}^{k-1} \|x_i - x'_i\| \\
 &\quad + \left(1 + \gamma^{1+\delta}\mathbf{L}\right)^k \left[\|z_k - z'_k\| + k \sum_{i=1}^{k-1} \|z_i - z'_i\| \right],
 \end{aligned}$$

and then conclude with (3.54). \square

3.4.3.2 Stability of maps conditioned by the total noise

In order to express (X_{k+1}, V_{k+1}) in terms of $\mathbf{G}^{(k+1)}$ and $\tilde{\mathbf{Z}}^{(k+1)}$ defined in (3.34) and (3.42), we perform a linear change of variables in the functions $\Gamma_\gamma^{(i)}$ defined in (3.30), for $i \in \{0, \dots, k+1\}$. More precisely, we aim at writing, for $i \in \{0, \dots, k+1\}$,

$$(X_i, V_i) = \tilde{\Gamma}_\gamma^{(i, k+1)} \left(X_0, V_0, \left\{ \sqrt{\gamma} \sigma_\gamma \tilde{Z}_j^{(k+1)} \right\}_{j=1}^{i \wedge (k-1)}, \{W_j\}_{j=1}^i, \sigma_\gamma \mathbf{G}^{(k+1)} \right). \tag{3.57}$$

3.4. Proof of the minorization condition

To this end, we define, for a given $(k, \gamma) \in \mathbf{E}_c$, the following functions, which take as arguments the initial condition $(x, v) \in \mathbb{R}^{2d}$, the realizations $(g_1, g_2) \in \mathbb{R}^{2d}$, $\{\tilde{z}_j\}_{j=1}^{k-1} \in \mathbb{R}^{d \times (k-1)}$ and $\{w_j\}_{j=1}^{k+1} \in \mathbf{W}^{k+1}$: for $i \in \{0, \dots, k+1\}$,

$$\tilde{\Gamma}_\gamma^{(i, k+1)} \left(x, v, \{\tilde{z}_j\}_{j=1}^{i \wedge (k-1)}, \{w_i\}_{j=1}^i, (g_1, g_2) \right) = \Gamma_\gamma^{(i)} \left(x, v, \{z_i\}_{j=1}^i, \{w_i\}_{j=1}^i \right), \quad (3.58)$$

where

$$z_i = \tilde{z}_i + \gamma \alpha_i(k, \gamma) g_1 + \gamma \beta_i(k, \gamma) g_2, \quad i \in \{1\} k - 1, \quad (3.59)$$

$$z_k = \frac{1}{\gamma} \left(g_1 - \sum_{i=0}^{k-2} \mathbf{g}_{1, i+1}^{(k+1)} z_{i+1} \right), \quad z_{k+1} = g_2 - \sum_{i=0}^{k-1} \mathbf{g}_{2, i+1}^{(k+1)} z_{i+1}. \quad (3.60)$$

Note that (3.57) holds in view of (3.42) and (3.35) (mind in particular the change of scaling in (3.59) compared to (3.42), which is due to the fact that \tilde{z}_j correspond to the increments $\sqrt{\gamma} \tilde{Z}_j^{(k)}$ in (3.57)). The final iterate can then be expressed in terms of the initial condition (X_0, V_0) and the noise $\mathbf{G}^{(k+1)}$ by reformulating (3.32) as

$$(X_{k+1}, V_{k+1}) = \bar{\Gamma}_{\gamma, \sqrt{\gamma} \sigma_\gamma \tilde{\mathbf{Z}}^{(k+1)}, \mathbf{W}^{(k+1)}}^{(k+1)} \left((X_0, V_0), \sigma_\gamma \mathbf{G}^{(k+1)} \right),$$

where $\mathbf{W}^{(k+1)} = (W_1, \dots, W_{k+1})$, and

$$\begin{aligned} & \bar{\Gamma}_{\gamma, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)} \left((x, v), (g_1, g_2) \right) \\ &= \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} + \gamma^\delta \begin{pmatrix} \mathbf{D}_\gamma \tilde{\Lambda}_{\{\tilde{z}_j\}_{j=1}^{k-1}}^{(k+1)}(g_1, g_2) \\ \mathbf{0}_d \end{pmatrix} + \Xi_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g_1, g_2), \end{aligned} \quad (3.61)$$

with

$$\begin{aligned} & \Xi_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g_1, g_2) = \mathbf{M}_\gamma^{k+1} \begin{pmatrix} x \\ v \end{pmatrix} \\ &+ \gamma \sum_{i=0}^k \tilde{\Theta}^{(i)} \left(\tilde{\Gamma}_\gamma^{(i, k+1)} \left(x, v, \{\tilde{z}_j\}_{j=1}^{i \wedge (k-1)}, \{w_j\}_{j=1}^i, (g_1, g_2) \right), \{\tilde{z}_i\}_{j=1}^{(i+1) \wedge (k-1)}, (g_1, g_2), w_{i+1} \right), \end{aligned} \quad (3.62)$$

\mathbf{M}_γ^{k+1} is given by (3.33), and, setting $\{z_i\}_{i=1}^{k+1}$ as in (3.59)-(3.60),

$$\tilde{\Theta}^{(i)} \left(x, v, \{\tilde{z}_j\}_{j=1}^{(i+1) \wedge (k-1)}, (g_1, g_2), w_{i+1} \right) = \Theta^{(i)}(x, v, z_{i+1}, w_{i+1}), \quad (3.63)$$

and

$$\tilde{\Lambda}_{\{\tilde{z}_j\}_{j=1}^{k-1}}^{(k+1)}(g_1, g_2) = \sum_{j=1}^{k+1} z_j. \quad (3.64)$$

We give in the next proposition key estimates which allow to prove that $\bar{\Gamma}_{\gamma,x,v,\{\tilde{z}_j\}_{j=1}^{k-1},\{w_j\}_{j=1}^{k+1}}^{(k+1)}$ is a C^1 -diffeomorphism on \mathbb{R}^{2d} (see Proposition 3.8 below). To state the result, we define, for any $(k, \gamma) \in \mathbf{E}_c$,

$$m(k, \gamma) = \max (\|\boldsymbol{\alpha}(k+1, \gamma)\|_\infty, \|\boldsymbol{\beta}(k+1, \gamma)\|_\infty),$$

where $\boldsymbol{\alpha}(k+1, \gamma), \boldsymbol{\beta}(k+1, \gamma)$ are defined by (3.44).

Lemma 14. *Assume that A2 holds and $\sup_{\gamma \in (0, \bar{\gamma}]} \tau_\gamma \leq 1$. For any parameter $\lambda > 0$, $(k, \gamma) \in \mathbf{E}_c$, initial condition $(x, v) \in \mathbb{R}^{2d}$, realizations $\bar{\mathbf{z}} = \{\tilde{z}_j\}_{j=1}^{k-1} \subset \mathbb{R}^d$ and $\mathbf{w} = \{w_j\}_{j=1}^{k+1} \in \mathbf{W}^{k+1}$, as well as $(g_1, g_2), (g'_1, g'_2) \in \mathbb{R}^{2d}$, it holds*

$$\begin{aligned} & \left\| \tilde{\Lambda}_{\bar{\mathbf{z}}}^{(k+1)}(g_1, g_2) - \tilde{\Lambda}_{\bar{\mathbf{z}}}^{(k+1)}(g'_1, g'_2) \right\| \\ & \leq \left(1 + m(k, \gamma)k\gamma + \frac{1 - \tau_\gamma}{\gamma} \left[1 + m(k, \gamma)(k\gamma)^2 \right] \right) (\|g_1 - g'_1\| + \|g_2 - g'_2\|), \end{aligned} \quad (3.65)$$

and

$$\left\| \Xi_{\gamma,x,v,\bar{\mathbf{z}},\mathbf{w}}^{(k+1)}(g_1, g_2) - \Xi_{\gamma,x,v,\bar{\mathbf{z}},\mathbf{w}}^{(k+1)}(g'_1, g'_2) \right\| \leq (2 + k\gamma)\text{LL}_{\Xi,k+1,\gamma,\lambda} (\|g_1 - g'_1\| + \|g_2 - g'_2\|), \quad (3.66)$$

with

$$\begin{aligned} \text{L}_{\Xi,k+1,\gamma,\lambda} &= \gamma^\delta \left(1 + (k\gamma)^2 m(k, \gamma) \right) \left(2 + \mathcal{D} + \left(1 + \gamma^\delta \text{L} \right)^k + \gamma \text{L} \right) + \gamma^{1+\delta} \left(1 + k\gamma m(k, \gamma) \right) \\ & \quad + k\gamma^2 m(k, \gamma) \left[k\text{M}_{\Gamma,\gamma}(\lambda) \text{L}_{\Gamma,\gamma}(\lambda)^k + \mathcal{L}_{k,\gamma,\lambda}^x + \gamma^\delta \left(1 + k\mathcal{L}_{k,\gamma,\lambda}^v \right) \right], \end{aligned}$$

where $\text{M}_{\Gamma,\gamma}(\lambda), \text{L}_{\Gamma,\gamma}(\lambda)$ and $\mathcal{L}_{k,\gamma,\lambda}^x, \mathcal{L}_{k,\gamma,\lambda}^v$ are defined in (3.48) and (3.51), respectively.

Proof. Fix $\lambda > 0$, $(k, \gamma) \in \mathbf{E}_c$, $(x, v) \in \mathbb{R}^{2d}$, $\bar{\mathbf{z}} = \{\tilde{z}_j\}_{j=1}^{k-1} \subset \mathbb{R}^d$, $\mathbf{w} = \{w_j\}_{j=1}^{k+1} \in \mathbf{W}^{k+1}$, and $(g_1, g_2), (g'_1, g'_2) \in \mathbb{R}^{2d}$. We introduce for $i \in \{0\}k$, $(x_i, v_i) = \tilde{\Gamma}_\gamma^{(i,k+1)}(x, v, \{\tilde{z}_j\}_{j=1}^{i \wedge (k-1)}, \{w_j\}_{j=1}^i, (g_1, g_2))$, and similarly $(x'_i, v'_i) = \tilde{\Gamma}_\gamma^{(i,k+1)}(x, v, \{\tilde{z}_j\}_{j=1}^{i \wedge (k-1)}, \{w_j\}_{j=1}^i, (g'_1, g'_2))$. With this notation and recalling (3.30), we obtain by the definition (3.58) of $\tilde{\Gamma}_\gamma^{(i,k)}$ that

$$(x_i, v_i) = \Gamma_\gamma^{(i)} \left(x, v, \{z_j\}_{j=1}^i, \{w_j\}_{j=1}^i \right), \quad (x'_i, v'_i) = \Gamma_\gamma^{(i)} \left(x, v, \{z'_j\}_{j=1}^i, \{w_j\}_{j=1}^i \right),$$

where $\{z_i\}_{i=1}^{k+1}$ and $\{z'_i\}_{i=1}^{k+1}$ are the sequences defined in (3.59) based on (g_1, g_2) and (g'_1, g'_2) , respectively. We start by estimating the difference $\|z_i - z'_i\|$ for $i \in \{1\}k + 1$. First, by (3.59), for any $i \in \{1, \dots, k-1\}$,

$$\|z_i - z'_i\| \leq \gamma m(k, \gamma) (\|g_1 - g'_1\| + \|g_2 - g'_2\|). \quad (3.67)$$

3.4. Proof of the minorization condition

Next, by (3.60) and (3.37) we get

$$\begin{aligned}
\|z_k - z'_k\| &\leq \frac{1}{\gamma} \|g_1 - g'_1\| + k \sum_{i=1}^{k-1} \|z_i - z'_i\| \\
&\leq \frac{1 + (k\gamma)^2 m(k, \gamma)}{\gamma} (\|g_1 - g'_1\| + \|g_2 - g'_2\|), \\
\|z_{k+1} - z'_{k+1}\| &\leq \|g_2 - g'_2\| + \sum_{i=1}^k \|z_i - z'_i\| \\
&\leq \|z_k - z'_k\| + [1 + k\gamma m(k, \gamma)] (\|g_1 - g'_1\| + \|g_2 - g'_2\|).
\end{aligned} \tag{3.68}$$

We are now in position to prove (3.65). By (3.64) and (3.31) and since $g_2 = \sum_{i=0}^k \mathbf{g}_{2,i+1}^{(k+1)} z_{i+1}$ by (3.60),

$$\begin{aligned}
\left\| \tilde{\Lambda}_{\{\tilde{z}_j\}_{j=1}^{k-1}}^{(k+1)}(g_1, g_2) - \tilde{\Lambda}_{\{\tilde{z}'_j\}_{j=1}^{k-1}}^{(k+1)}(g'_1, g'_2) \right\| &\leq \|g_2 - g'_2\| + \sum_{i=0}^{k-1} \left(1 - \mathbf{g}_{2,i+1}^{(k+1)}\right) \|z_{i+1} - z'_{i+1}\| \\
&\leq \|g_2 - g'_2\| + (1 - \tau_\gamma) \|z_k - z'_k\| + \sum_{i=0}^{k-2} \|z_{i+1} - z'_{i+1}\| \\
&\leq \|g_2 - g'_2\| + (1 - \tau_\gamma) \|z_k - z'_k\| + m(k, \gamma) k\gamma (\|g_1 - g'_1\| + \|g_2 - g'_2\|).
\end{aligned}$$

The estimate (3.65) follows by combining the last inequality with (3.68).

We next turn to the proof of (3.66). First, by (3.62)-(3.63) and the notation introduced at the beginning of the proof,

$$\begin{aligned}
&\left\| \Xi_{\gamma, x, v, \tilde{z}, \mathbf{w}}^{(k+1)}(g_1, g_2) - \Xi_{\gamma, x, v, \tilde{z}, \mathbf{w}}^{(k+1)}(g'_1, g'_2) \right\| \\
&\leq \gamma \sum_{i=0}^k \left\| \Theta^{(i)}(x_i, v_i, z_{i+1}, w_{i+1}) - \Theta^{(i)}(x'_i, v'_i, z'_{i+1}, w_{i+1}) \right\| \\
&\leq (2 + k\gamma) L\gamma \left(\sum_{i=1}^k \|x_i - x'_i\| + \gamma^\delta \sum_{i=1}^k \|v_i - v'_i\| + \gamma^\delta \sum_{i=0}^k \|z_{i+1} - z'_{i+1}\| \right),
\end{aligned}$$

where we used (3.36) for the last inequality, and the fact that $(x_0, v_0) = (x'_0, v'_0)$ to eliminate the term $i = 0$ in the first two sums on the right hand side of the last inequality. Therefore, in view of (3.49) and (3.50) in Lemma 13,

$$\begin{aligned}
&\left\| \Xi_{\gamma, x, v, \tilde{z}, \mathbf{w}}^{(k+1)}(g_1, g_2) - \Xi_{\gamma, x, v, \tilde{z}, \mathbf{w}}^{(k+1)}(g'_1, g'_2) \right\| \\
&\leq (2 + k\gamma) L\gamma \left(\gamma^\delta \|z_{k+1} - z'_{k+1}\| + \gamma^\delta \mathcal{H}_{k, \gamma} \|z_k - z'_k\| + \mathcal{B}_{k, \gamma, \lambda} \sum_{i=1}^{k-1} \|z_i - z'_i\| \right),
\end{aligned}$$

with $\mathcal{H}_{k, \gamma} = 1 + \mathcal{D} + \left(1 + \gamma^{1+\delta} \mathbf{L}\right)^k + \gamma \mathbf{L}$ and $\mathcal{B}_{k, \gamma, \lambda} = k M_{\Gamma, \gamma}(\lambda) L_{\Gamma, \gamma}(\lambda)^k + \mathcal{L}_{k, \gamma, \lambda}^x + \gamma^\delta (1 + k \mathcal{L}_{k, \gamma, \lambda}^v)$. The conclusion then follows from (3.67) and (3.68). \square

3.4.3.3 Estimates on diffeomorphisms defined for the total noise

We use in this section the stability results obtained in Lemmas 13 and 14 to prove that the function $\bar{\Gamma}_{\gamma, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}$ giving the $(k+1)$ -th iterate in (3.61), considered as a function of (g_1, g_2) for fixed initial condition (x, v) , is a C^1 -diffeomorphism. In addition, we also establish regularity properties of this function and its inverse with respect to the initial condition (x, v) . To state our results, we introduce the following functions for $(k, \gamma) \in \mathbb{E}_c$, and given realizations $\tilde{\mathbf{z}} = \{\tilde{z}_j\}_{j=1}^{k-1} \subset \mathbb{R}^d$, $\mathbf{w} = \{w_j\}_{j=1}^{k+1} \subset \mathbb{R}^m$:

- for a given initial condition $(x, v) \in \mathbb{R}^{2d}$, set $\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(k+1)}(g_1, g_2) = \bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}}^{(k+1)}((x, v), (g_1, g_2))$;
- for a given realization $(g_1, g_2) \in \mathbb{R}^{2d}$, set $\bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(k+1)}(x, v) = \bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}}^{(k+1)}((x, v), (g_1, g_2))$.

The first result is that $\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(k+1)}$ is a diffeomorphism when $k\gamma$ is a sufficiently small positive time and γ is not too large.

Proposition 3.8. *Assume that A1 and A2 hold. There exists $\bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$, there is $\bar{\gamma}_{t_0} > 0$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$, $(x, v) \in \mathbb{R}^{2d}$, $\tilde{\mathbf{z}} \in \mathbb{R}^{(k-1) \times d}$ and $\mathbf{w} \in \mathbb{R}^{(k+1) \times m}$*

(a) $\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)}$ is 3/2-Lipschitz and a C^1 -diffeomorphism on \mathbb{R}^{2d} satisfying, for any $(g_1, g_2), (g'_1, g'_2) \in \mathbb{R}^{2d}$,

$$\left\| \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)}(g_1, g_2) - \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)}(g'_1, g'_2) \right\| \geq \frac{1}{2} \|(g_1 - g'_1, g_2 - g'_2)\| ; \quad (3.69)$$

(b) the inverse of $\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor)}$, denoted by $\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor)}$, is 2-Lipschitz: for any $(u_1, u_2), (u'_1, u'_2) \in \mathbb{R}^{2d}$,

$$\left\| \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) - \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u'_1, u'_2) \right\| \leq 2 \|(u_1 - u'_1, u_2 - u'_2)\| .$$

Proof. We prove that there exists $\bar{t}_0 > 0$ such that, for any $t_0 \in (0, \bar{t}_0]$, there is $\bar{\gamma}_{t_0} > 0$ for which $(g_1, g_2) \mapsto \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor)}(g_1, g_2) - (g_1, g_2)$ is 1/2-Lipschitz. Item (a) is then a straightforward consequence of [39, Exercise 3.26]. First, for any $\lambda > 0$, $(k, \gamma) \in \mathbb{E}_c$, $(x, v) \in \mathbb{R}^{2d}$, $\tilde{\mathbf{z}} \in \mathbb{R}^{(k-1) \times d}$, $\mathbf{w} \in \mathbb{R}^{(k+1) \times m}$, $(g_1, g_2), (g'_1, g'_2) \in \mathbb{R}^{2d}$, by (3.61) and Lemma 14, we have

$$\begin{aligned} & \left\| \bar{\Gamma}_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g_1, g_2) - \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} - \bar{\Gamma}_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g'_1, g'_2) + \begin{pmatrix} g'_1 \\ g'_2 \end{pmatrix} \right\| \\ & \leq \gamma^\delta \mathcal{D} \left\| \tilde{\Lambda}_{\{\tilde{z}_j\}_{j=1}^{k-1}}^{(k+1)}(g_1, g_2) - \tilde{\Lambda}_{\{\tilde{z}_j\}_{j=1}^{k-1}}^{(k+1)}(g'_1, g'_2) \right\| \\ & \quad + \left\| \Xi_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g_1, g_2) - \Xi_{\gamma, x, v, \{\tilde{z}_j\}_{j=1}^{k-1}, \{w_j\}_{j=1}^{k+1}}^{(k+1)}(g'_1, g'_2) \right\| \\ & \leq C(k+1, \gamma, \lambda) (\|g_1 - g'_1\| + \|g_2 - g'_2\|) , \end{aligned}$$

3.4. Proof of the minorization condition

where we have set

$$C(k+1, \gamma, \lambda) = \gamma^\delta \mathcal{D} \left(1 + \mathfrak{m}(k, \gamma)k\gamma + \frac{1 - \tau\gamma}{\gamma} [1 + \mathfrak{m}(k, \gamma)(k\gamma)^2] \right) + (2 + k\gamma)\mathbb{L}\mathbb{L}_{\Xi, k+1, \gamma, \lambda}.$$

As noted previously, there is $\bar{t}_0^{(1)} \geq 0$ for which, for any $t_0 \in (0, \bar{t}_0^{(1)})$, there exists $\bar{\gamma}_{t_0}^{(1)}$ such that for any $\gamma \in (0, \bar{\gamma}_{t_0}^{(1)})$, we have $(\lfloor t_0/\gamma \rfloor, \gamma) \in \mathbb{E}_c$. Then, Lemma 12 implies that there exists $\bar{t}_0^{(2)} \in (0, \bar{t}_0^{(1)})$ such that for any $t_0 \in (0, \bar{t}_0^{(2)})$, $\limsup_{\gamma \downarrow 0} \mathfrak{m}(\lfloor t_0/\gamma \rfloor, \gamma) \leq K[t_0^{-2} \vee t_0^{-1}]$ for some constant $K \geq 0$. This implies, for any $t_0 \in (0, \bar{t}_0^{(2)})$ and $\lambda > 0$,

$$\begin{aligned} \limsup_{\gamma \downarrow 0} C(\lfloor t_0/\gamma \rfloor + 1, \gamma, \lambda) &= \limsup_{\gamma \downarrow 0} (2 + \lfloor t_0/\gamma \rfloor \gamma) \mathbb{L}[t_0/\gamma]^2 \gamma^2 \mathbb{M}_{\Gamma, \gamma}(\lambda) \mathbb{L}_{\Gamma, \gamma}(\lambda)^{\lfloor t_0/\gamma \rfloor} \\ &\quad \times \mathfrak{m}(\lfloor t_0/\gamma \rfloor, \gamma) \\ &= \lambda(2 + t_0) \mathbb{L} t_0^2 \exp\left(t_0 \left[\frac{1}{\lambda} + (1 + \lambda)\mathbb{L} \right]\right) \limsup_{\gamma \downarrow 0} \mathfrak{m}(\lfloor t_0/\gamma \rfloor, \gamma) \\ &\leq \lambda(2 + t_0) \mathbb{L} K [t_0 \vee 1] \exp\left(t_0 \left[\frac{1}{\lambda} + (1 + \lambda)\mathbb{L} \right]\right), \end{aligned}$$

where we have used for the penultimate inequality the expressions of $\mathbb{M}_{\Gamma, \gamma}(\lambda)$, $\mathbb{L}_{\Gamma, \gamma}(\lambda)$ provided by (3.48). The latter quantity is smaller than 1/4 for any $t_0 \in (0, \bar{t}_0^{(2)})$ upon choosing first $\lambda = 1/(16K\mathbb{L})$, and then $\bar{t}_0 \in (0, \bar{t}_0^{(2)})$ sufficiently small so that for any $t_0 \in (0, \bar{t}_0]$,

$$(2 + t_0)[t_0 \vee 1] \exp\left(t_0 \left[16K\mathbb{L} + \left(1 + \frac{1}{16K\mathbb{L}}\right)\mathbb{L} \right]\right) \leq 4.$$

This leads to the claimed statement, and therefore proves item (a). Item (b) next easily follows from (3.69). \square

The second result is that $\bar{\Gamma}_{\gamma, \bar{\mathbf{z}}, \mathbf{w}, g_1, g_2}$ is a diffeomorphism when $k\gamma$ is a sufficiently small positive time and γ is not too large.

Proposition 3.9. *Assume that A1 and A2 hold. There exists $\bar{t}_0 > 0$, such that, for any $t_0 \in (0, \bar{t}_0]$, there is $\bar{\gamma}_{t_0} \in (0, 1]$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$, $(g_1, g_2), (u_1, u_2) \in \mathbb{R}^{2d}$, $\bar{\mathbf{z}} \in \mathbb{R}^{(k-1) \times d}$ and $\mathbf{w} \in \mathbb{R}^{(k+1) \times m}$,*

(a) *the mapping $\bar{\Gamma}_{\gamma, \bar{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{\lfloor t_0/\gamma \rfloor + 1}$ is Lipschitz on \mathbb{R}^{2d} : for any $(x, v), (x', v') \in \mathbb{R}^{2d}$,*

$$\left\| \bar{\Gamma}_{\gamma, \bar{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{\lfloor t_0/\gamma \rfloor + 1}(x, v) - \bar{\Gamma}_{\gamma, \bar{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{\lfloor t_0/\gamma \rfloor + 1}(x', v') \right\| \leq e^{(1+2\mathbb{L})(t_0+\gamma)} (\|x - x'\| + \|v - v'\|);$$

(b) *the inverse of $\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{\lfloor t_0/\gamma \rfloor + 1}$ (see Proposition 3.8), namely $(x, v) \mapsto \bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2)$, is Lipschitz on \mathbb{R}^{2d} : for any $(x, v), (x', v') \in \mathbb{R}^{2d}$,*

$$\left\| \bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) - \bar{\Gamma}_{\gamma, x', v', \bar{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) \right\| \leq 2e^{(1+2\mathbb{L})(t_0+\gamma)} (\|x - x'\| + \|v - v'\|).$$

Proof. Let $\bar{t}_0 \geq 0$ such that for any $t_0 \in (0, \bar{t}_0]$, there exists $\bar{\gamma}_{t_0}$ for which, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$, $(\lfloor t_0/\gamma \rfloor, \gamma) \in \mathbf{E}_c$ and Proposition 3.8-(a)-(b) hold. To prove item (a), we note that (3.57)-(3.58) and (3.47) in Lemma 13 with $\lambda = 1$ imply for any $t_0 \in (0, \bar{t}_0]$, $\gamma \in (0, \bar{\gamma}_{t_0}]$, $(g_1, g_2) \in \mathbb{R}^{2d}$, $\tilde{\mathbf{z}} \in \mathbb{R}^{(k-1) \times d}$ and $\mathbf{w} \in \mathbb{R}^{(k+1) \times m}$, $(x, v), (x', v') \in \mathbb{R}^{2d}$,

$$\left\| \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)}(g_1, g_2) - \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)}(g_1, g_2) \right\| \leq L_{\Gamma, \gamma}(1)^{\lfloor t_0/\gamma \rfloor + 1} [\|x - x'\| + \|v - v'\|],$$

from which the result follows in view of the bound

$$L_{\Gamma, \gamma}(\lambda)^{\lfloor t_0/\gamma \rfloor + 1} \leq \exp \left((t_0 + \gamma) \left[\frac{1}{\lambda} + (1 + \lambda) \max \left(1, \frac{\gamma^\delta}{\lambda} \right) L \right] \right),$$

and the fact that $\gamma \leq \bar{\gamma}_{t_0} \leq 1$.

Regarding (b), let $(x, v), (x', v') \in \mathbb{R}^{2d}$ and denote by $(g_1, g_2) = \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor)}(u_1, u_2)$ and $(g'_1, g'_2) = \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor)}(u_1, u_2)$. We obtain using $(g'_1, g'_2) = \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor)}(\bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(\lfloor t_0/\gamma \rfloor)}(x, v))$ and Proposition 3.8-(b), for any $t_0 \in (0, \bar{t}_0]$, $\gamma \in (0, \bar{\gamma}_{t_0}]$, $(g_1, g_2) \in \mathbb{R}^{2d}$, $\tilde{\mathbf{z}} \in \mathbb{R}^{(k-1) \times d}$ and $\mathbf{w} \in \mathbb{R}^{(k+1) \times m}$,

$$\begin{aligned} \left\| \bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) - \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) \right\| &= \|(g_1, g_2) - (g'_1, g'_2)\| \\ &= \left\| \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(\bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(\lfloor t_0/\gamma \rfloor + 1)}(x', v')) - \bar{\Gamma}_{\gamma, x', v', \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(\bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(\lfloor t_0/\gamma \rfloor + 1)}(x, v)) \right\| \\ &\leq 2 \left\| \bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(\lfloor t_0/\gamma \rfloor + 1)}(x', v') - \bar{\Gamma}_{\gamma, \tilde{\mathbf{z}}, \mathbf{w}, g_1, g_2}^{(\lfloor t_0/\gamma \rfloor + 1)}(x, v) \right\|, \end{aligned}$$

which completes the proof using (a). \square

3.4.4 Proof of Theorem 3.1

We can now finally provide the proof of Theorem 3.1.

Proof of Theorem 3.1. For $(x, v) \in \mathbb{R}^{2d}$ and $B \in \mathcal{B}(\mathbb{R}^{2d})$, it holds, by the definition (3.61),

$$R_\gamma^{\lfloor t_0/\gamma \rfloor + 1}((x, v), B) = \mathbb{E} \left[\mathbb{1}_B \left\{ \bar{\Gamma}_{\gamma, x, v, \sqrt{\gamma} \sigma_\gamma \tilde{\mathbf{Z}}^{(\lfloor t_0/\gamma \rfloor + 1)}, \mathbf{w}^{(\lfloor t_0/\gamma \rfloor + 1)}} \left(\sigma_\gamma \mathbf{G}^{(\lfloor t_0/\gamma \rfloor + 1)} \right) \right\} \right].$$

Fix $M > 0$, and consider $\bar{t}_0 > 0$ such that the statements of Propositions 3.8 and 3.9 hold true. Introduce next $t_0 \in (0, \bar{t}_0]$ and the corresponding stepsize $\bar{\gamma}_{t_0}$ as given by Propositions 3.8 and 3.9.

The random variables $\sigma_\gamma \mathbf{G}^{(\lfloor t_0/\gamma \rfloor + 1)}$ and $\sqrt{\gamma} \sigma_\gamma \tilde{\mathbf{Z}}^{(\lfloor t_0/\gamma \rfloor + 1)}$ are independent Gaussian random variables by Lemma 11. Denoting their densities with respect to the Lebesgue measure by $\varphi_{t_0, \gamma}$ and $\psi_{t_0, \gamma}$ respectively, setting $n_0 = \lfloor t_0/\gamma \rfloor + 1$, and using a change of

3.4. Proof of the minorization condition

variable, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$

$$\begin{aligned} & R_\gamma^{[t_0/\gamma]+1}((x, v), B) \\ &= \int_{\mathbb{R}^{n_0 d} \times W^{n_0}} \mathbb{1}_B \left\{ \bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(g_1, g_2) \right\} \varphi_{t_0, \gamma}(g_1, g_2) \psi_{t_0, \gamma}(\bar{\mathbf{z}}) d\bar{\mathbf{z}} \mu_W^{\otimes [t_0/\gamma]+1}(d\mathbf{w}) dg_1 dg_2 \\ &= \int_{\mathbb{R}^{n_0 d} \times W^{n_0}} \mathbb{1}_B(u_1, u_2) J_{\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2) \varphi_{t_0, \gamma} \left(\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2) \right) \psi_{t_0, \gamma}(\bar{\mathbf{z}}) \\ & \hspace{20em} d\bar{\mathbf{z}} \mu_W^{\otimes [t_0/\gamma]+1}(d\mathbf{w}) du_1 du_2, \end{aligned}$$

where $\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}$ is the inverse of $\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}$ (well defined by Proposition 3.8) and $J_{\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2)$ is the absolute value of the determinant of the Jacobian matrix of this mapping.

We can now introduce the reference point $(\mathbf{0}_d, \mathbf{0}_d)$, and relate the transition probability starting from (x, v) in terms of transitions starting from this $(\mathbf{0}_d, \mathbf{0}_d)$. For reasons that will become clear below and similarly to what is done in Section 3.3, we replace $\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2)$ by $\sqrt{2} \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2)$ and not simply $\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2)$ in $\varphi_{t_0, \gamma}$, henceforth write for any $\gamma \in (0, \bar{\gamma}_{t_0}]$

$$\begin{aligned} & R_\gamma^{[t_0/\gamma]+1}((x, v), B) \\ &= \int_{\mathbb{R}^{n_0 d} \times W^{n_0}} \mathbb{1}_B(u_1, u_2) A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}(u_1, u_2) J_{\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2) \\ & \quad \times \varphi_{t_0, \gamma} \left(\sqrt{2} \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2) \right) \psi_{t_0, \gamma}(\bar{\mathbf{z}}) d\bar{\mathbf{z}} \mu_W^{\otimes [t_0/\gamma]+1}(d\mathbf{w}) du_1 du_2, \end{aligned} \tag{3.70}$$

where $A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}(u_1, u_2) = A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(1)}(u_1, u_2) A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2)$ with

$$\begin{aligned} A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(1)}(u_1, u_2) &= \frac{J_{\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2)}{J_{\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2)}, \\ A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2) &= \frac{\varphi_{t_0, \gamma} \left(\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2) \right)}{\varphi_{t_0, \gamma} \left(\sqrt{2} \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2) \right)}. \end{aligned}$$

We next bound $A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(1)}(u_1, u_2)$ and $A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2)$ from below. For the latter term, we assume that $\|x\| + \|v\| \leq M$.

By an application of Hadamard's inequality (see Proposition 3.10 below) and Proposition 3.8-(a), we obtain for any $(u_1, u_2) \in \mathbb{R}^{2d}$, $(\bar{\mathbf{z}}, \mathbf{w}) \in \mathbb{R}^{n_0 d} \times W^{n_0}$ and $\gamma \in (0, \bar{\gamma}_{t_0}]$,

$$J_{\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2) = \frac{1}{J_{\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}} \left(\bar{\Gamma}_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}(u_1, u_2) \right)} \geq \left(\frac{2}{3} \right)^d.$$

Similarly, $J_{\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \bar{\mathbf{z}}, \mathbf{w}}^{([t_0/\gamma]+1)}}(u_1, u_2) \leq 2^d$ by Proposition 3.8-(b), so that for any $(u_1, u_2) \in \mathbb{R}^{2d}$, $(\bar{\mathbf{z}}, \mathbf{w}) \in \mathbb{R}^{n_0 d} \times W^{n_0}$ and $\gamma \in (0, \bar{\gamma}_{t_0}]$,

$$A_{\gamma, x, v, \bar{\mathbf{z}}, \mathbf{w}}^{(1)}(u_1, u_2) \geq \frac{1}{3^d}. \tag{3.71}$$

We next recall that $\varphi_{t_0, \gamma}$ is a Gaussian density with mean 0 and covariance matrix $\sigma_\gamma^2 c^{(\lfloor t_0/\gamma \rfloor + 1)}$ (see (3.39)), and use the inequality (3.28) to write for any $(u_1, u_2) \in \mathbb{R}^{2d}$, $(\tilde{\mathbf{z}}, \mathbf{w}) \in \mathbb{R}^{n_0d} \times \mathbb{W}^{n_0}$, $\gamma \in (0, \bar{\gamma}_{t_0}]$,

$$\begin{aligned} & A_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2) \\ & \geq \exp \left(-\frac{1}{\sigma_\gamma^2} \left\| \left(c^{(\lfloor t_0/\gamma \rfloor + 1)} \right)^{-\frac{1}{2}} \left(\bar{\Gamma}_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) - \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) \right) \right\|^2 \right). \end{aligned}$$

Next, by Proposition 3.9-(b) and Lemma 9,

$$\begin{aligned} \liminf_{\tilde{\gamma} \downarrow 0} A_{\tilde{\gamma}, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2) & \geq \exp \left(-4e^{2(1+2L)t_0} \left\| \left[\boldsymbol{\Sigma}(t_0) \right]^{-1} \right\|_{\text{op}} (\|x\| + \|v\|)^2 \right) \\ & \geq \exp \left(-4e^{2(1+2L)t_0} \left\| \left[\boldsymbol{\Sigma}(t_0) \right]^{-1} \right\|_{\text{op}} M^2 \right) = \eta_{t_0, M}, \end{aligned}$$

where $\boldsymbol{\Sigma}^{(t_0)}$ is the matrix defined in (3.9). Upon reducing $\bar{\gamma}_{t_0}$, we can therefore assume that the following inequality holds: for any $(u_1, u_2) \in \mathbb{R}^{2d}$, $(\tilde{\mathbf{z}}, \mathbf{w}) \in \mathbb{R}^{n_0d} \times \mathbb{W}^{n_0}$, $\gamma \in (0, \bar{\gamma}_{t_0}]$,

$$A_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}^{(2)}(u_1, u_2) \geq \frac{\eta_{t_0, M}}{2}. \quad (3.72)$$

Now that the factor $A_{\gamma, x, v, \tilde{\mathbf{z}}, \mathbf{w}}(u_1, u_2)$ is bounded from below, we can consider the remaining terms in (3.70) in order to construct a reference minorization measure. More precisely, using again Proposition 3.8-(a), the change of variable $(g_1, g_2) = \sqrt{2} \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2)$ implies, for any $\gamma \in (0, \bar{\gamma}_{t_0}]$,

$$\begin{aligned} & \int_{\mathbb{R}^{n_0d} \times \mathbb{W}^{n_0}} \mathbb{1}_B(u_1, u_2) J_{\bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}}(u_1, u_2) \varphi_{t_0, \gamma} \left(\sqrt{2} \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \tilde{\mathbf{z}}, \mathbf{w}}^{(\leftarrow, \lfloor t_0/\gamma \rfloor + 1)}(u_1, u_2) \right) \psi_{t_0, \gamma}(\tilde{\mathbf{z}}) \\ & \qquad \qquad \qquad d\tilde{\mathbf{z}} \mu_W^{\otimes \lfloor t_0/\gamma \rfloor + 1}(d\mathbf{w}) du_1 du_2 \\ & = 2^{-d} \int_{\mathbb{R}^{n_0d} \times \mathbb{W}^{n_0}} \mathbb{1}_B \left\{ \bar{\Gamma}_{\gamma, \mathbf{0}_d, \mathbf{0}_d, \tilde{\mathbf{z}}, \mathbf{w}}^{(\lfloor t_0/\gamma \rfloor + 1)} \left(\frac{g_1}{\sqrt{2}}, \frac{g_2}{\sqrt{2}} \right) \right\} \varphi_{t_0, \gamma}(g_1, g_2) \psi_{t_0, \gamma}(\tilde{\mathbf{z}}) \\ & \qquad \qquad \qquad d\tilde{\mathbf{z}} \mu_W^{\otimes \lfloor t_0/\gamma \rfloor + 1}(d\mathbf{w}) dg_1 dg_2 \\ & = 2^{-d} \mu_{t_0, \gamma}(B), \end{aligned}$$

where $\mu_{t_0, \gamma}$ is a probability measure on $(\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d}))$. By combining the latter inequality and the lower bounds (3.71) and (3.72) in (3.70), we obtain the following result: for any $\gamma \in (0, \bar{\gamma}_{t_0}]$ and $(x, v) \in \mathbb{R}^{2d}$ such that $\|x\| + \|v\| \leq M$,

$$R_\gamma^{\lfloor t_0/\gamma \rfloor + 1}((x, v), B) \geq \varepsilon_{t_0, M} \mu_{t_0, \gamma}(B),$$

with $\varepsilon_{t_0, M} = 6^{-d} \eta_{t_0, M}/2$. The inequality (3.14) then follows from [37, Lemma 18.2.7] for instance. \square

3.5. Proof of Theorem 3.4

We conclude this section by recalling (and proving for completeness) a well-known result on the Jacobian determinant of C^1 Lipschitz function.

Proposition 3.10. *Let $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a continuously differentiable L -Lipschitz function, i.e. for any $x, y \in \mathbb{R}^d$,*

$$\|\Phi(x) - \Phi(y)\| \leq L \|x - y\| .$$

Then, $J_\Phi(x) \leq L^d$, where J_Φ denotes the absolute value of the determinant of the Jacobian matrix of Φ .

Proof. Denote by $\text{Jac}_\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ the Jacobian matrix of Φ . Since Φ is L -Lipschitz, it holds, for any $(x, h) \in \mathbb{R}^d \times \mathbb{R}^d$, $\|\text{Jac}_\Phi(x)h\| \leq L \|h\|$. By choosing $h = e_i$ for $i \in \{1, \dots, d\}$ (with e_1, \dots, e_d the canonical basis of \mathbb{R}^d), it follows that

$$\|\nabla \Phi_i\|^2 = \sum_{j=1}^d \left(\frac{\partial \Phi_i}{\partial x_j} \right)^2 \leq L^2,$$

where Φ_i is the i -th component of Φ . In addition, using Hadamard's inequality (see for instance [62, Example 4.18]), we obtain $|\det(\text{Jac}_\Phi(x))| \leq \prod_{i=1}^d \|\nabla \Phi_i\| \leq L^d$, which completes the proof. \square

3.5 Proof of Theorem 3.4

The function $\mathcal{W}_\gamma : \mathbb{R}^{2d} \rightarrow \mathbb{R}$ which appears in (3.15) is defined under **A1** and **D2** for $x, v \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}]$ as

$$\mathcal{W}_\gamma(x, v) = \frac{\kappa^2}{2} \|x\|^2 + \|v\|^2 + \frac{\kappa^2 \gamma (1 + \gamma^\delta \vartheta_\gamma)}{1 - \tau_\gamma} \langle x, v \rangle + 2\alpha_U U(x). \quad (3.73)$$

This function is indeed nonnegative by Lemma 16 below. The proof of Theorem 3.4 relies on a conditioned Lyapunov drift inequality for \mathcal{W}_γ (see Section 3.5.1), which motivates the choice of the various prefactors in (3.73) (see the discussion at the end of the proof of Lemma 15). This allows to write the proof of Theorem 3.4 in Section 3.5.2, with the help of some technical results postponed to Section 3.5.3.

3.5.1 Conditioned drift inequality

The following Lyapunov inequality is a key result to prove Theorem 3.4.

Lemma 15. *Assume that **A1**, **D1**, **D2** and **D3** hold. For any $x, v \in \mathbb{R}^d$, $w \in \mathbb{R}^{m_1+m_2}$ and $\gamma \in (0, \bar{\gamma}]$, set $(X_1^{x,v,w}, V_1^{x,v,w}) = \Gamma_\gamma(x, v, \sqrt{\gamma} \sigma_\gamma Z, w)$, where Γ_γ is defined by (3.13) and Z is a d -dimensional standard Gaussian random variable. Then, there exists $C \geq 0$ such that, for any $\gamma \in (0, \bar{\gamma} \wedge 1]$, $x, v \in \mathbb{R}^d$ and $w = (w_1, w_2) \in \mathbb{R}^{m_1+m_2}$,*

$$\begin{aligned} \mathbb{E} [\mathcal{W}_\gamma(X_1^{x,v,w}, V_1^{x,v,w})] &\leq \mathcal{W}_\gamma(x, v) - \gamma \kappa \|v\|^2 - \gamma \kappa \zeta_U \left[\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right] \\ &\quad + C \gamma \left[1 + \|\nabla U(x)\| + \|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) + \gamma^{\delta_U} \|x\| \|w_1\| \right], \end{aligned}$$

where the expectation is over the realizations of Z , and $\tilde{\mathcal{F}}(x, v, w) = \|\nabla U(x)\|^2 / L^2 + \|v\|^2 + \|w\|^2 + \|x\|$.

Proof. The finite nonnegative constant C in this proof may change from line to line but does not depend on γ , x , v or w . For ease of notation, we also simply denote $(X_1^{x,v,w}, V_1^{x,v,w})$ by (X_1, V_1) , and do not explicitly indicate that $\gamma \in (0, \bar{\gamma} \wedge 1]$. By definition,

$$\mathbb{E}[\mathcal{W}_\gamma(X_1, V_1)] = 2\alpha_U \mathbb{E}[U(X_1)] + \frac{\kappa^2}{2} \mathbb{E}[\|X_1\|^2] + \mathbb{E}[\|V_1\|^2] + \frac{\kappa^2 \gamma (\gamma^\delta \vartheta_\gamma + 1)}{1 - \tau_\gamma} \mathbb{E}[\langle X_1, V_1 \rangle]. \quad (3.74)$$

We successively bound each expectation on the right-hand side of the previous equality. To bound $\mathbb{E}[U(X_1)]$, we make use of the following estimates, which are straightforwardly obtained from **D2** and the equality $\mathbb{E}[\mathcal{F}(x, v, \sqrt{\gamma} \sigma_\gamma Z, w)] = \tilde{\mathcal{F}}(x, v, w) + \gamma \sigma_\gamma^2 d$:

$$\begin{aligned} \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\|^2 \right] &\leq C \left[1 + \gamma^{\delta v} \tilde{\mathcal{F}}(x, v, w) \right], \\ \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\| \right] &\leq C \left[1 + \gamma^{\delta v/2} \sqrt{\tilde{\mathcal{F}}(x, v, w)} \right], \end{aligned} \quad (3.75)$$

the second inequality being obtained from the first one by a Cauchy–Schwarz inequality and the bound $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$. Using [106, Lemma 1.2.3] and **D1**, we obtain that

$$\begin{aligned} \mathbb{E}[U(X_1)] &= \mathbb{E} \left[U \left(x + \gamma v + \gamma f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \gamma^{\delta+1/2} \sigma_\gamma \mathbf{D}_\gamma Z \right) \right] \\ &\leq U(x) + \mathbb{E} \left[\left\langle \nabla U(x), \gamma v + \gamma f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\rangle \right] \\ &\quad + \frac{L}{2} \mathbb{E} \left[\left\| \gamma v + \gamma f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \gamma^{\delta+1/2} \sigma_\gamma \mathbf{D}_\gamma Z \right\|^2 \right] \\ &\leq U(x) + \gamma \langle \nabla U(x), v \rangle + \gamma \|\nabla U(x)\| \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\| \right] \\ &\quad + \frac{3L}{2} \left(\gamma^2 \|v\|^2 + \gamma^2 \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\|^2 \right] + \gamma^{1+2\delta} \sigma_\gamma^2 \mathbb{E} \left[\|\mathbf{D}_\gamma Z\|^2 \right] \right). \end{aligned}$$

Note first that, in view of (3.75) and **A1**, and since $\|v\|^2 \leq \tilde{\mathcal{F}}(x, v, w)$,

$$\begin{aligned} \gamma^2 \|v\|^2 + \gamma^2 \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\|^2 \right] + \gamma^{1+2\delta} \sigma_\gamma^2 \mathbb{E} \left[\|\mathbf{D}_\gamma Z\|^2 \right] \\ \leq C \left[\gamma^{1+(2\delta) \wedge 1} + \gamma^2 \tilde{\mathcal{F}}(x, v, w) \right]. \end{aligned}$$

Moreover, still with (3.75), and since $\|\nabla U(x)\| \leq L \sqrt{\tilde{\mathcal{F}}(x, v, w)}$,

$$\begin{aligned} \|\nabla U(x)\| \mathbb{E} \left[\left\| f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\| \right] &\leq C \|\nabla U(x)\| \left[1 + \gamma^{\delta v/2} \sqrt{\tilde{\mathcal{F}}(x, v, w)} \right], \\ &\leq C \left[\|\nabla U(x)\| + \gamma^{\delta v/2} \tilde{\mathcal{F}}(x, v, w) \right]. \end{aligned} \quad (3.76)$$

3.5. Proof of Theorem 3.4

Since $\delta_U \leq 1$, this leads finally to

$$\mathbb{E} [U(X_1)] \leq U(x) + \gamma \langle \nabla U(x), v \rangle + C\gamma \left[\gamma^{(2\delta) \wedge 1} + \|\nabla U(x)\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right]. \quad (3.77)$$

Let us next bound $\mathbb{E} [\|X_1\|^2]$. Note first that **D2** implies that

$$\begin{aligned} \mathbb{E} \left[\langle x, f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) \rangle \right] \\ \leq \gamma^\delta \vartheta_\gamma \langle x, v \rangle + C_U \left[1 + \gamma^{\delta_U} \tilde{\mathcal{F}}(x, v, w) + \gamma^{1+2\delta+\delta_U} \sigma_\gamma^2 d + \gamma^{\delta_U} \|x\| \|w_1\| \right], \end{aligned}$$

so that, with **A1** and **D2**,

$$\begin{aligned} \mathbb{E} [\|X_1\|^2] &\leq \|x\|^2 + \gamma^2 \|v\|^2 + \gamma^2 \mathbb{E} \left[\left\| f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) \right\|^2 \right] \\ &\quad + 2\gamma \mathbb{E} \left[\langle x, f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) \rangle \right] + \gamma^{1+2\delta} \sigma_\gamma^2 \mathcal{D}^2 d + 2\gamma \langle x, v \rangle \\ &\quad + 2\gamma^2 \mathbb{E} \left[\langle v, f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) \rangle \right] \\ &\quad + 2\gamma^{3/2+\delta} \sigma_\gamma \mathbb{E} \left[\langle f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w), \mathbf{D}_\gamma Z \rangle \right] \\ &\leq \|x\|^2 + 2\gamma (1 + \gamma^\delta \vartheta_\gamma) \langle x, v \rangle + C\gamma \left[1 + \gamma^{\delta_U} \tilde{\mathcal{F}}(x, v, w) + \gamma^{\delta_U} \|x\| \|w_1\| \right]. \end{aligned} \quad (3.78)$$

To bound $\mathbb{E} [\|V_1\|^2]$, we rely on the following estimates, obtained from **D2**:

$$\begin{aligned} \mathbb{E} \left[\left\| g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) + \alpha_U \nabla U(x) \right\|^2 \right] &\leq C (1 + \gamma^{\delta_U} \tilde{\mathcal{F}}(x, v, w)), \\ \mathbb{E} \left[\left\| g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) + \alpha_U \nabla U(x) \right\| \right] &\leq C \left(1 + \gamma^{\delta_U/2} \sqrt{\tilde{\mathcal{F}}(x, v, w)} \right), \\ \mathbb{E} \left[\langle x, g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) \rangle \right] &\leq -\zeta_U \left[\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right] + C\gamma^{\delta_U} (1 + \tilde{\mathcal{F}}(x, v, w)). \end{aligned} \quad (3.79)$$

Therefore,

$$\begin{aligned} \mathbb{E} [\|V_1\|^2] &= \tau_\gamma^2 \|v\|^2 + \gamma^2 \alpha_U^2 \|\nabla U(x)\|^2 + \gamma^2 \mathbb{E} \left[\left\| g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) + \alpha_U \nabla U(x) \right\|^2 \right] \\ &\quad + \gamma \sigma_\gamma^2 d - 2\tau_\gamma \gamma \alpha_U \langle v, \nabla U(x) \rangle \\ &\quad + 2\tau_\gamma \gamma \langle v, \mathbb{E} [g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w)] + \alpha_U \nabla U(x) \rangle \\ &\quad - 2\gamma^2 \alpha_U \langle \nabla U(x), \mathbb{E} [g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w)] + \alpha_U \nabla U(x) \rangle \\ &\quad + 2\gamma^{3/2} \sigma_\gamma \mathbb{E} \left[\langle g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w) + \alpha_U \nabla U(x), Z \rangle \right]. \end{aligned}$$

The terms on the third and fourth lines of the above series of inequalities can be bounded as in (3.76) using **D2**. For instance, for the one in the third line:

$$\begin{aligned}
 \left| \left\langle v, \mathbb{E} \left[g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right] + \alpha_U \nabla U(x) \right\rangle \right| \\
 \leq \|v\| \mathbb{E} \left[\left\| g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \alpha_U \nabla U(x) \right\| \right] \\
 \leq C \left[\|v\| + \gamma^{\delta_U/2} \|v\| \sqrt{\tilde{\mathcal{F}}(x, v, w)} \right], \\
 \leq C \left[\|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right].
 \end{aligned}$$

We thus obtain with **A1** that

$$\begin{aligned}
 \mathbb{E} \left[\|V_1\|^2 \right] &\leq \tau_\gamma^2 \|v\|^2 - 2\tau_\gamma \gamma \alpha_U \langle v, \nabla U(x) \rangle + C\gamma \left\{ 1 + \|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right\} \\
 &\leq (1 - \kappa\gamma) \|v\|^2 - 2\gamma \alpha_U \langle v, \nabla U(x) \rangle + C\gamma \left\{ 1 + \|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right\},
 \end{aligned} \tag{3.80}$$

where we have used for the last inequality that $|\tau_\gamma - 1| \leq \kappa\gamma$ and $\tau_\gamma^2 = e^{-2\kappa\gamma} + \tau_\gamma^2 - e^{-2\kappa\gamma} \leq 1 - 2\gamma\kappa + 2(\kappa^2 + C_\kappa)\gamma^2$ by Lemma 8 (the term of order γ^2 going into the remainder term in the inequality). Finally, using **A1** and **D2**, as well as the estimates (3.75) and (3.79),

$$\begin{aligned}
 \mathbb{E} [\langle X_1, V_1 \rangle] &= \tau_\gamma \langle x, v \rangle + \gamma \mathbb{E} \left[\left\langle x, g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) \right\rangle \right] + \gamma \tau_\gamma \|v\|^2 \\
 &\quad + \gamma \tau_\gamma \mathbb{E} \left[\left\langle f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right), v \right\rangle \right] \\
 &\quad - \gamma^2 \alpha_U \mathbb{E} \left[\left\langle f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right), \nabla U(x) \right\rangle \right] \\
 &\quad + \gamma^2 \mathbb{E} \left[\left\langle f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right), g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \alpha_U \nabla U(x) \right\rangle \right] \\
 &\quad + \gamma^2 \mathbb{E} \left[\left\langle v, g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \alpha_U \nabla U(x) \right\rangle \right] - \gamma^2 \alpha_U \langle v, \nabla U(x) \rangle \\
 &\quad + \gamma^{3/2} \sigma_\gamma \mathbb{E} \left[\left\langle f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right), Z \right\rangle \right] \\
 &\quad + \gamma^{\delta+3/2} \sigma_\gamma \mathbb{E} \left[\left\langle \mathbf{D}_\gamma Z, g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma Z, w \right) + \alpha_U \nabla U(x) \right\rangle \right] \\
 &\quad + \gamma^{1+\delta} \sigma_\gamma^2 \mathbb{E} [\langle \mathbf{D}_\gamma Z, Z \rangle] \\
 &\leq \tau_\gamma \langle x, v \rangle - \gamma \zeta_U \left[\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right] + \gamma \tau_\gamma \|v\|^2 + C\gamma \left\{ 1 + \|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right\}.
 \end{aligned} \tag{3.81}$$

We can now come back to (3.74). In order to make apparent our choices of prefactors for the Lyapunov function (3.73), we introduce real numbers $c_1, c_2, c_3 > 0$ and $c_4 \in \mathbb{R}$, and define

$$\mathcal{W}_c(x, v) = c_1 U(x) + c_2 \|x\|^2 + c_3 \|v\|^2 + c_4 \langle x, v \rangle.$$

3.5. Proof of Theorem 3.4

By multiplying (3.77) by c_1 , (3.78) by c_2 , (3.80) by c_3 and (3.81) by c_4 , we obtain

$$\begin{aligned} \mathbb{E} [\mathcal{W}_c(X_1, V_1)] &\leq \mathcal{W}_c(x, v) - c_4 \gamma \zeta_U \left[\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right] - \gamma [2c_3 \kappa - c_4 \tau_\gamma] \|v\|^2 \\ &\quad + \left[c_4 (\tau_\gamma - 1) + 2c_2 \gamma (1 + \gamma^\delta \vartheta_\gamma) \right] \langle x, v \rangle + \gamma [c_1 - 2c_3 \alpha_U] \langle v, \nabla U(x) \rangle \\ &\quad + C\gamma \left[1 + \|\nabla U(x)\| + \|v\| + \gamma^{\delta_U} \|x\| \|w_1\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, w) \right]. \end{aligned}$$

We choose c_1, c_2, c_3, c_4 to cancel the prefactors of the scalar products in the second line and to ensure that the term $-[2\kappa c_3 - c_4 \tau_\gamma] \|v\|^2 \leq -\tilde{c} \|v\|^2 + C\gamma^2 \tilde{\mathcal{F}}(x, v, w)$ for some $\tilde{c} > 0$. Such requirements can be obtained with $\tilde{c} = \kappa$ by setting $c_1 = 2\alpha_U$, $c_2 = \kappa^2/2$, $c_3 = 1$ and $c_4 = \kappa^2 \gamma (1 + \gamma^\delta \vartheta_\gamma) / (1 - \tau_\gamma)$ and using $|c_4 - \kappa| \leq C\gamma$ by Lemma 8. \square

3.5.2 Proof of Theorem 3.4

We are now in position to prove Theorem 3.4. The finite nonnegative constant C in this proof may change from line to line but does not depend on γ, x, v or w . A first idea in the proof is to rewrite the Lyapunov condition to be shown for $\overline{\mathcal{W}}_{\gamma, \varpi}$ as a Lyapunov condition for the function

$$\phi_\gamma = \sqrt{1 + \mathcal{W}_\gamma}, \quad (3.82)$$

thanks to **D3**. In all this proof, the timestep γ belongs to $(0, \bar{\gamma}_{\mathcal{W}}]$ with $\bar{\gamma}_{\mathcal{W}} > 0$ defined below in (3.89). We also denote by $(X_1, V_1) = \Gamma_\gamma(x, v, \sqrt{\gamma} \sigma_\gamma Z, W)$, the output of one step of the Markov chain starting from a given configuration $x, v \in \mathbb{R}^d$, with Z a d -dimensional standard Gaussian random variable and $W = (W_1, W_2)$ a random variable independent of Z with distribution $\mu_w = \mu_{w_1} \otimes \mu_{w_2}$ (recall that Γ_γ is defined by (3.13)).

We start by bounding $R_\gamma \overline{\mathcal{W}}_{\gamma, \varpi}$ in terms of $\mathbb{E} [\phi_\gamma(X_1, V_1) | W_2]$. The first observation is that, in view of Lemmas 18 and 19, there exists a constant $\mathcal{L} \in \mathbb{R}_+$ such that the function $(z, w_1) \mapsto \phi_\gamma(\Gamma_\gamma(x, v, \sqrt{\gamma} \sigma_\gamma z, (w_1, w_2)))$ is Lipschitz with Lipschitz constant $\sqrt{\mathcal{L} \gamma}$, uniform in x, v and w_2 . Therefore, by **D3-1)-(a)**, for any $\varpi \in \mathbb{R}_+^*$,

$$R_\gamma \overline{\mathcal{W}}_{\gamma, \varpi}(x, v) = \mathbb{E} \left[\mathbb{E} \left[\overline{\mathcal{W}}_{\gamma, \varpi}(X_1, V_1) \mid W_2 \right] \right] \leq \mathbb{E} \left[\exp \left(\varpi \mathbb{E} [\phi_\gamma(X_1, V_1) | W_2] + \frac{\varpi^2 \mathcal{L} \gamma}{2} \right) \right]. \quad (3.83)$$

The next step of the proof is to obtain bounds on $\mathbb{E} [\phi_\gamma(X_1, V_1) | W_2]$. We first relate this quantity to $\mathbb{E} [\mathcal{W}_\gamma(X_1, V_1) | W_2]$ and then rely on Lemma 15. More precisely, using the bound $\sqrt{1+t} \leq 1+t/2$ for $t \geq -1$,

$$\begin{aligned} \frac{\mathbb{E} [\phi_\gamma(X_1, V_1) | W_2]}{\phi_\gamma(x, v)} &= \mathbb{E} \left[\sqrt{1 + \frac{\mathcal{W}_\gamma(X_1, V_1) - \mathcal{W}_\gamma(x, v)}{\phi_\gamma^2(x, v)}} \mid W_2 \right] \\ &\leq \mathbb{E} \left[1 + \frac{\mathcal{W}_\gamma(X_1, V_1) - \mathcal{W}_\gamma(x, v)}{2\phi_\gamma^2(x, v)} \mid W_2 \right] \\ &= 1 + \frac{\mathbb{E} [\mathcal{W}_\gamma(X_1, V_1) | W_2] - \mathcal{W}_\gamma(x, v)}{2\phi_\gamma(x, v)^2}. \end{aligned} \quad (3.84)$$

Since W_1 admits a first moment by **D3**, using that $a \leq (2\varepsilon)^{-1} + \varepsilon a^2/2$ for any $a, \varepsilon > 0$, there exists $\bar{\gamma}_U^{(1)} \in (0, \bar{\gamma}_W]$ such that for any $\gamma \in (0, \bar{\gamma}_U^{(1)}]$,

$$\begin{aligned} C\mathbb{E} \left[1 + \|\nabla U(x)\| + \|v\| + \gamma^{\delta_U/2} \tilde{\mathcal{F}}(x, v, W) + \gamma^{\delta_U} \|x\| \|W_1\| |W_2] \right. \\ \left. \leq C(1 + \gamma^{\delta_U/2} \|W_2\|^2) + \frac{\kappa}{2} \left[\|v\|^2 + \zeta_U \left(\frac{\|\nabla U(x)\|^2}{L^2} + \|x\| \right) \right]. \end{aligned}$$

Therefore, by Lemma 15 and since $-\|v\|^2 \leq -\|v\| + 1$,

$$\begin{aligned} \mathbb{E} [\mathcal{W}_\gamma(X_1, V_1) | W_2] &\leq \mathcal{W}_\gamma(x, v) - \frac{\gamma\kappa}{2} (\|v\|^2 + \zeta_U \|x\|) + C\gamma(1 + \|W_2\|^2) \\ &\leq \mathcal{W}_\gamma(x, v) - \frac{\gamma\kappa \min(1, \zeta_U)}{2} (\|v\| + \|x\|) + C\gamma(1 + \|W_2\|^2). \end{aligned}$$

Plugging this estimate in (3.84) and using Lemmas 17 and 16, as well as the inequality $\sqrt{a+b} \geq 2^{-1}(\sqrt{a} + \sqrt{b})$ for $a, b \geq 0$, we obtain

$$\begin{aligned} \mathbb{E} [\phi_\gamma(X_1, V_1) | W_2] &\leq \phi_\gamma(x, v) + \gamma \frac{-\kappa \min(1, \zeta_U)(\|x\| + \|v\|) + C(1 + \|W_2\|^2)}{4\phi_\gamma(x, v)} \\ &\leq \phi_\gamma(x, v) - \gamma \frac{\kappa \min(1, \zeta_U)(\|x\| + \|v\|)}{4[1 + \mathfrak{C}_\phi(\|x\| + \|v\|)]} + C\gamma \frac{1 + \|W_2\|^2}{1 + \sqrt{c_W/2}(\|x\| + \|v\|)}. \end{aligned} \quad (3.85)$$

Then, by (3.83) and (3.85), for any $\gamma \in (0, \bar{\gamma}_U^{(1)}]$, $\varpi > 0$, and $x, v \in \mathbb{R}^d$,

$$\begin{aligned} \frac{R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v)}{\mathcal{W}_{\gamma, \varpi}(x, v)} &\leq \mathbb{E} \left[\exp \left(\frac{\varpi^2 \mathcal{L}\gamma}{2} + \varpi C\gamma \frac{1 + \|W_2\|^2}{1 + \sqrt{c_W/2}(\|x\| + \|v\|)} \right) \right] \\ &\quad \times \exp \left(-\gamma\varpi \frac{\kappa \min(1, \zeta_U)(\|x\| + \|v\|)}{4[1 + \mathfrak{C}_\phi(\|x\| + \|v\|)]} \right). \end{aligned} \quad (3.86)$$

We now choose successively $R_U > 0$ sufficiently large and $\varpi > 0$ sufficiently small so that the Markov chain induces a contraction for the Lyapunov function $\bar{\mathcal{W}}_{\gamma, \varpi}$ on the set $\{(x, v) \in \mathbb{R}^{2d}, \|x\| + \|v\| \geq R_U\}$. We first need to this end a bound on exponential moments of W_2 . By Jensen's inequality and **D3**, for any $c > 0$ and $\gamma \in (0, \bar{\gamma}_W/c]$, setting $C_W = \mathbb{E} [e^{\bar{\gamma}_W \|W_2\|^2}]$,

$$\mathbb{E} [e^{c\gamma \|W_2\|^2}] = \mathbb{E} \left[\left(e^{\bar{\gamma}_W \|W_2\|^2} \right)^{c\gamma/\bar{\gamma}_W} \right] \leq C_W^{c\gamma/\bar{\gamma}_W}. \quad (3.87)$$

For any $\varpi > 0$, $\gamma \in (0, \bar{\gamma}_U^{(1)}]$ such that $\varpi\gamma \leq \bar{\gamma}_W(1 + \sqrt{c_W/2}R_U)/C$ and $R_U > 0$, $x, v \in \mathbb{R}^d$, $\|x\| + \|v\| \geq R_U$, it therefore holds by (3.86),

$$\frac{1}{\varpi\gamma} \log \left(\frac{R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v)}{\mathcal{W}_{\gamma, \varpi}(x, v)} \right) \leq \frac{\varpi\mathcal{L}}{2} + \frac{C}{1 + \sqrt{c_W/2}R_U} \left(1 + \frac{\log C_W}{\bar{\gamma}_W} \right) - \frac{\kappa \min(1, \zeta_U)R_U}{4[1 + \mathfrak{C}_\phi R_U]}.$$

3.5. Proof of Theorem 3.4

Setting

$$R_U = 1 \vee \left[\frac{8[1 + \mathfrak{C}_\phi]}{\kappa \min(1, \zeta_U)} \left(1 + \frac{\log C_W}{\bar{\gamma}_W} \right) \sqrt{2/c_W} \right], \quad \varpi_U = \frac{\kappa \min(1, \zeta_U) R_U}{8\mathcal{L}[1 + \mathfrak{C}_\phi R_U]},$$

and $\bar{\gamma}_U^{(2)} = \bar{\gamma}_W [1 \wedge \{(1 + \sqrt{c_W/2} R_U)/C\varpi_U\}]$, it follows that, for any $\gamma \leq \min(\bar{\gamma}_U^{(1)}, \bar{\gamma}_U^{(2)})$ and $(x, v) \in \mathbb{R}^{2d}$ with $\|x\| + \|v\| \geq R_U$,

$$R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v) \leq \lambda_U^\gamma \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v), \quad \lambda_U = \exp \left(-\frac{\kappa \min(1, \zeta_U) R_U}{16[1 + \mathfrak{C}_\phi R_U]} \right) < 1. \quad (3.88)$$

We finally consider the case $\|x\| + \|v\| < R_U$. First, we note that (3.85) implies that for any $\gamma \leq \bar{\gamma}_U^{(1)}$, $\mathbb{E}[\phi_\gamma(X_1, V_1) | W_2] \leq \phi_\gamma(x, v) + C\gamma(1 + \|W_2\|^2)$. By plugging this result in (3.83), we obtain

$$R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v) \leq \mathbb{E} \left[\exp \left(\frac{\gamma \varpi_U}{2} [\varpi_U \mathcal{L} + 2C(1 + \|W_2\|^2)] \right) \right] \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v).$$

When $\gamma \leq \bar{\gamma}_U = \min(\bar{\gamma}_U^{(1)}, \bar{\gamma}_U^{(2)}, \bar{\gamma}_U^{(3)})$, with $\bar{\gamma}_U^{(3)} = \bar{\gamma}_W/(\varpi_U C)$, the right-hand side can be bounded by (3.87) as

$$R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v) \leq e^{K\gamma} \bar{\mathcal{W}}_{\gamma, \varpi_U}(x, v), \quad K = \frac{\varpi_U}{2} \left[\varpi_U \mathcal{L} + 2C \left(1 + \frac{\log C_W}{\bar{\gamma}_W} \right) \right].$$

We can therefore write, in view of the inequality $e^t - e^s \leq (t - s)e^t$ for $t, s \in \mathbb{R}$, $s \leq t$, for any $x, v \in \mathbb{R}^d$, $\|x\| + \|v\| < R_U$, $\gamma \leq \bar{\gamma}_U$,

$$\begin{aligned} R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) &\leq \lambda_U^\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) + (e^{K\gamma} - \lambda_U^\gamma) \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) \\ &\leq \lambda_U^\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) + \gamma(K - \log \lambda_U) e^{K\gamma} \bar{\mathcal{W}}_{\gamma, \varpi}(x, v). \end{aligned}$$

We finally obtain the following bound for any $x, v \in \mathbb{R}^d$, $\|x\| + \|v\| < R_U$, $\gamma \leq \bar{\gamma}_U$, using Lemma 17:

$$R_\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) \leq \lambda_U^\gamma \bar{\mathcal{W}}_{\gamma, \varpi}(x, v) + \gamma b_U, \quad b_U = (K - \log \lambda_U) e^{K\bar{\gamma}_U + \varpi_U(1 + \mathfrak{C}_\phi R_U)}.$$

Combining this bound with (3.88) completes the proof.

3.5.3 Supporting lemmas

The proofs of the following technical lemmas are postponed to the appendix. The first technical result ensures that \mathcal{W}_γ is non-negative.

Lemma 16. *Assume A1 and D2. Then, for any $x, v \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}_W]$ with*

$$\bar{\gamma}_W = \min \left(1, \bar{\gamma}, \left(\frac{c_W}{\kappa \vartheta + (1 + \vartheta)(2C_\kappa + \kappa^2)} \right)^{(\delta \wedge 1)^{-1}} \right), \quad c_W = \frac{1}{2} \min \left(\frac{\kappa^2}{6}, \frac{1}{4} \right), \quad (3.89)$$

it holds $\mathcal{W}_\gamma(x, v) \geq c_W(\|x\|^2 + \|v\|^2) + 2\alpha_U U(x)$, where \mathcal{W}_γ is defined in (3.73).

The next result provides an upper bound on ϕ_γ , itself obtained from an upper bound on \mathcal{W}_γ .

Lemma 17. *Assume **A1**, **D1** and **D2**. Then, for any $\gamma \in (0, \min(1, \bar{\gamma})]$ and $x, v \in \mathbb{R}^d$,*

$$\mathcal{W}_\gamma(x, v) \leq \mathfrak{C}_\phi^2(\|x\|^2 + \|v\|^2), \quad \phi_\gamma(x, v) \leq 1 + \mathfrak{C}_\phi(\|x\| + \|v\|), \quad (3.90)$$

where

$$\mathfrak{C}_\phi = \sqrt{\max\left(1, \frac{\kappa^2}{2} + \alpha_U L\right) + \frac{1}{2}(1 + \bar{\vartheta})(\kappa^2 + \kappa + 2C_\kappa)}.$$

The last two results provide Lipschitz bounds which allow to obtain (3.83).

Lemma 18. *Assume that **A1**, **D1** and **D2** hold. For any $\gamma \in (0, \bar{\gamma}_\mathcal{W}]$, the function $\phi_\gamma \in C^1(\mathbb{R}^{2d}, \mathbb{R})$ defined by (3.82) is Lipschitz continuous, and its Lipschitz constant is uniformly bounded by*

$$\mathcal{L}_\phi = \frac{1}{\sqrt{c_\mathcal{W}}} \max\left(2, 2\alpha_U L + \kappa^2, (1 + \bar{\vartheta})(\kappa^2 + \kappa + C_\kappa)\right),$$

where $\bar{\gamma}_\mathcal{W}, c_\mathcal{W}$ are defined in (3.89).

Lemma 19. *Assume that **A1** and **D3** hold, and recall that Γ_γ is defined in (3.13). Then, for any $\gamma \in (0, \min(1, \bar{\gamma})]$, $x, v \in \mathbb{R}^d$ and $w_2 \in \mathbb{R}^m$, the function $(z, w_1) \mapsto \Gamma_\gamma(x, v, \sqrt{\gamma}\sigma_\gamma z, (w_1, w_2))$ is Lipschitz continuous, and its Lipschitz constant is bounded by $\sqrt{2\bar{\gamma}}[2\bar{L}\max(1, \bar{\sigma}) + \mathcal{D}\bar{\sigma} + \bar{\sigma}]$.*

3.6 Postponed proofs

3.6.1 Proof of Lemma 6

Note that by (3.9) and using that for any $t \in \mathbb{R}_+$, $t - t^2/2 + t^3/6 - t^4/24 \leq 1 - e^{-t} \leq t - t^2/2 + t^3/6$, we have for any $t_0 \in \mathbb{R}_+$,

$$\sigma^2 t_0^3/3 - \sigma^2 \kappa t_0^4/3 \leq \Sigma_1^{(t_0)} \leq \sigma^2 t_0^3/3 + \sigma^2 \kappa t_0^4/12.$$

Similarly using that for any $t \in \mathbb{R}_+$, $t - t^2/2 \leq 1 - e^{-t} \leq t$, we have for any $t_0 \in \mathbb{R}_+$,

$$\sigma^2 t_0^2/2 - \sigma^2 \kappa t_0^3/2 \leq \Sigma_2^{(t_0)} \leq \sigma^2 t_0^2/2, \quad \sigma^2 t_0 - \sigma^2 \kappa t_0^2 \leq \Sigma_3^{(t_0)} \leq \sigma^2 t_0.$$

Then, since $1/3 - 1/4 > 0$, taking \bar{t}_0 sufficiently small completes the proof.

3.6.2 Proof of Lemma 8

The first estimate is a direct consequence of the following inequality for $0 \leq a \leq b$ and $\ell \geq 1$:

$$0 \leq b^\ell - a^\ell = \ell \int_a^b x^{\ell-1} dx \leq (b-a)\ell b^{\ell-1},$$

together with the bound $\max(\tau_\gamma, e^{-\kappa\gamma}) \leq 1$. The bound on $|\tau_\gamma - 1|$ follows from the fact that this quantity is bounded by $1 - e^{-\kappa\gamma} + C_\kappa\gamma^2$ in view of **A1**, together with the inequality $1 - e^{-\kappa\gamma} \leq \kappa\gamma$. For the final estimate, we write

$$\left| \frac{\gamma}{1 - \tau_\gamma} - \frac{1}{\kappa} \right| = \frac{|\tau_\gamma - 1 + \kappa\gamma|}{\kappa(1 - \tau_\gamma)} \leq \frac{|\tau_\gamma - e^{-\kappa\gamma}| + |e^{-\kappa\gamma} - 1 + \kappa\gamma|}{\kappa(1 - \tau_\gamma)}.$$

The first term in the last numerator is bounded by $C_\kappa\gamma^2$ in view of **A1**. For the second one and the denominator, we use the inequality $-t^2/2 \leq 1 - t - e^{-t} \leq 0$ for any $t \geq 0$ to write $|e^{-\kappa\gamma} - 1 + \kappa\gamma| \leq \kappa^2\gamma^2/2$ and

$$\frac{1 - \tau_\gamma}{\gamma} \geq \frac{1 - e^{-\kappa\gamma} - C_\kappa\gamma^2}{\gamma} \geq \kappa - \left(\frac{\kappa^2}{2} + C_\kappa \right) \gamma \geq \frac{\kappa}{2},$$

where the last inequality follows from the bound $\gamma \leq \bar{\gamma} \leq (\kappa + 2C_\kappa/\kappa)^{-1}$ in **A1**. This finally leads to (3.38).

3.6.3 Proof of Lemma 16

Let $\gamma \in (0, \bar{\gamma}_W]$. Consider $V_0(x, v) = \kappa^2 \|x\|^2/2 + \|v\|^2 + \kappa \langle x, v \rangle + 2\alpha_U U(x)$. The Cauchy–Schwarz inequality and Lemma 8 give

$$\begin{aligned} |\mathcal{W}_\gamma(x, v) - V_0(x, v)| &\leq \kappa \left| \frac{\kappa\gamma(1 + \gamma^\delta \vartheta_\gamma)}{1 - \tau_\gamma} - 1 \right| \frac{\|x\|^2 + \|v\|^2}{2} \\ &\leq \kappa \left(\gamma^\delta \vartheta_\gamma + (1 + \gamma^\delta \vartheta_\gamma) \left| \frac{\kappa\gamma}{1 - \tau_\gamma} - 1 \right| \right) \frac{\|x\|^2 + \|v\|^2}{2} \\ &\leq \kappa \left(\gamma^\delta \bar{\vartheta} + (1 + \gamma^\delta \bar{\vartheta}) \left[\frac{2C_\kappa}{\kappa} + \kappa \right] \gamma \right) \frac{\|x\|^2 + \|v\|^2}{2} \\ &\leq c_W (\|x\|^2 + \|v\|^2), \end{aligned}$$

where the last inequality follows from the definition of $\bar{\gamma}_W$. In addition, using the Cauchy–Schwarz inequality again, we get for any $\eta \in (1/2, 1)$,

$$V_0(x, v) \geq \frac{\kappa^2}{2} (1 - \eta) \|x\|^2 + \left(1 - \frac{1}{2\eta} \right) \|v\|^2 + 2\alpha_U U(x) \geq 2c_W (\|x\|^2 + \|v\|^2) + 2\alpha_U U(x),$$

where the last inequality is obtained with $\eta = 2/3$. The combination of the two previous inequalities finally gives the claimed result.

3.6.4 Proof of Lemma 17

By [106, Lemma 1.2.3] and **D1**, it holds $U(x) \leq L\|x\|^2/2$. Moreover, the Cauchy–Schwarz inequality, Lemma 8 and the last condition in **D2** lead to

$$\begin{aligned} \frac{\kappa^2\gamma(1+\gamma^\delta\vartheta_\gamma)}{1-\tau_\gamma} \langle x, v \rangle &\leq \frac{\kappa^2}{2}(1+\bar{\vartheta}) \left(\frac{1}{\kappa} + \left[\frac{2C_\kappa}{\kappa^2} + 1 \right] \gamma \right) (\|x\|^2 + \|v\|^2) \\ &\leq \frac{1}{2}(1+\bar{\vartheta})(\kappa^2 + \kappa + 2C_\kappa)(\|x\|^2 + \|v\|^2), \end{aligned} \quad (3.91)$$

where we used $\gamma \leq 1$ in the last inequality. This finally implies the first inequality in (3.90) by the definition (3.73) of \mathcal{W}_γ . The proof of the second one is concluded with the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$.

3.6.5 Proof of Lemma 18

In view of the definitions (3.73) and (3.82), it holds

$$\nabla_x \phi_\gamma(x, v) = \frac{1}{2\phi_\gamma(x, v)} \left[\kappa^2 x + \frac{\kappa^2\gamma(1+\gamma^\delta\vartheta_\gamma)}{1-\tau_\gamma} v + 2\alpha_U \nabla U(x) \right],$$

so that, by a triangle inequality and upon bounding the prefactor of v as in (3.91), and using also Lemma 16 and the inequality $\|\nabla U(x)\| \leq L\|x\|$, we obtain

$$\begin{aligned} \|\nabla_x \phi_\gamma(x, v)\| &\leq \frac{\kappa^2 \|x\| + (1+\bar{\vartheta})(\kappa^2 + \kappa + 2C_\kappa) \|v\| + 2\alpha_U \|\nabla U(x)\|}{2\sqrt{1+c_{\mathcal{W}}\{\|x\|^2 + \|v\|^2\}}} \\ &\leq \frac{1}{\sqrt{c_{\mathcal{W}}}} \max \left(\alpha_U L + \frac{\kappa^2}{2}, \frac{1}{2}(1+\bar{\vartheta})(\kappa^2 + \kappa + 2C_\kappa) \right). \end{aligned}$$

Similarly, for any $x, v \in \mathbb{R}^d$,

$$\|\nabla_v \phi_\gamma(x, v)\| \leq \frac{1}{\sqrt{c_{\mathcal{W}}}} \max \left(1, \frac{1}{2}(1+\bar{\vartheta})(\kappa^2 + \kappa + 2C_\kappa) \right).$$

The conclusion then follows from the inequality $\|\nabla \phi_\gamma(x, v)\| \leq 2 \max(\|\nabla_x \phi_\gamma(x, v)\|, \|\nabla_v \phi_\gamma(x, v)\|)$ and the mean value theorem.

3.6.6 Proof of Lemma 19

In view of (3.13), A1 and D3-1)-(b), we can write, for any $x, v, z, z' \in \mathbb{R}^d$, $w_1, w'_1 \in \mathbb{R}^{m_1}$ and $w_2 \in \mathbb{R}^{m_2}$,

$$\begin{aligned}
& \left\| \Gamma_\gamma(x, v, \sqrt{\gamma}\sigma_\gamma z, (w_1, w_2)) - \Gamma_\gamma(x, v, \sqrt{\gamma}\sigma_\gamma z', (w'_1, w_2)) \right\| \\
& \leq \gamma \left\| f_\gamma(x, \gamma^\delta v, \gamma^{1/2+\delta}\sigma_\gamma z, (w_1, w_2)) - f_\gamma(x, \gamma^\delta v, \gamma^{1/2+\delta}\sigma_\gamma z', (w'_1, w_2)) \right\| \\
& \quad + \gamma \left\| g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2}\sigma_\gamma z, (w_1, w_2)) - g_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2}\sigma_\gamma z', (w'_1, w_2)) \right\| \\
& \quad + \gamma^{\delta+1/2}\sigma_\gamma \left\| \mathbf{D}_\gamma(z - z') \right\| + \sigma_\gamma \sqrt{\gamma} \|z - z'\| \\
& \leq \left(2\tilde{L}\gamma \max(1, \gamma^{\delta+1/2}\bar{\sigma}) + \gamma^{\delta+1/2}\mathcal{D}\bar{\sigma} + \bar{\sigma}\sqrt{\gamma} \right) (\|z - z'\| + \|w_1 - w'_1\|), \\
& \leq \sqrt{2} \left(2\tilde{L}\gamma \max(1, \gamma^{\delta+1/2}\bar{\sigma}) + \gamma^{\delta+1/2}\mathcal{D}\bar{\sigma} + \bar{\sigma}\sqrt{\gamma} \right) \|(z, w_1) - (z', w'_1)\|,
\end{aligned}$$

which completes the proof.

3.6.7 Proof of Proposition 3.2

In view of (3.12) and since $b = -\nabla U$,

$$f_\gamma(x, v, z, w) = \mathcal{C}_{1,\gamma}v - \frac{\gamma}{2}\nabla U\left(x + \mathcal{C}_{2,\gamma}v + \mathcal{C}_{3,\gamma}z + \gamma^{3/2}\mathcal{C}_{4,\gamma}w\right) + 2\gamma^{3/2}\mathcal{C}_{4,\gamma}w,$$

with

$$\mathcal{C}_{1,\gamma} = \frac{e^{-\kappa\gamma/2} - 1}{\gamma}, \quad \mathcal{C}_{2,\gamma} = \frac{e^{-\kappa\gamma/2}}{2}, \quad \mathcal{C}_{3,\gamma} = \frac{e^{-\kappa\gamma/2}}{2(1 + e^{-\kappa\gamma})}, \quad \mathcal{C}_{4,\gamma} = \sqrt{\frac{\tilde{\sigma}_{\gamma/2}^2}{8(1 + e^{-\kappa\gamma})}}.$$

The coefficients $\mathcal{C}_{i,\gamma}$ (for $1 \leq i \leq 4$) are uniformly bounded in γ for $\gamma \in (0, \bar{\gamma}]$, and we denote by $\bar{\mathcal{C}}$ their maximal value:

$$\bar{\mathcal{C}} = \sup_{1 \leq i \leq 4} \sup_{\gamma \in (0, \bar{\gamma}]} \mathcal{C}_{i,\gamma} < +\infty. \quad (3.92)$$

A Cauchy–Schwarz inequality gives

$$\begin{aligned}
\left\| f_\gamma(x, \gamma^\delta v, \gamma^{\delta+1/2}\sigma_\gamma z, w) \right\|^2 & \leq 3\gamma^{2\delta}\bar{\mathcal{C}}^2 \|v\|^2 + 12\gamma^3\bar{\mathcal{C}}^2 \|w\|^2 \\
& \quad + \frac{3\gamma^2}{4} \left\| \nabla U\left(x + \gamma^\delta\mathcal{C}_{2,\gamma}v + \gamma^{\delta+1/2}\sigma_\gamma\mathcal{C}_{3,\gamma}z + \gamma^{3/2}\mathcal{C}_{4,\gamma}w\right) \right\|^2.
\end{aligned}$$

It therefore suffices to bound the term on the second line of the previous inequality. To this end, we note that D1 and a Cauchy–Schwarz inequality imply that, for any $h \in \mathbb{R}^d$,

$$\|\nabla U(x + h)\|^2 \leq 2\|\nabla U(x)\|^2 + 2L^2\|h\|^2. \quad (3.93)$$

The first condition in D2 is then easily seen to hold upon setting $h = \gamma^\delta\mathcal{C}_{2,\gamma}v + \gamma^{\delta+1/2}\sigma_\gamma\mathcal{C}_{3,\gamma}z + \gamma^{3/2}\mathcal{C}_{4,\gamma}w$.

To prove that the second condition in **D2** holds, we need to be careful about the dependence of our estimates on $\|x\|$. We rely on (3.18), which implies that there exist $a > 0$ and $b \in \mathbb{R}$ such that

$$-\langle x, \nabla U(x) \rangle \leq -a \left(\|x\| + \|\nabla U(x)\|^2 \right) + b.$$

An inequality similar to (3.93) can also be written for any $h \in \mathbb{R}^d$:

$$-2 \|\nabla U(x+h)\|^2 \leq -\|\nabla U(x)\|^2 + 2L^2 \|h\|^2.$$

We therefore obtain, using a Cauchy–Schwarz inequality, for any $h \in \mathbb{R}^d$,

$$\begin{aligned} -\langle x, \nabla U(x+h) \rangle &= -\langle x+h, \nabla U(x+h) \rangle + \langle h, \nabla U(x+h) \rangle \\ &\leq -a \left(\|x+h\| + \|\nabla U(x+h)\|^2 \right) + b + \frac{a}{2} \left(\frac{\|h\|^2}{a^2} + \|\nabla U(x+h)\|^2 \right) \\ &\leq -a \left(\|x\| + \frac{1}{4} \|\nabla U(x)\|^2 \right) + \tilde{b} \left(1 + \|h\|^2 \right), \end{aligned} \quad (3.94)$$

for some constant $\tilde{b} \in \mathbb{R}$. The second condition in **D2** then follows from the above inequality, the fact that

$$\left\langle x, f_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w \right) \right\rangle = \gamma^\delta \mathcal{C}_{1,\gamma} \langle x, v \rangle + 2\gamma^{3/2} \mathcal{C}_{4,\gamma} \langle x, w \rangle - \frac{\gamma}{2} \langle x, \nabla U(x+h) \rangle$$

with $h = \gamma^\delta \mathcal{C}_{2,\gamma} v + \gamma^{\delta+1/2} \sigma_\gamma \mathcal{C}_{3,\gamma} z + \gamma^{3/2} \mathcal{C}_{4,\gamma} w$ and using (3.92).

Let us next check that the conditions in **D2** involving g_γ are satisfied with $\alpha_U = 1$. In view of (3.12) and since $b = -\nabla U$,

$$g_\gamma(x, v, z, w) = -\mathcal{G}_{1,\gamma} \nabla U \left(x + \mathcal{G}_{2,\gamma} v + \mathcal{G}_{3,\gamma} z + \gamma^{3/2} \mathcal{G}_{4,\gamma} w \right),$$

with

$$\mathcal{G}_{1,\gamma} = e^{-\kappa\gamma/2}, \quad \mathcal{G}_{2,\gamma} = \frac{e^{-\kappa\gamma/2}}{2}, \quad \mathcal{G}_{3,\gamma} = \frac{e^{-\kappa\gamma/2}}{2(1+e^{-\kappa\gamma})}, \quad \mathcal{G}_{4,\gamma} = \sqrt{\frac{\tilde{\sigma}_{\gamma/2}^2}{8(1+e^{-\kappa\gamma})}}.$$

The coefficients $\mathcal{G}_{i,\gamma}$ (for $1 \leq i \leq 4$) are uniformly bounded in γ for $\gamma \in (0, \bar{\gamma}]$, and we denote by $\bar{\mathcal{G}}$ their maximal value:

$$\bar{\mathcal{G}} = \sup_{1 \leq i \leq 4} \sup_{\gamma \in (0, \bar{\gamma}]} \mathcal{G}_{i,\gamma} < +\infty. \quad (3.95)$$

Note also that, there exists $K \geq 0$ such that for any $\gamma \in (0, \bar{\gamma}]$,

$$|\mathcal{G}_{1,\gamma} - 1| \leq K\gamma, \quad (3.96)$$

3.7. Complementary and technical results

so that we bound using **D1** the term involving g_γ in the first condition as

$$\begin{aligned} \left\| g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w \right) + \nabla U(x) \right\|^2 &\leq 2(1 - \mathcal{G}_{1,\gamma})^2 \|\nabla U(x)\|^2 \\ &\quad + 2\overline{\mathcal{G}}^2 \|\nabla U(x+h) - \nabla U(x)\|^2 \\ &\leq 2K^2\gamma^2 \|\nabla U(x)\|^2 + 2\overline{\mathcal{G}}^2 L^2 \|h\|^2, \end{aligned}$$

with $h = \gamma^\delta \mathcal{G}_{2,\gamma} v + \gamma^{\delta+1/2} \sigma_\gamma \mathcal{G}_{3,\gamma} z + \gamma^{3/2} \mathcal{G}_{4,\gamma} w$, which easily implies that the second condition holds by (3.95). Moreover, with the same definition of h ,

$$\left\langle x, g_\gamma \left(x, \gamma^\delta v, \gamma^{\delta+1/2} \sigma_\gamma z, w \right) \right\rangle = -\mathcal{G}_{1,\gamma} \langle x, \nabla U(x+h) \rangle,$$

from which the third condition easily follows in view of (3.94), (3.95) and (3.96).

3.7 Complementary and technical results

Lemma 20. *For any $\kappa, \sigma, \gamma > 0$, $\Sigma^{(\gamma)} \otimes \text{Id}$ in (3.9) is invertible.*

Proof. Note first that we only need to consider the case $\sigma = \kappa = 1$. Let $\gamma > 0$. Set $M = \Sigma^{(\gamma)} \otimes \text{Id}$. We show that $\det(M) > 0$. Since by [62, Exercise 45, Chapter 1], $\det(M) = \det(\Sigma^{(\gamma)})^d$ where $\Sigma^{(\gamma)}$ is given by (3.9), it suffices to show that $\det(\Sigma^{(\gamma)}) > 0$. Denote for any $t > 0$ and square-integrable functions $h_1, h_2 : [0, t] \rightarrow \mathbb{R}$,

$$\langle h_1, h_2 \rangle_{L^2([0,t])} = \int_0^t h_1(s) h_2(s) ds, \quad \|h_1\|_{L^2([0,t])} = \sqrt{\langle h_1, h_1 \rangle_2}.$$

By (3.9),

$$\begin{aligned} \det(\Sigma^{(\gamma)}) &= \det \begin{pmatrix} \|h_1\|_{L^2([0,\gamma])}^2 & \langle h_1, h_2 \rangle_{L^2([0,\gamma])} \\ \langle h_1, h_2 \rangle_{L^2([0,\gamma])} & \|h_2\|_{L^2([0,\gamma])}^2 \end{pmatrix} \\ &= \|h_1\|_{L^2([0,\gamma])}^2 \|h_2\|_{L^2([0,\gamma])}^2 - \langle h_1, h_2 \rangle_{L^2([0,\gamma])}^2, \end{aligned}$$

where for any $s \in [0, \gamma]$, $h_1(s) = 1 - e^{-(\gamma-s)}$ and $h_2(s) = e^{-(\gamma-s)}$. The result follows by a Cauchy–Schwarz inequality since h_1, h_2 are linearly independent. \square

Lemma 21. *Let $x, v \in \mathbb{R}^d$ and $\kappa > 0$. Consider for any $t \in \mathbb{R}_+$,*

$$\begin{aligned} \tilde{\mathbf{X}}_t &= x + \frac{1 - e^{-\kappa t}}{\kappa} v + \frac{\kappa t + e^{-\kappa t} - 1}{\kappa^2} b(x) + \sigma \int_0^t \frac{1 - e^{-\kappa(t-s)}}{\kappa} dB_s, \\ \tilde{\mathbf{V}}_t &= e^{-\kappa t} v + \frac{1 - e^{-\kappa t}}{\kappa} b(x) + \sigma \int_0^t e^{-\kappa(t-s)} dB_s, \end{aligned} \tag{3.97}$$

where $(B_t)_{t \geq 0}$ is a standard d -dimensional Brownian motion. The process $(\tilde{\mathbf{X}}_t, \tilde{\mathbf{V}}_t)_{t \geq 0}$ is the unique solution of the SDE,

$$\tilde{\mathbf{X}}_t = x + \int_0^t \tilde{\mathbf{V}}_s ds, \quad \tilde{\mathbf{V}}_t = v + \int_0^t \left\{ -\kappa \tilde{\mathbf{V}}_s + b(x) \right\} ds + \sigma B_t.$$

Proof. For any $t \in \mathbb{R}_+$, by (3.97), Lemma 22 and linearity,

$$\begin{aligned} \int_0^t \tilde{\mathbf{V}}_s \, ds &= \int_0^t e^{-\kappa s} v + \frac{1 - e^{-\kappa s}}{\kappa} b(x) \, ds + \sigma \int_0^t \int_0^s e^{-\kappa(s-r)} \, dB_r \, ds \\ &= \frac{1 - e^{-\kappa t}}{\kappa} v + \frac{\kappa t + e^{-\kappa t} - 1}{\kappa^2} b(x) + \sigma \int_0^t \int_r^t e^{-\kappa(s-r)} \, ds \, dB_r \\ &= \tilde{\mathbf{X}}_t - x. \end{aligned} \quad (3.98)$$

In addition, using (3.98), we obtain

$$\begin{aligned} \int_0^t -\kappa \tilde{\mathbf{V}}_s + b(x) \, ds &= (e^{-\kappa t} - 1)v - \frac{e^{-\kappa t} - 1}{\kappa} b(x) + \sigma \int_0^t (e^{-\kappa(t-s)} - 1) \, dB_s \\ &= \tilde{\mathbf{V}}_t - v - \sigma \int_0^t \, dB_s, \end{aligned}$$

which completes the proof. \square

The following Fubini-type result for stochastic integrals is established in [72, Theorem 1] (see also [116, Chapter IV, Exercise (5.17)]), but an alternative proof is given here for completeness.

Lemma 22. *For any $f \in C^1(\mathbb{R})$, $g \in C^0(\mathbb{R})$, $u, v \geq 0$,*

$$\int_u^v \int_u^v \mathbb{1}_{\mathbb{R}_+}(t-s) f(s) g(t) \, dB_s \, dt = \int_u^v \int_u^v \mathbb{1}_{\mathbb{R}_+}(t-s) f(s) g(t) \, dt \, dB_s. \quad (3.99)$$

Proof. Consider $f \in C^1(\mathbb{R})$, $g \in C^0(\mathbb{R})$, $v \geq 0$. Without loss of generality, it is sufficient to show (3.99) for $u = 0$. Introduce, for any $w \geq 0$, $G(w) = \int_0^w g(t) \, dt$ and $M_w = f(w)B_w$. By integration by parts [116, Chapter IV, Proposition (3.1)], we have, for any $w \geq 0$,

$$M_w = \int_0^w f'(s) B_s \, ds + \int_0^w f(s) \, dB_s, \quad (3.100)$$

$$G(w)M_w = \int_0^w g(s) f(s) B_s \, ds + \int_0^w G(s) \, dM_s. \quad (3.101)$$

Then, by (3.100),

$$\begin{aligned} \int_0^v \int_0^v \mathbb{1}_{\mathbb{R}_+}(t-s) f(s) g(t) \, dB_s \, dt &= \int_0^v \left(\int_0^t f(s) \, dB_s \right) g(t) \, dt \\ &= \int_0^v \left(M_t - \int_0^t f'(s) B_s \, ds \right) g(t) \, dt, \end{aligned} \quad (3.102)$$

and

$$\begin{aligned} \int_0^v \int_0^v \mathbb{1}_{\mathbb{R}_+}(t-s) f(s) g(t) \, dt \, dB_s &= \int_0^v (G(v) - G(s)) f(s) \, dB_s \\ &= \int_0^v (G(v) - G(s)) \, dM_s \\ &\quad - \int_0^v (G(v) - G(s)) f'(s) B_s \, ds. \end{aligned} \quad (3.103)$$

3.7. Complementary and technical results

By Fubini's theorem, almost surely we have $\int_0^v \int_0^t f'(s) B_s ds g(t) dt = \int_0^v (G(v) - G(s)) f'(s) B_s ds$. Therefore using this result and (3.102)-(3.103), (3.99) holds if $\int_0^v M_t g(t) dt = \int_0^v (G(v) - G(t)) dM_t$. which follows from (3.101) as

$$\int_0^v (G(v) - G(t)) dM_t = G(v)M_v - \int_0^v G(t) dM_t = \int_0^v g(t)f(t)B_t dt = \int_0^v M_t g(t) dt.$$

This allows to conclude the proof. □

CHAPTER 4

Discrete sticky couplings of functional autoregressive processes

ALAIN DURMUS¹, ANDREAS EBERLE², AURÉLIEN ENFROY³, ARNAUD GUILLIN⁴
AND PIERRE MONMARCHÉ⁵

In this paper, we provide bounds in Wasserstein and total variation distances between the distributions of the successive iterates of two functional autoregressive processes with isotropic Gaussian noise of the form $Y_{k+1} = T_\gamma(Y_k) + \sqrt{\gamma\sigma^2}Z_{k+1}$ and $\tilde{Y}_{k+1} = \tilde{T}_\gamma(\tilde{Y}_k) + \sqrt{\gamma\sigma^2}\tilde{Z}_{k+1}$. More precisely, we give non-asymptotic bounds on $\rho(\mathcal{L}(Y_k), \mathcal{L}(\tilde{Y}_k))$, where ρ is an appropriate weighted Wasserstein distance or a V -distance, uniformly in the parameter γ , and on $\rho(\pi_\gamma, \tilde{\pi}_\gamma)$, where π_γ and $\tilde{\pi}_\gamma$ are the respective stationary measures of the two processes. The class of considered processes encompasses the Euler-Maruyama discretization of Langevin diffusions and its variants. The bounds we derive are of order γ as $\gamma \rightarrow 0$. To obtain our results, we rely on the construction of a discrete sticky Markov chain $(W_k^{(\gamma)})_{k \in \mathbb{N}}$ which bounds the distance between an appropriate coupling of the two processes. We then establish stability and quantitative convergence results for this process uniformly on γ . In addition, we show that it converges in distribution to the continuous sticky process studied in [51, 68]. Finally, we apply our result to Bayesian inference of ODE parameters and numerically illustrate them on two particular problems.

¹Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, F-91190 Gif-sur-Yvette, France.
Email: alain.durmus@ens-paris-saclay.fr

²Institute for Applied Mathematics - University of Bonn, Germany. Email: eberle@uni-bonn.de

³Samovar, Télécom SudParis, département CITI, TIPIC, Institut Polytechnique de Paris, Palaiseau.
Email: aurelien.enfroy@ens-paris-saclay.fr

⁴Laboratoire de Mathématiques Blaise Pascal - Université Clermont-Auvergne, France.
Email: arnaud.guillin@uca.fr

⁵LJLL - Sorbonne Université, France. Email: pierre.monmarche@sorbonne-universite.fr

4.1 Introduction

We are interested in this paper in Markov chains $(Y_k)_{k \in \mathbb{N}}$ starting from $y \in \mathbb{R}^d$ and defined by recursions of the form

$$Y_{k+1} = T_\gamma(Y_k) + \sigma\sqrt{\gamma}Z_{k+1}, \quad (4.1)$$

where $\sigma > 0$, $\gamma \in (0, \bar{\gamma}]$, for some $\bar{\gamma} > 0$, $\{T_\gamma : \gamma \in (0, \bar{\gamma}]\}$ is a family of continuous functions from \mathbb{R}^d to \mathbb{R}^d and $(Z_k)_{k \geq 1}$ is a sequence of i.i.d. d -dimensional standard Gaussian random variables. Note that the Euler-Maruyama discretization of overdamped Langevin diffusions or of general Komolgorov processes and its variants belong to this class of processes and in that setting γ corresponds to the discretization step size. Indeed, the Euler scheme consists in taking for any $\gamma \in (0, \bar{\gamma}]$, $T_\gamma(y) = y + \gamma b(y)$ for some $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$. When $b = -\nabla U$ for some potential U , these methods are now popular Markov Chain Monte Carlo algorithms to sample from the target density $x \mapsto e^{-U(x)} / \int_{\mathbb{R}^d} e^{-U(y)} dy$. However, in some applications, explicitly computing ∇U is not an option and further numerical methods must be implemented which come with additional bias since only approximations of ∇U can be used in (4.1). In this paper, we precisely study this additional source of error. In particular, based on a chain defined by (4.1), we consider a second Markov chain $(\tilde{Y}_k)_{k \in \mathbb{N}}$ defined by the recursion

$$\tilde{Y}_{k+1} = \tilde{T}_\gamma(\tilde{Y}_k) + \sigma\sqrt{\gamma}\tilde{Z}_{k+1}, \quad (4.2)$$

where $\{\tilde{T}_\gamma : \gamma \in (0, \bar{\gamma}]\}$ is a family of functions from \mathbb{R}^d to \mathbb{R}^d such that for any γ , \tilde{T}_γ is an approximation of T_γ in a sense specified below, and $(\tilde{Z}_k)_{k \geq 1}$ is a sequence of i.i.d. d -dimensional standard Gaussian random variables potentially correlated with $(Z_k)_{k \geq 1}$.

We will enforce below conditions that ensure that both $(Y_k)_{k \in \mathbb{N}}$ and $(\tilde{Y}_k)_{k \geq 1}$ are geometrically ergodic, and denote by π_γ and $\tilde{\pi}_\gamma$ their invariant probability measures respectively. If for any $\gamma > 0$, \tilde{T}_γ is close in some sense to T_γ , the overall process $(\tilde{Y}_k)_{k \in \mathbb{N}}$ can be seen as a perturbed version of $(Y_k)_{k \in \mathbb{N}}$, and $\tilde{\pi}_\gamma$ is expected to be close to π_γ . The main goal of this paper is to establish quantitative bounds on the Wasserstein and total variation distance between the finite-time laws of the two processes and between their equilibria. The study of perturbation of Markov processes has been the subject of many existing works; see e.g. [71, 97, 102, 123, 125] and the references therein. However, it turns out that these existing results do not apply as such. We pay particular attention to the dependency of these estimates on γ . Indeed, in the case of the Euler scheme of a continuous-time diffusion, π_γ and the law of $Y_{[t/\gamma]}$ for some $t > 0$ converge to the invariant measure and law at time t of the continuous-time process, and similarly for the perturbed chain. Hence, as $\gamma \rightarrow 0$, our estimates should not degenerate, but rather yield quantitative estimates for the continuous time process. More precisely, the present paper is the discrete-time counterpart of the study conducted by [51] in the continuous-time case, and as γ vanishes we recover estimates that are consistent with those of [51].

As in [51], our results are based on the construction of a suitable coupling of the processes, i.e. a simultaneous construction of a pair $(Y_k, \tilde{Y}_k)_{k \in \mathbb{N}}$ of non-independent

chains that marginally follow (4.1) and (4.2) respectively and are designed to get and stay close to each other. We use the maximal reflection coupling for Gaussian laws, namely at each step the two chains are coupled to merge with maximal probability and, otherwise, we use a reflection (see Section 4.2.2 below). Estimates on the laws of the chains then follow from the study of $(\|Y_k - \tilde{Y}_k\|)_{k \in \mathbb{N}}$, which is itself based on the analysis of a Markov chain $(W_k)_{k \in \mathbb{N}}$ on $[0, +\infty)$ that is such that, by design of the coupling, $\|Y_k - \tilde{Y}_k\|_k \leq W_k$ for all $k \in \mathbb{N}$. Thus, the question of establishing bounds between the laws of two d -dimensional Markov chains is reduced to the study of a single one-dimensional chain. Besides, together with the Markov property, the auxiliary chain has some nice features. At first, it is stochastically monotonous, *i.e.* if $(W'_k)_{k \in \mathbb{N}}$ is a Markov chain associated to the same Markov kernel as $(W_k)_{k \in \mathbb{N}}$ and such that $W_0 \leq W'_0$, then for any $k \in \mathbb{N}$, W'_k is stochastically dominated by W_k , *i.e.* for any $t \geq 0$, $\mathbb{P}(W_k \leq t) \geq \mathbb{P}(W'_k \leq t)$. Secondly, $(W_k)_{k \in \mathbb{N}}$ has an atom at 0.

The main results and main steps of this study are the following. First, we prove that $(W_k)_{k \in \mathbb{N}}$ admits a unique invariant measure and that, independently of γ , the moments and mass on $(0, +\infty)$ of this equilibrium are small when the difference between T_γ and \tilde{T}_γ is small. Secondly, we establish the geometric convergence of the chain towards its equilibrium, at an explicit rate (stable as $\gamma \rightarrow 0$). Finally, we prove that, as $\gamma \rightarrow 0$, the chain $(W_k)_{k \in \mathbb{N}}$ converges in law to the continuous-time sticky diffusion that played the same role in [51]. This last part is not necessary to get estimates on the finite-time and equilibrium laws of (4.1) and (4.2) for a given $\gamma > 0$, but it sheds some new light on the limit sticky process which, in [51], is constructed as the limit of continuous-time diffusions with diffusion coefficients that vanish at zero, rather than discrete-time chains. In some sense, $(W_k)_{k \in \mathbb{N}}$ can be seen as a discretization scheme for the sticky process, see also [6] on this topic.

Besides the obvious continuous/discrete time difference between [51] and the present work, let us emphasize a few other distinctions. First, in [51], the one-dimensional sticky process has an explicit invariant measure. This is not the case in our framework, which makes the derivation of the bounds on the moments of the equilibrium a bit more involved. Secondly, in [51], although it is proven that the mass at zero and the first moment of the law of the sticky diffusion converge to their value at equilibrium (which is sufficient to get estimates on the laws of the two initial d -dimensional processes), the question of long-time convergence is not addressed for the sticky diffusion, whereas our long-time convergence results for $(W_k)_{k \geq 0}$ together with its convergence as $\gamma \rightarrow +\infty$ furnish an explicit convergence rate for the sticky diffusion. The proof of the stability of the mass at zero and of the first moment in [51] relies on a concave modification of the distance (such as used *e.g.* in [48]), which is contracted by the chain before it hits zero. This method does not apply to, say, the second moment of the process. As a consequence, the results of [51] only concern the total variation and \mathbf{W}_1 Wasserstein distances, while we consider a broader class of distances.

Finally, our theoretical results are illustrated through numerical experiments. In particular, we study the influence of the discretization scheme generally needed to perform Bayesian inference for parameters of Ordinary Differential Equations (ODEs).

Notation and convention

We denote by $\mathcal{B}(\mathbb{R}^d)$, the Borel σ -field of \mathbb{R}^d endowed with the Euclidean distance and by φ_{σ^2} the density of the one-dimensional Gaussian distribution with zero-mean and variance $\sigma^2 > 0$. In the case $\sigma = 1$, we simply denote this density by φ . $\Delta_{\mathbb{R}^d}$ stands for the subset $\{(x, x) \in \mathbb{R}^{2d} : x \in \mathbb{R}^d\}$ of \mathbb{R}^d and for any $A \subset \mathbb{R}^d$, A^c for its complement. Let μ and ν be two σ -finite measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. If ν is absolutely continuous with respect to μ , we write $\nu \ll \mu$. We say that ν and μ are equivalent if and only if $\nu \ll \mu$ and $\mu \ll \nu$. We denote by $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ the floor and ceiling function respectively. For $d, n \in \mathbb{N}^*$, $\mathcal{M}_{d,n}(\mathbb{R})$ stands for the set of $d \times n$ real matrices. We denote by $C^k(\mathcal{U}, \mathcal{A})$ the set of k times continuously differentiable functions from an open set $\mathcal{U} \subset \mathbb{R}^m$ to $\mathcal{A} \subset \mathbb{R}^p$. We use the convention $\sum_{k=n}^p = 0$ and $\prod_{k=n}^p = 1$ for $p < n$, $n, p \in \mathbb{N}$, and $a/0 = +\infty$ for $a > 0$.

4.2 Sticky reflection coupling

4.2.1 Main result

The Markov kernels R_γ associated with $(Y_k)_{k \in \mathbb{N}}$ defined in (4.1) are given for any $y \in \mathbb{R}^d$, $A \in \mathcal{B}(\mathbb{R}^d)$ by

$$R_\gamma(y, A) = (2\pi\sigma^2\gamma)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_A(y') \exp \left\{ -\|y' - T_\gamma(y)\|^2 / (2\sigma^2\gamma) \right\} dy'.$$

Note that \tilde{R}_γ associated with $(\tilde{Y}_k)_{k \in \mathbb{N}}$ is given by the same expression upon replacing T_γ by \tilde{T}_γ . We consider the following assumption on the family $\{T_\gamma : \gamma \in (0, \bar{\gamma}]\}$. This condition will ensure that R_γ is geometrically ergodic (see Proposition 4.1) and it will be important to derive our main results regarding the distance of R_γ^k and \tilde{R}_γ^k , for $k \in \mathbb{N}$.

H1. *There exist $R_1, L \geq 0$ and $m > 0$ such that for any $\gamma \in (0, \bar{\gamma}]$, there exists a non-decreasing function $\tau_\gamma : [0, +\infty) \rightarrow [0, +\infty)$ satisfying $\tau_\gamma(0) = 0$, $\|T_\gamma(x) - T_\gamma(\tilde{x})\| \leq \tau_\gamma(\|x - \tilde{x}\|)$ for any $x, \tilde{x} \in \mathbb{R}^d$, and*

$$\sup_{r \in (0, +\infty)} \{\tau_\gamma(r)/r\} \leq 1 + \gamma L, \quad \sup_{r \in (R_1, +\infty)} \{\tau_\gamma(r)/r\} \leq 1 - \gamma m. \quad (4.3)$$

In addition, $\sup_{\gamma \in (0, \bar{\gamma}]} \|T_\gamma(0)\| < +\infty$.

Note that the condition that for any $\gamma \in (0, \bar{\gamma}]$, τ_γ is non-decreasing can be omitted upon replacing in our study τ_γ by the affine majorant

$$\bar{\tau}_\gamma : r \mapsto \begin{cases} (1 + L\gamma)r & \text{if } r \in [0, R_1], \\ (1 + L\gamma)R_1 + (1 - m\gamma)(r - R_1) & \text{otherwise.} \end{cases} \quad (4.4)$$

Indeed, by definition and (4.3), for any $r \in [0, +\infty)$, $\tau_\gamma(r) \leq \bar{\tau}_\gamma(r)$, therefore for any $x, \tilde{x} \in \mathbb{R}^d$, $\|T_\gamma(x) - T_\gamma(\tilde{x})\| \leq \bar{\tau}_\gamma(\|x - \tilde{x}\|)$. In addition, an easy computation leads to

4.2. Sticky reflection coupling

setting $R_2 = 2R_1(L + m)/m$,

$$\sup_{r \in (0, +\infty)} \{\bar{\tau}_\gamma(r)/r\} \leq 1 + \gamma L, \quad \sup_{r \in (R_2, +\infty)} \{\bar{\tau}_\gamma(r)/r\} \leq 1 - \gamma m/2.$$

Then, $\bar{\tau}_\gamma$ satisfies **H1** and is non-decreasing.

Note that **H1** implies that for any $r \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$, $\tau_\gamma(r) \leq (1 + \gamma L)r$, therefore T_γ is $(1 + \gamma L)$ -Lipschitz. The second condition in (4.3) ensures that for any $\gamma \in (0, \bar{\gamma}]$, T_γ is a *contraction at large distances*, i.e. for any $x, \tilde{x} \in \mathbb{R}^d$, $\|T_\gamma(x) - T_\gamma(\tilde{x})\| \leq (1 - \gamma m) \|x - \tilde{x}\|$, if $\|x - \tilde{x}\| \geq R_1$.

The assumption **H1** holds for the Euler scheme applied to diffusions with scalar covariance matrices, i.e. (4.1) with $T_\gamma(x) = x + \gamma b(x)$ and a drift function $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$, if, for some $L_b, m_b, R_b > 0$, b is L_b -Lipschitz continuous and satisfies

$$\langle x - y, b(x) - b(y) \rangle \leq -m_b \|x - y\|^2,$$

for all $x, y \in \mathbb{R}^d$ with $\|x - y\| \geq R_b$. Indeed, this implies that for any $x, y \in \mathbb{R}^d$, $\|T_\gamma(x) - T_\gamma(y)\| \leq (1 + L_b \gamma) \|x - y\|$ and, provided $\gamma \in (0, m_b/L^2)$ and $\|x - y\| \geq R_b$, $\|T_\gamma(x) - T_\gamma(y)\|^2 \leq (1 - m_b \gamma) \|x - y\|^2$. Therefore, it suffices to consider τ_γ defined by (4.4) with $L = L_b$, $m = m_b/2$ and $R_1 = R_b$.

Our results will be stated in term of Wasserstein distances and V -norms, whose definitions are the following. Consider a measurable cost function $\mathbf{c} : \mathbb{R}^{2d} \rightarrow [0, \infty)$. Then the associated Wasserstein distance \mathbf{W}_c is given for two probability measures μ, ν on \mathbb{R}^d by

$$\mathbf{W}_c(\nu, \mu) = \inf_{\pi \in \Pi(\nu, \mu)} \int_{\mathbb{R}^{2d}} \mathbf{c}(x, y) \pi(dx, dy),$$

where $\Pi(\nu, \mu)$ is the set of transference plans or couplings between ν and μ , namely the set of probability measures on \mathbb{R}^d whose first and second d -dimensional marginals are ν and μ respectively. In the particular case where $\mathbf{c}(x, y) = \mathbb{1}_{\Delta_{\mathbb{R}^d}^c}(x, y)$, \mathbf{W}_c is simply the total variation distance $\|\cdot\|_{TV}$. For $V : \mathbb{R}^d \rightarrow [1, +\infty)$, the choice $\mathbf{c}(x, y) = \mathbb{1}_{\Delta_{\mathbb{R}^d}^c}(x, y) \{V(x) + V(y)\}$ yields the V -norm (see [37, Theorem 19.1.7]), i.e. $\mathbf{W}_c(\nu, \mu) = \|\nu - \mu\|_V$. Finally, for $\mathbf{c}(x, y) = \|x - y\|^p$ with $p \in [1, +\infty)$, \mathbf{W}_c is the p -th power of the usual Wasserstein distance of order p .

Following the same line as the proof of [33, Theorem 15]¹, we can show that **H1** implies that the Markov kernel R_γ is V_c -uniformly geometrically ergodic where for any $c > 0$ and $x \in \mathbb{R}^d$, $V_c(x) = \exp(c\|x\|^2)$, with a convergence rate that scales linearly with the step size γ .

Proposition 4.1. *Assume **H1**. Then, setting $\bar{\gamma}_1 = \bar{\gamma} \wedge \{1/m\}$, for any $\gamma \in (0, \bar{\gamma}_1]$, R_γ admits a unique stationary distribution π_γ . In addition, there exist $c > 0$, $\rho \in [0, 1)$ and $C \geq 0$ such that for any $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}_1]$, $\|\delta_x R_\gamma^k - \pi_\gamma\|_{V_c} \leq C \rho^{k\gamma} V_c(x)$.*

¹[33, Theorem 15] consider the case where T_γ comes from the Euler discretization scheme and has form $T_\gamma(x) = x + \gamma b(x)$.

Proof. This result is a simple consequence of [33, Corollary 11]. For completeness, its proof is included in Section 4.5.1.1. \square

Note that this result can be made quantitative, and other convergence results in total variation and Wasserstein distance of order $p \in [1, +\infty)$ can also be established following the same lines as the proof of [33, Corollary 14]. However, these results are out of the scope of the present paper and would be simple adaptations of those in [33] or [50].

We now consider an assumption which quantifies the perturbation associated with \tilde{T}_γ relatively to T_γ , for $\gamma \in (0, \bar{\gamma}]$.

H2. *There exists $c_\infty > 0$ such that $\sup_{x \in \mathbb{R}^d} \|T_\gamma(x) - \tilde{T}_\gamma(x)\| \leq \gamma c_\infty$ for all $\gamma \in (0, \bar{\gamma}]$.*

Example 4.2. *The assumption H2 holds for the Euler scheme applied to diffusions with scalar covariance matrices, i.e. (4.1) and (4.2) with*

$$T_\gamma(x) = x + \gamma b(x) \quad \text{and} \quad \tilde{T}_\gamma(x) = x + \gamma \tilde{b}(x), \quad (4.5)$$

under the condition that $\sup_{x \in \mathbb{R}^d} \|b(x) - \tilde{b}(x)\| \leq c_\infty$. This setting is exactly the one we introduced to motivate our study. In particular, in the case where $b = -\nabla U$ for some potential U , \tilde{b} may correspond to a numerical approximation of this gradient.

Note that compared to T_γ , $\gamma \in (0, \bar{\gamma}]$, we do not assume any smoothness condition on \tilde{T}_γ . More precisely, we do not assume that \tilde{T}_γ satisfies H1. Regarding the ergodicity properties of \tilde{R}_γ associated with \tilde{T}_γ , $\gamma \in (0, \bar{\gamma}]$, we have the following result.

Proposition 4.3. *Assume H1 and H2 and set $\bar{\gamma}_1 = \bar{\gamma} \wedge \{1/m\}$. Then, for any $\gamma \in (0, \bar{\gamma}_1]$, \tilde{R}_γ admits a unique stationary distribution $\tilde{\pi}_\gamma$. In addition, there exists $c > 0$ such that for any $\gamma \in (0, \bar{\gamma}_1]$, there exist $\rho_\gamma \in [0, 1)$ and $C_\gamma \geq 0$ such that for any $x \in \mathbb{R}^d$ $\|\delta_x \tilde{R}_\gamma^k - \tilde{\pi}_\gamma\|_{V_c} \leq C_\gamma \rho_\gamma^k V_c(x)$, where $V_c(x) = \exp(c\|x\|^2)$.*

Proof. The proof is postponed to Section 4.5.1.2. \square

Similarly to Proposition 4.1 with respect to R_γ , Proposition 4.3 implies that \tilde{R}_γ is V_c -uniformly geometrically ergodic. However in contrast to Proposition 4.1, the dependency of the rate of convergence with respect to the step size γ is not explicit anymore since the results and the method employed in [33] or [50] cannot be applied anymore.

Note that Proposition 4.1 and Proposition 4.3 imply that R_γ and \tilde{R}_γ converge to π_γ and $\tilde{\pi}_\gamma$ respectively in total variation and Wasserstein metric of any order $p \in [1, +\infty)$.

Based on the two assumptions above, we can now state one of our main results. Our goal is to quantify the distance between the laws of the iterates of the two chains $(Y_k)_{k \in \mathbb{N}}$ and (\tilde{Y}_k) , in particular starting from the same initial point $x \in \mathbb{R}^d$ or at equilibrium. Indeed, remark that, in view of Propositions 4.1 and 4.3, letting $k \rightarrow +\infty$ in the next statement yields quantitative bounds on $\mathbf{W}_c(\pi_\gamma, \tilde{\pi}_\gamma)$ for any $\gamma \in (0, \bar{\gamma} \wedge \{1/m\}]$.

Theorem 4.4. *Assume H1 and H2 hold and let*

$$(\tilde{\mathbf{c}}, \mathcal{V}) \in \{(\mathbb{1}_{(0,+\infty)}, |\cdot|), (|\cdot|, |\cdot|), (\mathbb{1}_{(0,+\infty)} \exp(|\cdot|), \mathbb{1}_{(0,+\infty)} \exp(|\cdot|))\}.$$

4.2. Sticky reflection coupling

Then, there exist some explicit constants $C, c \geq 0, \rho \in [0, 1)$ such that for any $k \in \mathbb{N}$, $\gamma \in (0, \bar{\gamma}]$ and $x, \tilde{x} \in \mathbb{R}^d$,

$$\mathbf{W}_{\mathbf{c}}(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k) \leq C \rho^{\gamma k} \mathcal{V}(\|x - \tilde{x}\|) + cc_\infty.$$

where $\mathbf{c}(x, \tilde{x}) = \tilde{\mathbf{c}}(\|x - \tilde{x}\|)$.

Remark 4.5. It is also possible to treat the case of cost functions \mathbf{c} of the form $\mathbf{c}(x, y) = \tilde{\mathbf{c}}(\|x - y\|) (V(x) + V(y))$ with $\tilde{\mathbf{c}}$ as in Theorem 4.4 and V a positive function, simply by using Hölder's inequality. Indeed, for $p, q > 1$ with $1/p + 1/q = 1$, we can bound

$$\mathbf{W}_{\mathbf{c}}(\nu, \mu) \leq (\mathbf{W}_{\mathbf{c}_p}(\nu, \mu))^{1/p} \left((\nu(V^q))^{1/q} + (\mu(V^q))^{1/q} \right)$$

with $\mathbf{c}_p(x, y) = \tilde{\mathbf{c}}^p(\|x - y\|)$. Bounds on the $\mathbf{W}_{\mathbf{c}_p}$ distance can then be established as in Theorem 4.4, while bounds on expected values of V^q , independent of γ , are classically obtained through Lyapunov arguments (see e.g. the proof of Proposition 4.1 in Section 4.5.1.1).

The rest of this section is devoted to the proof of Theorem 4.4. In particular, we define in the following the main object of this paper.

4.2.2 The discrete sticky kernel

We define a Markovian coupling of the two chains $(Y_k)_{k \in \mathbb{N}}$ and $(\tilde{Y}_k)_{k \in \mathbb{N}}$ defined in (4.1) and (4.2) by using, at each step, the maximal reflection coupling of the two Gaussian proposals, which is optimal for the total variation distance (i.e that maximizes the probability of coalescence). Let $(U_k)_{k \geq 1}$ be a sequence of *i.i.d.* uniform random variables on $[0, 1]$ independent of $(Z_k)_{k \geq 1}$ which we recall is a sequence of *i.i.d.* d -dimensional standard Gaussian random variables. We define the discrete sticky Markov coupling K_γ of R_γ and \tilde{R}_γ as the Markov kernel associated with the Markov chain on \mathbb{R}^{2d} given for $k \in \mathbb{N}$ by

$$\begin{aligned} X_{k+1} &= \mathbf{T}_\gamma(X_k) + (\sigma^2 \gamma)^{1/2} Z_{k+1} \\ \tilde{X}_{k+1} &= X_{k+1} B_{k+1} + (1 - B_{k+1}) \mathbf{F}_\gamma(X_k, \tilde{X}_k, Z_{k+1}), \end{aligned} \quad (4.6)$$

where $B_{k+1} = \mathbb{1}_{[0, +\infty)}(p_\gamma(X_k, \tilde{X}_k, Z_{k+1}) - U_{k+1})$ and

$$\begin{aligned} \mathbf{F}_\gamma(x, \tilde{x}, z) &= \tilde{\mathbf{T}}_\gamma(\tilde{x}) + (\sigma^2 \gamma)^{1/2} \left\{ \text{Id} - 2\mathbf{e}(x, \tilde{x}) \mathbf{e}(x, \tilde{x})^\top \right\} z, \\ \mathbf{E}(x, \tilde{x}) &= \tilde{\mathbf{T}}_\gamma(\tilde{x}) - \mathbf{T}_\gamma(x), \quad \mathbf{e}(x, \tilde{x}) = \begin{cases} \frac{\mathbf{E}(x, \tilde{x})}{\|\mathbf{E}(x, \tilde{x})\|} & \text{if } \mathbf{E}(x, \tilde{x}) \neq 0 \\ \mathbf{e}_0 & \text{otherwise,} \end{cases} \\ p_\gamma(x, \tilde{x}, z) &= 1 \wedge \left[\frac{\varphi_{\sigma^2 \gamma} \left\{ \|\mathbf{E}(x, \tilde{x})\| - (\sigma^2 \gamma)^{1/2} \langle \mathbf{e}(x, \tilde{x}), z \rangle \right\}}{\varphi_{\sigma^2 \gamma} \left\{ (\sigma^2 \gamma)^{1/2} \langle \mathbf{e}(x, \tilde{x}), z \rangle \right\}} \right], \end{aligned} \quad (4.7)$$

where $\mathbf{e}_0 \in \mathbb{R}^d$ is an arbitrary unit-vector, *i.e.* $\|\mathbf{e}_0\| = 1$, and $\varphi_{\sigma^2\gamma}$ is the density of the one-dimensional Gaussian distribution with mean 0 and variance $\sigma^2\gamma$. In other words, K_γ is given for any $\gamma \in (0, \bar{\gamma}]$, $(x, y) \in \mathbb{R}^{2d}$ and $\mathbf{A} \in \mathcal{B}(\mathbb{R}^{2d})$ by

$$\begin{aligned} K_\gamma((x, \tilde{x}), \mathbf{A}) &= \int_{\mathbb{R}^d} \mathbb{1}_{\mathbf{A}}(\mathbf{T}_\gamma(x) + (\sigma^2\gamma)^{1/2}z, \mathbf{T}_\gamma(x) + (\sigma^2\gamma)^{1/2}z) p_\gamma(x, \tilde{x}, z) \frac{e^{-\|z\|^2/2}}{(2\pi)^{d/2}} dz \\ &\quad + \int_{\mathbb{R}^d} \mathbb{1}_{\mathbf{A}}(\mathbf{T}_\gamma(x) + (\sigma^2\gamma)^{1/2}z, \mathbf{F}_\gamma(x, \tilde{x}, z)) (1 - p_\gamma(x, \tilde{x}, z)) \frac{e^{-\|z\|^2/2}}{(2\pi)^{d/2}} dz. \end{aligned}$$

In words, from the initial conditions (x, \tilde{x}) , this coupling works as follows: first, a Gaussian variable Z_{k+1} is drawn for the fluctuations of X_{k+1} . Then, \tilde{X}_{k+1} is made equal to X_{k+1} with probability $p(x, \tilde{x}, Z_{k+1})$ and, otherwise, the random variable \tilde{Z}_{k+1} which determines the fluctuations of \tilde{X}_{k+1} with respect to its average $\tilde{\mathbf{T}}_\gamma(\tilde{x})$ is given by the orthogonal reflection of Z_{k+1} in the direction $\tilde{\mathbf{T}}_\gamma(\tilde{x}) - \mathbf{T}_\gamma(x)$. It is well known that for any $(x, \tilde{x}) \in \mathbb{R}^d$, $K_\gamma((x, \tilde{x}), \mathbf{A} \times \mathbb{R}^d) = R_\gamma(x, \mathbf{A})$ and $K_\gamma((x, \tilde{x}), \mathbb{R}^d \times \mathbf{A}) = \tilde{R}_\gamma(x, \mathbf{A})$, see e.g. [21, Section 3.3], [45, Section 4.1], [50] or [33].

The starting point of our analysis is the next result, which will enable to compare the coupling difference process $\|X_{k+1} - \tilde{X}_{k+1}\|$ with a Markov chain on $[0, +\infty)$. Define $(G_k)_{k \geq 1}$ for any $k \geq 1$ by

$$G_k = \langle \mathbf{e}(X_{k-1}, \tilde{X}_{k-1}), Z_k \rangle, \quad (4.8)$$

where \mathbf{e} is given by (4.7). For any $a \geq 0$, $g \in \mathbb{R}$, $u \in [0, 1]$ and $\gamma \in (0, \bar{\gamma}]$ define

$$\mathcal{H}_\gamma(a, g, u) = \mathbb{1}_{[0, +\infty)}(u - \bar{p}_{\sigma^2\gamma}(a, g)) \left(a - 2(\sigma^2\gamma)^{1/2}g \right), \quad (4.9)$$

where

$$\bar{p}_{\sigma^2\gamma}(a, g) = 1 \wedge \frac{\varphi_{\sigma^2\gamma}(a - (\sigma^2\gamma)^{1/2}g)}{\varphi_{\sigma^2\gamma}((\sigma^2\gamma)^{1/2}g)}. \quad (4.10)$$

Proposition 4.6. *Assume **H1** and **H2** hold. Then for any $\gamma \in (0, \bar{\gamma}]$, $k \in \mathbb{N}$, almost surely, we have*

$$\|X_{k+1} - \tilde{X}_{k+1}\| \leq \mathcal{G}_\gamma(\|X_k - \tilde{X}_k\|, G_{k+1}, U_{k+1}), \quad (4.11)$$

where $(X_k, \tilde{X}_k)_{k \in \mathbb{N}}$ are defined by (4.6), and for any $w \in [0, +\infty)$, $g \in \mathbb{R}$ and $u \in [0, 1]$,

$$\mathcal{G}_\gamma(w, g, u) = \mathcal{H}_\gamma(\tau_\gamma(w) + \gamma c_\infty, g, u).$$

In addition, for any $g \in \mathbb{R}^d$ and $u \in [0, 1]$, $w \mapsto \mathcal{G}_\gamma(w, g, u)$ is non-decreasing.

Proof. The proof is postponed to Section 4.5.1.3. □

Consider now the stochastic process $(W_k)_{k \in \mathbb{N}}$ starting from $\|X_0 - \tilde{X}_0\|$ and defined by induction on k as follows,

$$\begin{aligned} W_{k+1} &= \mathcal{G}_\gamma(W_k, G_{k+1}, U_{k+1}) \\ &= \begin{cases} \tau_\gamma(W_k) + \gamma c_\infty - 2\sigma\sqrt{\gamma}G_k & \text{if } U_{k+1} \geq \bar{p}_{\sigma^2\gamma}(\tau_\gamma(W_k) + \gamma c_\infty, G_k) \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (4.12)$$

4.2. Sticky reflection coupling

By definition (4.8) and (4.7), an easy induction implies that $(G_k)_{k \geq 1}$ and $(U_k)_{k \geq 1}$ are independent, $(G_k)_{k \geq 1}$ are i.i.d. standard Gaussian random variables and $(U_k)_{k \geq 1}$ are i.i.d. uniform random variables on $[0, 1]$. Therefore, $(W_k)_{k \in \mathbb{N}}$ is a Markov chain with Markov kernel Q_γ defined for $w \in [0, +\infty)$ and $\mathbf{A} \in \mathcal{B}([0, +\infty))$ by

$$\begin{aligned} Q_\gamma(w, \mathbf{A}) &= \delta_0(\mathbf{A}) \int_{\mathbb{R}} \bar{p}_{\sigma^2 \gamma}(\tau_\gamma(w) + \gamma c_\infty, g) \varphi(g) dg \\ &+ \int_{\mathbb{R}} \mathbb{1}_{\mathbf{A}} \left(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\gamma^{1/2}g \right) \{1 - \bar{p}_{\sigma^2 \gamma}(\tau_\gamma(w) + \gamma c_\infty, g)\} \varphi(g) dg, \end{aligned} \quad (4.13)$$

where φ is the density of the standard Gaussian distribution on \mathbb{R} . By Proposition 4.6, we have almost surely for any $k \in \mathbb{N}$,

$$\|X_k - \tilde{X}_k\| \leq W_k. \quad (4.14)$$

Another consequence of Proposition 4.6 is that Q_γ is stochastically monotonous (see e.g. [89] or [122]), more precisely if $(W_k)_{k \in \mathbb{N}}$ and $(\tilde{W}_k)_{k \in \mathbb{N}}$ are two chains given by (4.12) with the same variables $(G_k, U_k)_{k \in \mathbb{N}}$ with $W_0 \leq \tilde{W}_0$, then almost surely $W_k \leq \tilde{W}_k$ for all $k \in \mathbb{N}$. This nice property will be used several times in the analysis of this chain.

The main consequence of (4.14) is the following result.

Corollary 2. *Assume H1 and H2 hold. Let $\mathbf{c} : \mathbb{R}^{2d} \rightarrow [0, +\infty)$ of the form $\mathbf{c}(x, y) = \tilde{\mathbf{c}}(\|x - y\|)$ for some non-decreasing function $\tilde{\mathbf{c}} : [0, +\infty) \rightarrow [0, +\infty)$, $\tilde{\mathbf{c}}(0) = 0$. For any $x, \tilde{x} \in \mathbb{R}^d$ and $k \in \mathbb{N}$,*

$$\mathbf{W}_{\mathbf{c}}(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k) \leq \int_{\mathbb{R}^{2d}} \mathbf{c}(y, \tilde{y}) K_\gamma^k((x, \tilde{x}), d(y, \tilde{y})) \leq \int_0^{+\infty} \tilde{\mathbf{c}}(\tilde{w}) Q_\gamma^k(\|x - \tilde{x}\|, d\tilde{w}).$$

Proof. Let $k \in \mathbb{N}$. By (4.14) and since $\tilde{\mathbf{c}}$ is non-decreasing, we get almost surely $\tilde{\mathbf{c}}(\|X_k - \tilde{X}_k\|) \leq \tilde{\mathbf{c}}(W_k)$. Taking the expectation concludes the proof. \square

From Corollary 2, the question to get bounds on $\mathbf{W}_{\mathbf{c}}(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k)$ boils down to the study of the Markov kernel Q_γ on $[0, +\infty)$, which is the main part of our work.

4.2.3 Analysis of the auxiliary Markov chain

We start with a Lyapunov/drift result.

Proposition 4.7. *Assume H1 and H2 hold. Then for any $w \geq 0$,*

$$Q_\gamma \mathcal{V}_1^*(w) \leq (1 - \gamma \mathbf{m}) \mathcal{V}_1^*(w) \mathbb{1}_{(R_1, +\infty)}(w) + (1 + \gamma \mathbf{L}) \mathcal{V}_1^*(w) \mathbb{1}_{(0, R_1]}(w) + \gamma c_\infty,$$

where Q_γ is defined by (4.13) and for any $w \in \mathbb{R}$, $\mathcal{V}_1^*(w) = |w|$.

Proof. The proof is postponed to Section 4.5.1.4. \square

Proposition 4.7 implies in particular that for any $w \in \mathbb{R}$,

$$Q_\gamma \mathcal{V}_1^*(w) \leq (1 - \gamma \mathfrak{m}) \mathcal{V}_1^*(w) + \gamma[(\mathbf{L} + \mathfrak{m})R_1 + c_\infty].$$

Then, a straightforward induction shows that for any $k \in \mathbb{N}$,

$$Q_\gamma^k \mathcal{V}_1^*(w) \leq (1 - \gamma \mathfrak{m})^k \mathcal{V}_1^*(w) + [(\mathbf{L} + \mathfrak{m})R_1 + c_\infty]/\mathfrak{m},$$

and therefore by Corollary 2 taking $\tilde{\mathbf{c}}(t) = t$,

$$\mathbf{W}_1(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k) \leq (1 - \gamma \mathfrak{m})^k \|x - \tilde{x}\| + [(\mathbf{L} + \mathfrak{m})R_1 + c_\infty]/\mathfrak{m}. \quad (4.15)$$

However, this result is not sharp as $k \rightarrow +\infty$. Indeed, in the case $c_\infty = 0$, $R_\gamma = \tilde{R}_\gamma$ and by Proposition 4.1, it holds that $\mathbf{W}_1(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k) \rightarrow 0$ as $k \rightarrow +\infty$, while the right-hand side of (4.15) converges to $(\mathbf{L} + \mathfrak{m})R_1/\mathfrak{m} \neq 0$. In particular, that is why adapting existing results, such as the one established in [123], is not an option here. We need to refine our results in order to fill this gap. To this end, we need to analyze more precisely the long-time behavior of Q_γ . A first step is to show that it is ergodic.

Proposition 4.8. *Assume H1 and H2 hold. For any $\gamma \in (0, \bar{\gamma}]$, Q_γ admits a unique invariant probability measure μ_γ and is geometrically ergodic. In addition, $\mu_\gamma(\{0\}) > 0$ and μ_γ is absolutely continuous with respect to the measure $\delta_0 + \text{Leb}$ on $([0, +\infty), \mathcal{B}([0, +\infty)))$. Finally, in the case $c_\infty \neq 0$, μ_γ and $\delta_0 + \text{Leb}$ are equivalent.*

Proof. The proof is postponed to Section 4.5.1.5. □

Corollary 3. *Assume H1 and H2 hold. Let $\mathbf{c} : \mathbb{R}^{2d} \rightarrow [0, +\infty)$ of the form $\mathbf{c}(x, y) = \tilde{\mathbf{c}}(\|x - y\|)$ for some non-decreasing function $\tilde{\mathbf{c}} : [0, +\infty) \rightarrow [0, +\infty)$, $\tilde{\mathbf{c}}(0) = 0$. For any $x, \tilde{x} \in \mathbb{R}^d$ and $k \in \mathbb{N}$,*

$$\mathbf{W}_{\mathbf{c}}(\delta_x R_\gamma^k, \delta_{\tilde{x}} \tilde{R}_\gamma^k) \leq \int_0^{+\infty} \tilde{\mathbf{c}}(\tilde{w}) \{Q_\gamma^k(\|x - \tilde{x}\|, \cdot) - \mu_\gamma\}(\mathrm{d}\tilde{w}) + \mu_\gamma(\tilde{\mathbf{c}}). \quad (4.16)$$

where μ_γ is the stationary distribution of Q_γ given by (4.13). In particular, if $x = \tilde{x}$, $\mathbf{W}_{\mathbf{c}}(\delta_x R_\gamma^k, \delta_x \tilde{R}_\gamma^k) \leq \mu_\gamma(\tilde{\mathbf{c}})$.

Proof. The proof of (4.16) is a consequence of Proposition 4.8 and Corollary 2. The last statement follows from the fact that Q_γ is stochastically monotonous. Indeed, by Proposition 4.6, for any $w, \tilde{w} \in [0, +\infty)$, $w \leq \tilde{w}$, and $a \in [0, +\infty)$, $Q_\gamma(w, [0, a]) \geq Q_\gamma(\tilde{w}, [0, a])$. Therefore, for any $a \in [0, +\infty)$, $w \mapsto Q_\gamma(w, [0, a])$ is non-increasing on $[0, +\infty)$ and for any non-increasing bounded function f , $Q_\gamma f(w) \geq Q_\gamma f(\tilde{w})$ for any $w, \tilde{w} \in [0, +\infty)$, $w \leq \tilde{w}$. As a result, a straightforward induction shows that for any $k \in \mathbb{N}$, $w, \tilde{w} \in [0, +\infty)$, $w \leq \tilde{w}$, and $a \in [0, +\infty)$, $Q_\gamma^k(w, [0, a]) \geq Q_\gamma^k(\tilde{w}, [0, a])$. Then, we obtain $Q_\gamma^k(0, [0, a]) \geq \int_0^{+\infty} \mu_\gamma(\mathrm{d}w) Q_\gamma^k(w, [0, a]) = \mu_\gamma([0, a])$. Since $\tilde{\mathbf{c}}$ is non-decreasing on $[0, +\infty)$, we get $Q_\gamma \tilde{\mathbf{c}}(0) \leq \mu_\gamma(\tilde{\mathbf{c}})$, which combined with (4.16) completes the proof. □

Corollary 3 then naturally brings us to derive moment bounds for the stationary distribution μ_γ , $\gamma \in (0, \bar{\gamma}]$ and quantitative convergence bounds for Q_γ to μ_γ . Our next results address these two problems.

4.2. Sticky reflection coupling

Theorem 4.9. *Assume **H1** and **H2** hold. For any $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$ and $\gamma \in (0, \bar{\gamma}]$,*

$$\int_{[0, +\infty)} w \mu_\gamma(dw) \leq c_\infty c_1, \quad \mu_\gamma((0, +\infty)) \leq c_\infty c_2, \quad (4.17)$$

where μ_γ is the stationary distribution of Q_γ given by (4.13), and, considering ζ given below in (4.55),

$$\begin{aligned} c_1 &= \eta_1 R_1 (1 + \mathbf{L}/\mathbf{m}) + 1/\mathbf{m}, \\ c_2 &= e^{(\bar{\delta} + \bar{\gamma})\mathbf{L}} (c_1 (1 + \bar{\gamma}\mathbf{L})/\bar{\delta}^{1/2} + [\bar{\delta} + \bar{\gamma}]^{1/2}) / (\sqrt{2\pi\sigma}) + 2\zeta [\bar{\delta} + \bar{\gamma}]^{1/2} e^{3(\bar{\delta} + \bar{\gamma})\mathbf{L}} / \sigma^3, \\ \eta_1 &= [\bar{\delta} + \bar{\gamma}]^{1/2} \left[\frac{\zeta e^{3(\bar{\delta} + \bar{\gamma})\mathbf{L}}}{\sigma^3} + \frac{e^{(\bar{\delta} + \bar{\gamma})\mathbf{L}}}{2\sqrt{2\pi\sigma}} \right] / \Phi \left(-\frac{(1 + \bar{\gamma}\mathbf{L})R_1 + (\bar{\delta} + \bar{\gamma})c_\infty}{2\bar{\delta}^{1/2}\sigma e^{-(\bar{\delta} + \bar{\gamma})\mathbf{L}}} \right). \end{aligned}$$

Proof. The proof is postponed to Section 4.5.1.6. □

Theorem 4.10. *Assume **H1** and **H2** hold. For any $a > 0$ and $\gamma \in (0, \bar{\gamma}]$,*

$$\int_0^{+\infty} \mathbb{1}_{(0, +\infty)}(w) \exp(aw) d\mu_\gamma(w) \leq c_\infty c_3,$$

where c_3 is explicitly given in the proof and μ_γ is the stationary distribution of Q_γ given by (4.13).

Proof. The proof is postponed to Section 4.5.1.7. □

We now specify the convergence of Q_γ to μ_γ for any $\gamma \in (0, \bar{\gamma}]$.

Theorem 4.11. *Assume **H1** and **H2** hold. There exist explicit constants $\rho \in [0, 1)$ and $C \geq 0$ such that for any $\gamma \in (0, \bar{\gamma}]$, $w \geq 0$,*

$$\|\delta_w Q_\gamma^k - \mu_\gamma\|_{\mathcal{V}} \leq C \rho^{\gamma k} \mathcal{V}(w),$$

where $\mathcal{V}(w) = 1 + |w|$ or $\mathcal{V}(w) = \exp(a|w|)$, for $a > 0$.

Proof. The proof is postponed to Section 4.5.1.8. □

Combining the results of Corollary 3, Theorem 4.9, Theorem 4.10 and Theorem 4.11 allows to address the main questions raised in this section and prove Theorem 4.4.

Discussion on the bounds provided by Theorem 4.9 In this paragraph, we discuss how the constants c_1, c_2 given in Theorem 4.9 behaves with respect to the parameters $R_1, \mathbf{L}, \mathbf{m}$ in the limit $c_\infty \rightarrow 0$ and $\bar{\gamma} \rightarrow 0$. For ease of presentation, we also only consider the case $\sigma = 1$.

- (1) First consider the case $R_1 = 0$. As $\mathbf{m} \rightarrow 0$, c_1, c_2 are of order \mathbf{m}^{-1} and $1/[\mathbf{m}\bar{\delta}^{1/2}] + \bar{\delta}^{1/2}$ respectively for $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$. Since \mathbf{L} can be taken arbitrarily small (as $R_1 = 0$), choosing $\bar{\delta} = \mathbf{m}^{-1}$, we obtain that c_2 is of order $\mathbf{m}^{-1/2}$. Note that the dependency of c_1, c_2 with respect to \mathbf{m} is sharp; see Example 4.12 below.

- (2) We now consider the case $R_1 \geq 1$, $L = 0$. Note that in this case $\bar{\delta}$ can be chosen arbitrarily in $(0, 1)$. Then, for some universal constants C_1, C_2, C_3 , $\eta_1 \geq C_1 \bar{\delta}^{1/2} / \Phi\{C_2 R_1 / \bar{\delta}^{1/2} + C_3 c_\infty \bar{\delta}^{1/2}\}$. Therefore, taking $\bar{\delta} = m^{-1} \vee R_1^2$, we get that for some universal constants $D_1, D_2, E \geq 0$, $c_1 \leq D_1[(R_1 \vee m^{-1/2}) + m^{-1}]$, $c_2 \leq E m^{-1/2} \vee R_1$. Note that the bound of c_2 with respect to R_1 and m is consistent with the results obtained in [51] (see [51, Lemma 1]) for the stationary distributions of continuous sticky processes. Note that it is shown in [51, Example 2] that this bound is sharp with respect to R_1 and m .
- (3) In the case $R_1 \wedge L \geq 1$, taking $\bar{\delta} = L^{-1}$ since we are in the regime $c_\infty \rightarrow 0$, we get that up to logarithmic term and using $\bar{\gamma} \leq L^{-1}$, c_1, c_2 are smaller than $C \exp[e^4(R_1 L^{1/2} + c_\infty)^2]$ for some universal constant $C \geq 0$. The estimate for c_2 is also consistent with [51, Lemma 1] which holds for stationary distributions of continuous sticky processes.

Example 4.12. Consider the particular example of two auto-regressive processes for which $T_\gamma(y) = (1 - \varrho\gamma)y$ and $\tilde{T}_\gamma(y) = (1 - \varrho\gamma)y + \gamma\varrho a$ for $\gamma \in (0, \varrho^{-1})$ and $a, \varrho > 0$. Then, on the one hand, **H1** and **H2** are satisfied with $R_1 = 0$, $m = \varrho$ and $c_\infty = \varrho a$ which lead to $c_1 c_\infty \sim a$ and $c_2 c_\infty \sim Ca/\varrho^{1/2}$, as $\varrho \rightarrow 0$, for some universal constant $C \geq 0$. On the other hand, an easy computation (see e.g. [42]) shows that the stationary distributions π_γ and $\tilde{\pi}_\gamma$ provided by Proposition 4.1 and Proposition 4.3 are $N(0, \varrho^{-1}(2 - \gamma\varrho\gamma)^{-1})$ and $N(a, \varrho^{-1}(2 - \gamma\varrho\gamma)^{-1})$ respectively. Therefore, we get $\mathbf{W}_1(\pi_\gamma, \tilde{\pi}_\gamma) = a$ and $\|\pi_\gamma - \tilde{\pi}_\gamma\|_{TV} \sim Ca/\varrho^{1/2}$ as $\varrho \rightarrow 0$.

4.3 Continuous-time limit

In the case where T_γ and \tilde{T}_γ are specified by (4.5), then under appropriate conditions on b and \tilde{b} , it can be shown, see e.g. [33, Proposition 25], that for any $T \geq 0$ and $x \in \mathbb{R}^d$,

$$\lim_{m \rightarrow +\infty} \{\|\delta_x R_{T/m}^m - \delta_x P_T\|_V + \|\delta_x \tilde{R}_{T/m}^m - \delta_x \tilde{P}_T\|_V\} = 0, \quad (4.18)$$

for some measurable function $V : \mathbb{R}^d \rightarrow [1, +\infty)$ and where $(P_t)_{t \geq 0}$ and $(\tilde{P}_t)_{t \geq 0}$ are the Markov semigroup corresponding to (4.1) and (4.2). Then, this naturally implies convergence in total variation and also Wasserstein distance of order p if $\inf_{x \in \mathbb{R}^d} \{V(x)/\|x\|^p\} > 0$. As a consequence, results of Section 4.2 immediately transfer to the continuous-time processes. More precisely, let $\mathbf{c} : \mathbb{R}^{2d} \rightarrow [0, +\infty)$ of the form $\mathbf{c}(x, y) = \tilde{\mathbf{c}}(\|x - y\|)$ for some non-decreasing function $\tilde{\mathbf{c}} : [0, +\infty) \rightarrow [0, +\infty)$, $\tilde{\mathbf{c}}(0) = 0$. If (4.18) holds and $\sup_{x, y \in \mathbb{R}^d} \{\mathbf{c}(x, y)/\{V(x) + V(y)\}\} < +\infty$, we get by the triangle inequality that for any $x, \tilde{x} \in \mathbb{R}^d$, $T > 0$, $\mathbf{W}_\mathbf{c}(\delta_x P_T, \delta_{\tilde{x}} \tilde{P}_T) \leq \liminf_{m \rightarrow +\infty} \mathbf{W}_\mathbf{c}(\delta_x R_{T/m}^m, \delta_{\tilde{x}} \tilde{R}_{T/m}^m)$. Then, results of Section 4.2 can be applied implying if **H1** and **H2** holds, that for any $x, \tilde{x} \in \mathbb{R}^d$ there exist $C_1, C_2 \geq 0$ such that for any $T \geq 0$, $\mathbf{W}_\mathbf{c}(\delta_x P_T, \delta_{\tilde{x}} \tilde{P}_T) \leq C_1 \rho^T + C_2 c_\infty$. We therefore generalize the result provided in [51] which is specific to the total variation distance. We do not give a specific statement for this result which is mainly technical

and is not the main subject of this paper. Instead, the goal of this section is to study the continuous-time limit of the coupling (4.6) (and not only of its marginals) toward some continuous-time sticky diffusion.

More precisely, let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence of step sizes such that $\lim_{n \rightarrow +\infty} \gamma_n = 0$ and $w_0 \geq 0$. Then, consider the sequence of Markov chains $\{(W_k^{(n)})_{k \in \mathbb{N}} : n \in \mathbb{N}\}$ for any $n \in \mathbb{N}$, $(W_k^{(n)})_{k \in \mathbb{N}}$ is the Markov chain defined by (4.12) with $W_0^{(n)} = w_0$, $\gamma = \gamma_n$ and therefore associated with the Markov kernel Q_{γ_n} . Let $\{(\mathbf{W}_t^{(n)})_{t \in (0, +\infty)} : n \in \mathbb{N}\}$ be the continuous linear interpolation of $\{(W_k^{(n)})_{k \in \mathbb{N}} : n \in \mathbb{N}\}$, i.e. the sequence of continuous processes defined for any $n \in \mathbb{N}$, $t \in (0, +\infty)$ by

$$\mathbf{W}_t^{(n)} = W_{\lfloor t/\gamma_n \rfloor}^{(n)} + \{W_{\lfloor t/\gamma_n \rfloor}^{(n)} - W_{\lfloor t/\gamma_n \rfloor}^{(n)}\} \{t/\gamma_n - \lfloor t/\gamma_n \rfloor\}. \quad (4.19)$$

Note that for any $k \in \mathbb{N}$ and $h \in [0, \gamma_n]$, $\mathbf{W}_{k\gamma_n+h}^{(n)} = W_k + (h/\gamma_n)\{W_{k+1}^{(n)} - W_k^{(n)}\}$. We denote by $\mathbb{W} = C([0, +\infty), \mathbb{R})$ endowed with the uniform topology on compact sets, \mathcal{W} its corresponding σ -field and $(W_t)_{t \geq 0}$ the canonical process defined for any $t \in (0, +\infty)$ and $\omega \in \mathbb{W}$ by $W_t(\omega) = \omega_t$. Denote by $(\mathcal{W}_t)_{t \geq 0}$ the filtration associated with $(W_t)_{t \geq 0}$. Note that $\{(\mathbf{W}_t^{(n)})_{t \in (0, +\infty)} : n \in \mathbb{N}\}$ is a sequence of \mathbb{W} -valued random variables. The main result of this section concerns the convergence in distribution of this sequence.

We consider the following assumption on the function τ_γ .

A 1. *There exists a function $\kappa : [0, +\infty) \rightarrow [0, +\infty)$ such that for any $\gamma \in (0, \bar{\gamma}]$, $\tau_\gamma(w) = w + \gamma\kappa(w)$ and $\kappa(0) = 0$. In addition, κ is L_κ -Lipschitz: for any $w_1, w_2 \in (0, +\infty)$, $|\kappa(w_1) - \kappa(w_2)| \leq L_\kappa |w_1 - w_2|$.*

This is not a restrictive condition since, under **H1**, up to a possible modification of τ_γ , it is always possible to ensure **A1**.

Under **A1**, we consider a sticky process [51, 134, 135], which solves the stochastic differential equation

$$d\mathbf{W}_t = \{\kappa(\mathbf{W}_t) + c_\infty\}dt + 2\sigma \mathbb{1}_{(0, +\infty)}(\mathbf{W}_t)dB_t, \quad (4.20)$$

where $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion. Note that for any initial distribution $\boldsymbol{\mu}_0$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}^d))$, (4.20) admits a unique weak solution by [51, Lemma 6].

The main result of this section is the following.

Theorem 4.13. *Assume **A1**. Then, the sequence $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ defined by (4.19) converges in distribution to the solution $(\mathbf{W}_t)_{t \geq 0}$ of the SDE (4.20).*

The proof of this theorem follows the usual strategy employed to show convergence of a sequence of continuous processes to a Markov process. A first step is to show that under **A1**, $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ is uniformly bounded in L^q for some $q \geq 2$, on $[0, T]$ for any $T \geq 0$.

Proposition 4.14. *Assume **A1**. Then for any $T \geq 0$, there exists $C_T \geq 0$ such that $\sup_{n \in \mathbb{N}} \mathbb{E}[\sup_{t \in [0, T]} \{\mathbf{W}_t^{(n)}\}^4] \leq C_T$ where $(\mathbf{W}_t^{(n)})_{t \geq 0}$ is defined by (4.19).*

Proof. The proof is postponed to Section 4.5.2.1. \square

Then, we are able to obtain the tightness of the sequence of stochastic processes $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$.

Proposition 4.15. *Assume A 1. Then, $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ is tight in \mathbb{W} .*

Proof. The proof is postponed to Section 4.5.2.2. \square

Denote for any $n \in \mathbb{N}$, μ_n the distribution of $(\mathbf{W}_t^{(n)})_{t \geq 0}$ on \mathbb{W} . Then, by Prohorov's Theorem [13, Theorem 5.1, 5.2], $(\mu_n)_{n \in \mathbb{N}}$ admits a limit point. If we now show that every limit point associated with $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ is a solution of the SDE (4.20) using that $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ is tight again and since (4.20) admits a unique weak solution, the proof of Theorem 4.13 will be completed. To establish this result, we use the characterization of solutions of SDEs through martingale problems. More precisely by [27, Theorem 1.27], the distribution μ on \mathbb{W} of $(\mathbf{W}_t)_{t \geq 0}$, solution of (4.20), is the unique solution to the martingale problem associated with μ_0 , the drift function $w \mapsto \kappa(w) + c_\infty$ and the variance function $2\sigma \mathbb{1}_{(0, +\infty)}$, *i.e.* it is the unique probability measure satisfying on the filtered probability space $(\mathbb{W}, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, \mu)$:

- (a) the distribution of W_0 is μ_0 ;
- (b) the processes $(M_t)_{t \geq 0}$, $(N_t)_{t \geq 0}$ defined for any $t \geq 0$ by

$$M_t = W_t - W_0 - \int_0^t \{c_\infty + \kappa(W_u)\} du, \quad N_t = M_t^2 - 4\sigma^2 \int_0^t \mathbb{1}_{(0, +\infty)}(W_u) du, \quad (4.21)$$

are $(\mathcal{W}_t)_{t \geq 0}$ -local martingales.

In other words, it corresponds in showing that $(M_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -local martingales and by [115, Theorem 1.8] identifying its quadratic variation $(\langle M \rangle_t)_{t \geq 0}$ as the process $(4\sigma^2 \int_0^t \mathbb{1}_{(0, +\infty)}(W_u) du)_{t \geq 0}$. Therefore, Theorem 4.13 is a direct consequence of the following result.

Theorem 4.16. *Assume A 1. Let μ_∞ be a limit point of $(\mu_n)_{n \in \mathbb{N}}$. Then, the two processes $(M_t)_{t \geq 0}$ and $(N_t)_{t \geq 0}$ defined by (4.21) are $(\mathcal{W}_t)_{t \geq 0}$ -martingales on $(\mathbb{W}, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, \mu_\infty)$.*

Proof. The proof is postponed to Section 4.5.2.4. \square

Consider the differential operators $\mathcal{A}, \tilde{\mathcal{A}}$ defined for any $\psi \in C^2(\mathbb{R})$ by

$$\begin{aligned} \mathcal{A}\psi(w) &= \{\kappa(w) + c_\infty\}\psi'(w) + 2\mathbb{1}_{(0, +\infty)}(w)\sigma^2\psi''(w) \\ \tilde{\mathcal{A}}\psi(w) &= \{\kappa(w) + c_\infty\}\psi'(w) + 2\sigma^2\psi''(w), \end{aligned} \quad (4.22)$$

where κ is arbitrary extended on \mathbb{R} . Note that \mathcal{A} is the *extended* generator associated with (4.20). A crucial step in the proof of Theorem 4.16 is the following.

Proposition 4.17. *Let $\varphi \in C^3(\mathbb{R})$, satisfying*

$$\sup_{w \in \mathbb{R}} \{ |\varphi(w)/(1+w^2) + |\varphi'(w)/(1+|w|) + |\varphi''(w) + |\varphi^{(3)}(w)| \} < +\infty. \quad (4.23)$$

Then, for any $N \in \mathbb{N}$, $(t_1, \dots, t_N, s, t) \in [0, +\infty)^{N+2}$, $0 \leq t_1 \leq \dots \leq t_N \leq s < t$, $\psi : [0, +\infty)^N \rightarrow \mathbb{R}$, positive, continuous and bounded, it holds that

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left(\varphi(\mathbf{W}_t^{(n)}) - \varphi(\mathbf{W}_s^{(n)}) - \int_s^t \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du \right) \psi(\mathbf{W}_{t_1}^{(n)}, \dots, \mathbf{W}_{t_N}^{(n)}) \right] = 0. \quad (4.24)$$

In addition, if $\varphi''(w) \geq 0$ for any $w \in \mathbb{R}$, it holds that

$$\limsup_{n \rightarrow +\infty} \mathbb{E} \left[\left(\varphi(\mathbf{W}_t^{(n)}) - \varphi(\mathbf{W}_s^{(n)}) - \int_s^t \tilde{\mathcal{A}}\varphi(\mathbf{W}_u^{(n)}) du \right) \psi(\mathbf{W}_{t_1}^{(n)}, \dots, \mathbf{W}_{t_N}^{(n)}) \right] \leq 0. \quad (4.25)$$

Proof. The proof is postponed to Section 4.5.2.3. □

Note that while Proposition 4.17-(4.24) is in general sufficient to conclude on the convergence of the sequence of processes $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ (see e.g. [54]), in our setting, it is not enough to complete the proof of Theorem 4.13 since the diffusion coefficient associated with \mathcal{A} is discontinuous. To circumvent this issue, we adapt to our sequence $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$ the same strategy employed in [113, Proposition 6].

4.4 An application in Bayesian statistics: parameter estimation in an ODE

4.4.1 Setting and verifying the assumptions

Consider an ordinary differential equation ODE on \mathbb{R}^n of the form

$$\dot{x}_\theta(t) = f_\theta(x_\theta(t), t), \quad x_\theta(0) = x_0 \in \mathbb{R}^n, \quad (4.26)$$

where $\{f_\theta : \theta \in \mathbb{R}^d\}$ is a family of function from $\mathbb{R}^n \times [0, +\infty)$ to \mathbb{R}^n parametrized by some parameter $\theta \in \mathbb{R}^d$. In all this section $x_0 \in \mathbb{R}^n$ is assumed to be fixed and we consider the following assumption.

AO1. *For all $\theta \in \mathbb{R}^d$ there exists a unique solution of (4.26) defined for all positive times, which we denote by $(x_\theta(t))_{t \geq 0}$. In addition, the functions $(\theta, x, t) \in \mathbb{R}^d \times \mathbb{R}^n \times [0, +\infty) \mapsto f_\theta(x, t)$ and $(\theta, t) \in \mathbb{R}^d \times [0, +\infty) \mapsto x_\theta(t)$ are continuously differentiable.*

In fact the continuous differentiability of $(\theta, t) \mapsto x_\theta(t)$ is a consequence of the one of $(\theta, x, t) \mapsto f_\theta(x, t)$, see e.g. [138, Theorem 4.D].

To fix ideas, throughout this section, we will repeatedly discuss the following case of a logistic equation.

Example 4.18. For $r \in C^1(\mathbb{R}, \mathbb{R}_+)$, set $f_\theta(x) = x(1 - r(\theta)x)$ for any $\theta, x \in \mathbb{R}$, so that (4.26) reads

$$\dot{x}_\theta(t) = x_\theta(t)(1 - r(\theta)x_\theta(t)), \quad x_\theta(0) = x_0,$$

with $x_0 \geq 0$. In this example, **AO1** holds and, r and x_0 being positive, for all $\theta \in \mathbb{R}$, the solution of (4.26) is such that $x_\theta(t) \in [0, e^t x_0]$ for all $t \geq 0$. Indeed, $x \mapsto x(1 - r(\theta))$ is locally Lipschitz continuous, which yields existence and uniqueness of a maximal solution. Since 0 is always an equilibrium, solutions stay positive, from which $x'_\theta(t) \leq x_\theta(t)$ for all $t \geq 0$, implying that $x_\theta(t) \leq e^t x_0$ for all $t \geq 0$. This also implies non-explosion, hence the solution is defined on $[0, +\infty)$.

We consider the problem of estimating θ based on some observation of a trajectory of the ODE. More precisely, for $T > 0$, $N \in \mathbb{N}^*$, $(t_1, \dots, t_N) \in \mathbb{R}^N$, $0 < t_1 < \dots < t_N = T$, the statistical model corresponding to the observation $\mathbf{y} = (y_i)_{i \in \{1, \dots, N\}} \in (\mathbb{R}^n)^N$ is given by

$$y_i = x_\theta(t_i) + \varepsilon_i, \quad (4.27)$$

for $\theta \in \mathbb{R}^d$ and where $(\varepsilon_i)_{i \in \{1, \dots, N\}}$ are independent and identically random variables on \mathbb{R}^n distributed according to some known positive density φ_ε with respect to the Lebesgue measure. Given a prior distribution with positive density π_0 on \mathbb{R}^d , the a posteriori distribution for this model admits a positive density π with respect to the Lebesgue measure which is characterized by the potential U given by (up to an additive constant)

$$-\log \pi(\theta) = U(\theta) = -\ln \pi_0(\theta) - \sum_{i=1}^N \ln \varphi_\varepsilon(y_i - x_\theta(t_i)).$$

We consider the following assumption on π_0 and φ_ε setting $-\log(\pi_0) = U_0$.

AO 2. The functions π_0 and φ_ε are twice continuously differentiable and there exist $\mathfrak{m}_U > 0$, $L_U, R_U \geq 0$ such that ∇U_0 is L_U -Lipschitz continuous and for any $\theta, \tilde{\theta} \in \mathbb{R}^d$ with $\|\theta - \tilde{\theta}\| \geq R_U$,

$$\langle \theta - \tilde{\theta}, \nabla U_0(\theta) - \nabla U_0(\tilde{\theta}) \rangle \geq \mathfrak{m}_U \|\theta - \tilde{\theta}\|^2.$$

In practice, expectations with respect to the posterior distribution can be approximated by ergodic means of the Unadjusted Langevin Algorithm (ULA), namely the Markov chain

$$X_{k+1} = X_k - \gamma \nabla U(X_k) + \sqrt{2\gamma} Z_{k+1}, \quad (4.28)$$

where $\gamma > 0$ and $(Z_k)_{k \in \mathbb{N}}$ are independent and identically standard Gaussian variables. The long-time convergence of this algorithm and the numerical bias on the invariant measure due to the time discretization are well understood, see e.g. [15, 30, 32, 43, 44] and references therein. However, in the present case, it is not possible to sample this Markov chain, as the exact computation of

$$\nabla U(\theta) = -\nabla_\theta \ln \pi_0(\theta) + \sum_{i=1}^N \nabla_\theta x_\theta(t_i) \nabla_x \ln \varphi_\varepsilon(y_i - x_\theta(t_i)), \quad (4.29)$$

is not possible in most cases because of the term involving x_θ and $\nabla_\theta x_\theta$. Here ∇_θ and ∇_x denote the gradient operator with respect to θ and x respectively. Therefore, only approximations of these two functions can be used in place of $(x_\theta(t_i), \nabla_\theta x_\theta(t_i))_{i \in [0, N]}$, which leads to an additional discretization bias. Our results based on the sticky coupling yields a quantitative bound on this error (with respect to the ideal ULA above). Let us detail this statement.

First, remark that $t \mapsto z_\theta(t) = (x_\theta(t), \nabla_\theta x_\theta(t))$ solves

$$\dot{z}_\theta(t) = F_\theta(z_\theta(t), t) \quad z_\theta(0) = z_0 = (x_0, 0) \quad (4.30)$$

on $\mathbb{R}^n \times \mathcal{M}_{d,n}(\mathbb{R})$ with for any $x \in \mathbb{R}^n$, $\mathbf{A} \in \mathcal{M}_{d,n}(\mathbb{R})$, $\theta \in \mathbb{R}^d$, $t \geq 0$,

$$F_\theta((x, \mathbf{A}), t) = (f_\theta(x, t), \nabla_\theta f_\theta(x, t) + \mathbf{A} \nabla_x f_\theta(x, t)) . \quad (4.31)$$

Provided f_θ , $\nabla_\theta f_\theta$ and $\nabla_x f$ are computable, in practice this ODE can be approximated by standard numerical schemes. For instance, a basic explicit Euler discretization with time-step $h > 0$ is given by

$$\tilde{z}_\theta^h(0) = z_0, \quad \tilde{z}_\theta^h((k+1)h) = \tilde{z}_\theta^h(kh) + hF_\theta(\tilde{z}_\theta^h(kh), kh) \quad \forall k \in \mathbb{N}$$

and

$$\tilde{z}_\theta^h(t) = \tilde{z}_\theta^h(kh) + (t - kh)F_\theta(\tilde{z}_\theta^h(kh), kh), \quad t \in [kh, (k+1)h). \quad (4.32)$$

To establish the consistency of this approximation (with some uniformity in θ), we consider the following condition.

AO3. *There exist $L_F, L'_F, C_F, \delta > 0$ and a compact set $\mathbf{K} \subset \mathbb{R}^n \times \mathcal{M}_{d,n}(\mathbb{R})$ such that the following holds. For all $t \in [0, T]$ and $\theta \in \mathbb{R}^d$, the ball centered at $z_\theta(t)$ and radius δ is included in \mathbf{K} . Moreover, for all $z, \tilde{z} \in \mathbf{K}$, $t, s \in [0, T]$ and $\theta, \tilde{\theta} \in \mathbb{R}^d$, $\|F_\theta(z, t)\| \leq C_F$ and*

$$\|F_\theta(z, t) - F_{\tilde{\theta}}(\tilde{z}, s)\| \leq L_F \|\theta - \tilde{\theta}\| + L'_F (\|z - \tilde{z}\| + |t - s|). \quad (4.33)$$

Proposition 4.19. *Assume **AO1** and **AO3**. There exist $\bar{h}, C > 0$ such that for all $h \in (0, \bar{h}]$, $\theta \in \mathbb{R}^d$, and $t \in [0, T]$, $\tilde{z}_\theta^h(t) \in \mathbf{K}$ and $\sum_{i=1}^N \|z_\theta(t_i) - \tilde{z}_\theta^h(t_i)\| \leq Ch$, where z_θ solves (4.30) and \tilde{z}_θ^h is given by (4.32).*

Proof. The proof is postponed to Section 4.5.3. □

Example 4.20 (Continuation of Example 4.18). *Let us check for instance that **AO3** is satisfied for Example 4.18 provided that r is twice continuously differentiable on $[0, +\infty)$ with, for some $L_r, L'_r, L''_r > 0$,*

$$r, r' \text{ and } r'' \text{ uniformly bounded respectively by } L_r, L'_r \text{ and } L''_r . \quad (4.34)$$

We may consider for example $r : \theta \mapsto a_1 \theta^2 / (\theta^2 + a_2)$ for $a_1, a_2 \in (0, +\infty)$. Recall that $f_\theta(x) = x(1 - r(\theta)x)$ and $x_\theta(t) \in [0, e^t x_0]$ for all $\theta \in \mathbb{R}$ and all $t \in [0, T]$, so that for any $t \geq 0$ and $\theta \in \mathbb{R}^d$,

$$|\partial_\theta f_\theta(x_\theta(t))| = |r'(\theta)x_\theta^2(t)| \leq L'_r e^{2t} x_0^2.$$

for all $t \geq 0$. Notice that $1/r(\theta)$ is an equilibrium of the equation, so that it cannot be crossed by other solutions. Hence, on the one hand, if $1 \leq r(\theta)x_0$ then x_θ is non-increasing (in particular $x_\theta(t) \leq x_0$ for all $t \geq 0$) while, on the other hand, if $1 \geq r(\theta)x_0$, then $1 \geq r(\theta)x_\theta(t)$ for all $t \geq 0$. In both cases, we get that for all $t \geq 0$,

$$|\partial_x f_\theta(x_\theta(t))| = |1 - 2r(\theta)x_\theta(t)| \leq 1 + 2L_r x_0.$$

Combining the two previous bounds,

$$\|\partial_\theta x_\theta(t)\| \leq \int_0^t \left(L'_r e^{2s} x_0^2 + (1 + 2L_r x_0) \|\partial_\theta x_\theta(s)\| \right) ds,$$

and thus by Grönwall's inequality, $\|\partial_\theta x_\theta(t)\| \leq M_t = L'_r x_0^2 t e^{(3+2L_r x_0)t}$, for all $t \in [0, T]$ and $\theta \in \mathbb{R}$. Then, for any $\delta > 0$, **AO 3** is satisfied with $\mathbf{K} = [-\delta, e^T x_0 + \delta] \times [-M_T - \delta, M_T + \delta]$. Since \mathbf{K} is compact and as, by (4.31), for any $x \in \mathbb{R}^n$, $\mathbf{A} \in \mathcal{M}_{d,n}(\mathbb{R})$, $\theta \in \mathbb{R}^d$, $t \geq 0$,

$$F_\theta((x, \mathbf{A}), t) = \left(x(1 - r(\theta)x), r'(\theta)x^2 + \mathbf{A}(1 - 2r(\theta)x) \right),$$

then (4.33) easily follows from the condition (4.34).

Since other schemes may be used, in particular higher-order ones, we consider more generally in the following a solver $\Psi^h : \mathbb{R}^d \rightarrow (\mathbb{R}^n \times \mathcal{M}_{d,n}(\mathbb{R}))^N$ for $h > 0$ satisfying the condition:

AO 4. There exist $\bar{h}, \mathbf{C}_\Psi, \alpha > 0$ such that, for any $\theta \in \mathbb{R}^d$ and $h \in (0, \bar{h}]$,

$$\sum_{i=1}^N \|z_\theta(t_i) - \Psi_i^h(\theta)\| \leq \mathbf{C}_\Psi h^\alpha,$$

where z_θ is a solution of (4.30) and $\Psi_i^h : \mathbb{R}^d \rightarrow \mathbb{R}^n \times \mathcal{M}_{d,n}(\mathbb{R})$ is the i -th component of Ψ^h .

When **AO 3** and **AO 4** are both satisfied, without loss of generality, we assume furthermore that \bar{h} is sufficiently small so that $\mathbf{C}_\Psi \bar{h}^\alpha \leq \delta$. This implies that $\Psi_i^h(\theta) \in \mathbf{K}$ for all $\theta \in \mathbb{R}^d$, $i \in \{1, \dots, N\}$ and $h \in (0, \bar{h}]$.

Writing $\Psi_i^h(\theta) = (\tilde{x}_\theta^h(t_i), G_\theta^h(t_i))$, we can consider for any $\theta \in \mathbb{R}^d$,

$$\tilde{b}_h(\theta) = -\nabla_\theta \ln \pi_0(\theta) + \sum_{i=1}^N G_\theta^h(t_i) \cdot \nabla_x \ln \varphi_\varepsilon \left(y_i - \tilde{x}_\theta^h(t_i) \right), \quad (4.35)$$

as an approximation of ∇U (4.29). Remark that, now, in contrast to $b(\theta) = \nabla U(\theta)$, it is possible in practice to evaluate $\tilde{b}_h(\theta)$ for $\theta \in \mathbb{R}^d$, provided $\nabla_x \ln \varphi_\varepsilon$ and $\nabla_\theta \ln \pi_0$ can be evaluated. We now assess the error due to the use of \tilde{b}_h in place of the exact gradient in (4.28) by verifying that the assumption of Section 4.2 are satisfied. For $\gamma, h > 0$ and $\theta \in \mathbb{R}^d$, denote

$$\mathbf{T}_\gamma(\theta) = \theta - \gamma \nabla U(\theta), \quad \tilde{\mathbf{T}}_{\gamma,h}(\theta) = \theta - \gamma \tilde{b}_h(\theta). \quad (4.36)$$

When **AO 2** and **AO 3** are both satisfied, there exist $C_\varphi, L_\varphi > 0$ such that for all $i \in \{1, \dots, N\}$, the function s_i on the compact set \mathbf{K} given for any $x \in \mathbb{R}^n$ and $\mathbf{A} \in \mathcal{M}_{d,n}(\mathbb{R})$, by

$$\mathbf{s}_i(x, \mathbf{A}) = \mathbf{A} \nabla_x \ln \varphi_\varepsilon(y_i - x) \quad (4.37)$$

is bounded by C_s and L_s -Lipschitz continuous on $\mathbb{R}^n \times \mathcal{M}_{d,n}(\mathbb{R})$.

Proposition 4.21. *Under **AO 1**, **AO 2**, **AO 3** and **AO 4**, for any $h \in (0, \bar{h})$, the functions T_γ and $\tilde{T}_{\gamma,h}$ given by (4.36) satisfy for any $\bar{\gamma} > 0$, **H 1** and **H 2**, with*

$$c_\infty = C_\Psi L_s h^\alpha, \quad R_1 = \frac{2NC_s}{m_U} \vee R_U, \quad m = \frac{m_U}{2}, \quad L = L_U + L_s L_F \sum_{i=1}^N t_i e^{L'_F t_i}.$$

Proof. The proof is postponed to Section 4.5.3. □

Under the conditions of Proposition 4.21 and using the results of Section 4.2, we get that the Markov chains $(X_k)_{k \in \mathbb{N}}$ and $(\tilde{X}_k)_{k \in \mathbb{N}}$ associated to T_γ and $\tilde{T}_{\gamma,h}$ given by (4.36) have unique invariant measure π_γ and $\tilde{\pi}_{\gamma,h}$, and that there exist $\bar{\gamma}, \bar{h}, C > 0$ such that for all $\gamma \in (0, \bar{\gamma}]$ and $h \in (0, \bar{h}]$,

$$\|\pi_\gamma - \tilde{\pi}_{\gamma,h}\|_{\text{TV}} \leq Ch. \quad (4.38)$$

Example 4.22 (Continuation of Example 4.18). *As a conclusion, consider the logistic case of Example 4.18 with the Euler scheme (4.32), assuming that $r \in C^2(\mathbb{R}, [0, +\infty))$ satisfies (4.34). Then, by Example 4.18 and Example 4.20, **AO 1**, **AO 3** and **AO 4** hold. Assuming moreover that π_0 and φ_ε are Gaussian, then **AO 2** also holds and we obtain (4.38).*

4.4.2 Numerical results

We illustrate our findings on two particular ODEs. First, we consider the ODE associated with the Van der Pol oscillator corresponding to the second order ODE:

$$\ddot{x}_\theta(t) - \theta(1 - x_\theta(t)^2)\dot{x}_\theta(t) + x_\theta(t) = 0, \quad (4.39)$$

where $\theta \in \mathbb{R}$ is the parameter to infer. It corresponds to (4.26) with $f_\theta(x_1, x_2) = (x_2, \theta(1 - x_1^2)x_2 - x_1)$. We generate synthetic data solving (4.39) using the 4th Runge-Kutta method for $T = 10$ and $\theta = 1$. We then select $(x_\theta(t_i))_{i=1}^{25}$ for $(t_i)_{i=1}^{25}$ uniformly chosen in $[0, T]$. The observations $\mathbf{y} = (y_i)_{i=1}^{25}$ are obtained from $(x_\theta(t_i))_{i=1}^{25}$ adding i.i.d. zero-mean Gaussian noise with variance 0.5. We consider the corresponding statistical model (4.27) where $(\varepsilon_i)_{i=1}^{25}$ are i.i.d. zero-mean Gaussian random variables with variance 0.5. We consider as prior π_0 , the zero-mean Gaussian distribution with variance 0.5. We then use ULA with $\gamma = 10^{-2}$, for which the gradient is estimated using the Euler method with the time steps $h \in \{0.05, 0.01, 0.001\}$. Figure 4.1a represents the histograms corresponding to the different Markov chains after 10^5 iterations with a burn-in of 10^4 steps. Gaussian kernel density approximation of these histograms are estimated

and used as proxy for the density of the invariant distributions $\pi_{\gamma,h}$ of the Markov chain $(\tilde{X}_k)_{k \in \mathbb{N}}$ associated to $\tilde{T}_{\gamma,h}$ given by (4.36). To obtain a proxy for the density of π_γ , the stationary distribution of $(X_k)_{k \in \mathbb{N}}$ associated to T_γ , we use the same procedure but using the Euler method with $h = 0.0001$. We then estimate the total variation between π_γ and $\pi_{\gamma,h}$ for $h \in \{0.005, 0.004, 0.003, 0.0025, 0.0015, 0.001, 0.00075, 0.0005\}$ using numerical integration. The corresponding results over 10 replications are reported in Figure 4.1b. We can observe that the total variation distance linearly decreases with h which supports our findings.

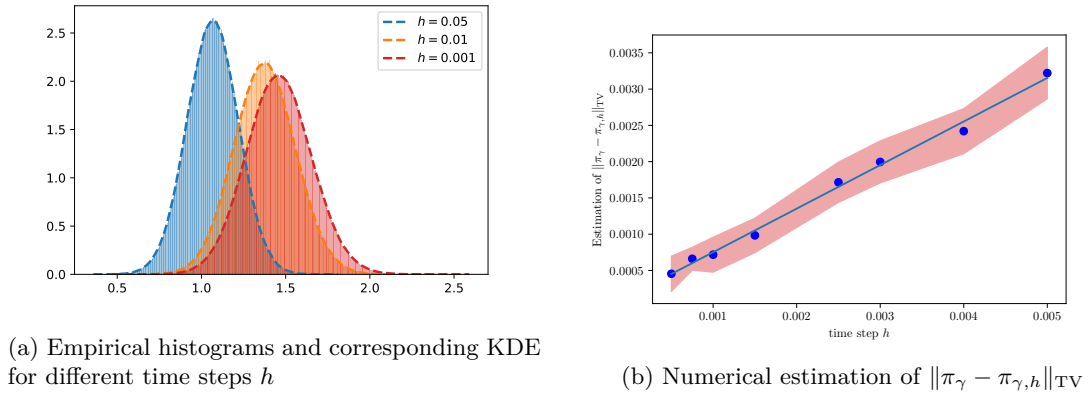


Figure 4.1: Numerical illustrations for the Van der Pol oscillator (4.39)

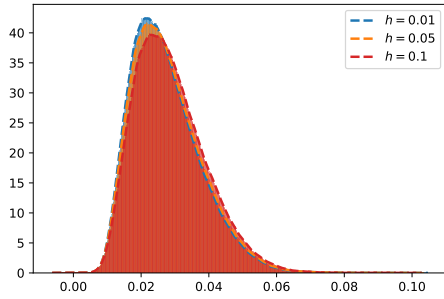
For our second experiment, we consider the Lotka-Volterra model describing the evolution of the population of two interacting biological species denoted by $t \mapsto x_\theta(t) = (u_\theta(t), v_\theta(t))$. The dynamics of these two populations are assumed to be governed by the system of equations given by:

$$\dot{u}_\theta(t) = (\alpha - \beta v_\theta(t))u_\theta(t), \quad \dot{v}_\theta(t) = (-\gamma + \delta u_\theta(t))v_\theta(t), \quad (4.40)$$

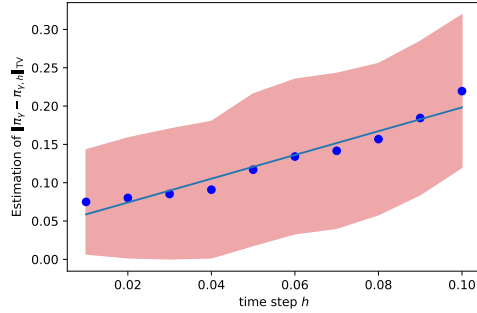
where $\theta = (\alpha, \beta, \gamma, \delta)$ is the parameter to infer. We follow the same methodology presented in [96, Chapter 16]. For this experiment, we consider an other statistical model as previously and generate synthetic data $\mathbf{y} = (y_i)_{i=1}^{50}$ accordingly and associated with observation times $(t_i)_{i=1}^{50}$ uniformly spaced in $[0, T]$ for $T = 10$ and the true parameter $\theta_0 = (0.6, 0.025, 0.8, 0.025)$. More precisely, for any $i \in \{1, \dots, 50\}$, $y_i = (u_i^y, v_i^y)$ with $u_i^y = u_\theta(t_i)e^{\varepsilon_{u,i}}$, $v_i^y = v_\theta(t_i)e^{\varepsilon_{v,i}}$ and $(\varepsilon_{u,j}, \varepsilon_{v,j})_{j=1}^{50}$ are i.i.d. one dimensional zero-mean Gaussian random variables with covariance matrix I_2 . The prior π_0 is set to be the Gaussian distribution on \mathbb{R}^4 with means $(1, 0.05, 1, 0.05)$ and standard deviations $(0.5, 0.05, 0.5, 0.05)$. The posterior distribution is then given by $\pi(\theta|\mathbf{y}) \propto \exp(-U(\theta))$, where

$$U(\theta) = -\log \pi_0(\theta) + \sum_{i=1}^{50} \left(\frac{(\log u_i^y - \log u_\theta(t_i))^2 + (\log v_i^y - \log v_\theta(t_i))^2}{2\zeta^2} \right).$$

We then use ULA with $\gamma = 5 \times 10^{-5}$, for which the gradient is estimated using the Euler method. We focus here on the second component of the chain. The results for the other components are similar. Figure 4.2a represents the histograms for the second component corresponding to the different Markov chains after 10^7 iterations with a burn-in of 10^3 steps. Gaussian kernel density approximation of these histograms are estimated and used as proxy for the marginal density of the invariant distributions $\pi_{\gamma,h}$ of the Markov chain $(\tilde{X}_k)_{k \in \mathbb{N}}$ associated to $\tilde{T}_{\gamma,h}$ given by (4.36). To obtain a proxy for the marginal density of π_γ , the stationary distribution of $(X_k)_{k \in \mathbb{N}}$ associated to T_γ , we use the same procedure but using the Euler method with $h = 0.0001$. We then estimate the total variation between π_γ and $\pi_{\gamma,h}$ for $h \in \{k \times 10^{-2} : k \in \{1, \dots, 10\}\}$ using numerical integration. The corresponding results over 10 replications are reported in Figure 4.2b. We can observe that the total variation distance still linearly decreases with h .



(a) Empirical histograms and corresponding KDE for different time steps h



(b) Numerical estimation of $\|\pi_\gamma - \pi_{\gamma,h}\|_{TV}$

Figure 4.2: Numerical illustrations for the Lotka-Volterra model (4.40)

4.5 Postponed proofs

4.5.1 Postponed proofs of Section 4.2

4.5.1.1 Proof of Proposition 4.1

Recall that under **H1**, we have for any $\gamma \in (0, \bar{\gamma}]$,

$$\begin{aligned} \sup_{x \in \mathbb{R}^d, x \neq 0} \{ \|T_\gamma(x) - T_\gamma(0)\| / \|x\| \} &\leq (1 + \gamma L), \\ \sup_{x \in \mathbb{R}^d, \|x\| \geq R_1} \{ \|T_\gamma(x) - T_\gamma(0)\| / \|x\| \} &\leq (1 - \gamma m). \end{aligned} \tag{4.41}$$

Define for any $x \in \mathbb{R}^d$

$$\bar{V}_{c,\gamma}(x) = \exp(c\|x - T_\gamma(0)\|^2), \text{ for } c = m/(16\sigma^2). \tag{4.42}$$

Note that since $\sup_{\gamma \in (0, \bar{\gamma}]} \|\mathbb{T}_\gamma(0)\| < +\infty$ by **H1**, $\bar{V}_{c,\gamma}$ goes to $+\infty$ at infinity and

$$\lim_{\|x\| \rightarrow +\infty} \left\{ \sup_{\gamma \in (0, \bar{\gamma}]} \bar{V}_{c,\gamma}/V_c \right\}(x) = 1 = \lim_{\|x\| \rightarrow +\infty} \left\{ \inf_{\gamma \in (0, \bar{\gamma}]} \bar{V}_{c,\gamma}/V_c \right\}(x) = 1.$$

Therefore, by [33, Corollary 11], it is sufficient to show that there exist $\bar{\gamma}_1 \in (0, \bar{\gamma}]$, $\lambda \in (0, 1)$, $A \geq 0$ such that for any $\gamma \in (0, \bar{\gamma}_1]$ and $x \in \mathbb{R}^d$,

$$R_\gamma \bar{V}_{c,\gamma}(x) \leq \lambda^\gamma \bar{V}_{c,\gamma}(x) + \gamma A. \quad (4.43)$$

We show that it holds with

$$\begin{aligned} \bar{\gamma}_1 &= \bar{\gamma} \wedge \{1/\mathfrak{m}\}, \quad \lambda = \exp(-\mathfrak{c}M^2/4), \quad M = R_1 \vee (8d\sigma^2/\mathfrak{m})^{1/2}, \\ A &= \lambda^{\bar{\gamma}} \exp(cM^2 + \bar{\gamma}\{B_1M^2 + B_2 - \log(\lambda)\})\{B_1M^2 + B_2 - \log(\lambda)\}, \quad B_2 = 2dc\sigma^2, \\ B_1 &= 4C_1, \quad C_1 = C_2 \vee C_2^2\bar{\gamma}, \quad C_2 = (2L) \vee (L\bar{\gamma}) \vee (8c\sigma^2). \end{aligned} \quad (4.44)$$

Define for any $x \in \mathbb{R}^d$, $\bar{\mathbb{T}}_\gamma(x) = \mathbb{T}_\gamma(x) - \mathbb{T}_\gamma(0)$. Note that for any $\gamma \in (0, \bar{\gamma}_1]$, $2c\sigma^2\gamma \leq 1$ by definition of c (4.42) and $\bar{\gamma}_1$. Let $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}_1]$. Then, we obtain using that $\int_{\mathbb{R}} e^{az+bz^2-z^2/2} dz = (2\pi(1-2b)^{-1})^{d/2} e^{a^2/(2(1-2b))}$ for any $a \in \mathbb{R}$ and $b \in [0, 1/2]$,

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp(c\|\bar{\mathbb{T}}_\gamma(x) + (\sigma^2\gamma)^{1/2}z\|^2 - \|z\|^2/2) dz \\ &= (1 - 2c\sigma^2\gamma)^{-d/2} \exp\{c(1 - 2c\sigma^2\gamma)^{-1}\|\bar{\mathbb{T}}_\gamma(x)\|^2\}. \end{aligned}$$

We now distinguish the case $\|x\| \geq M$ and $\|x\| < M$.

In the first case, we get by (4.41), $(1 - 2c\sigma^2\gamma)^{-1} \leq 8c\sigma^2\gamma$ and $(1 - \mathfrak{m}\gamma)^2 \leq 1 - \mathfrak{m}\gamma$ since $2c\sigma^2\gamma \leq 1/2$ and $\gamma \leq 1/\mathfrak{m}$, by definition of c and $\bar{\gamma}_1$ (4.42)-(4.44),

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &\leq (1 - 2c\sigma^2\gamma)^{-d/2} \exp\left[c(1 - \mathfrak{m}\gamma)(1 + 8c\sigma^2\gamma)\|x\|^2\right] \\ &\leq \exp\left[c(1 - \mathfrak{m}\gamma/4)\|x\|^2 + 2dc\sigma^2\gamma - \mathfrak{m}c\gamma\|x\|^2/4\right] \leq \lambda^\gamma \bar{V}_{c,\gamma}(x), \end{aligned} \quad (4.45)$$

where we used for the penultimate inequality that $-\log(1-t) \leq 2t$ for $t \in [0, 1/2]$.

For the case $\|x\| \leq M$, by (4.41), $(1 - 2c\sigma^2\gamma)^{-1} \leq 8c\sigma^2\gamma$ and $(1 - \mathfrak{m}\gamma)^2 \leq 1 - \mathfrak{m}\gamma$ since $2c\sigma^2\gamma \leq 1/2$ and $\gamma \leq 1/\mathfrak{m}$, by (4.42)-(4.44), and using $-\log(1-t) \leq 2t$ for $t \in [0, 1/2]$ again,

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &\leq (1 - 2c\sigma^2\gamma)^{-d/2} \exp\left[c(1 + L\gamma)^2(1 + 8c\sigma^2\gamma)\|x\|^2\right] \\ &\leq \exp\left[c(1 + \gamma B_1)\|x\|^2 + \gamma B_2\right]. \end{aligned}$$

Using that $e^t - 1 \leq te^t$ for $t \geq 0$, we obtain that

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &= \lambda^\gamma \bar{V}_{c,\gamma}(x) + \lambda^\gamma \bar{V}_{c,\gamma}(x) \left\{ \exp\left[c\gamma B_1\|x\|^2 + \gamma B_2 - \gamma \log(\lambda)\right] - 1 \right\} \\ &\leq \lambda^\gamma \bar{V}_{c,\gamma}(x) + \gamma \lambda^{\bar{\gamma}} \bar{V}_{c,\gamma}(x) \left\{ B_1\|x\|^2 + B_2 - \log(\lambda) \right\} \exp\left[\bar{\gamma}(B_1\|x\|^2 + B_2 - \log(\lambda))\right], \end{aligned}$$

which combined with (4.45) completes the proof of (4.43).

4.5.1.2 Proof of Proposition 4.3

First note that under **H1** and **H2**, for all $x \in \mathbb{R}^d$, since

$$\left\| \tilde{\mathbb{T}}_\gamma(x) - \mathbb{T}_\gamma(0) \right\| \leq \gamma c_\infty + \|\mathbb{T}_\gamma(x) - \mathbb{T}_\gamma(0)\|$$

and from (4.41),

$$\sup_{x \in \mathbb{R}^d, \|x\| \geq R_1} \left[\|\tilde{\mathbb{T}}_\gamma(x) - \mathbb{T}_\gamma(0)\| / \{(1 - \gamma \mathfrak{m})\|x\| + \gamma c_\infty\} \right] \leq 1. \quad (4.46)$$

Define for any $x \in \mathbb{R}^d$

$$\bar{V}_{c,\gamma}(x) = \exp(c\|x - \mathbb{T}_\gamma(0)\|^2), \text{ for } c = m/(16\sigma^2). \quad (4.47)$$

Note that $\lim_{\|x\| \rightarrow +\infty} \bar{V}_{c,\gamma}(x) = +\infty$ and $\lim_{\|x\| \rightarrow +\infty} \{\bar{V}_{c,\gamma}/V_c\}(x) = 1$. In addition, by **H1** and **H2**, \tilde{R}_γ is strongly aperiodic, Leb-irreducible and all compact sets are 1-small. It is therefore sufficient to show by [99, Theorem 16.0.1] that there exists $c > 0$, such that for any $\gamma \in (0, \bar{\gamma}_1]$ (recall $\bar{\gamma}_1 = \bar{\gamma} \wedge \{1/m\}$), there exists $\lambda_\gamma \in [0, 1)$ and $A_\gamma \geq 0$, such that

$$\tilde{R}_\gamma V_c \leq \lambda_\gamma V_c + A_\gamma. \quad (4.48)$$

Define for any $x \in \mathbb{R}^d$, $\bar{\mathbb{T}}_\gamma(x) = \tilde{\mathbb{T}}_\gamma(x) - \mathbb{T}_\gamma(0)$. Note that for any $\gamma \in (0, \bar{\gamma}_1]$, $2c\sigma^2\gamma \leq 1$ by definition of c (4.47) and $\bar{\gamma}_1$. Let $x \in \mathbb{R}^d$ and $\gamma \in (0, \bar{\gamma}_1]$. Then, we obtain using that $\int_{\mathbb{R}} e^{az+bz^2-z^2/2} dz = (2\pi(1-2b)^{-1})^{d/2} e^{a^2/(2(1-2b))}$ for any $a \in \mathbb{R}$ and $b \in [0, 1/2)$,

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \exp(c\|\bar{\mathbb{T}}_\gamma(x) + (\sigma^2\gamma)^{1/2}z\|^2 - \|z\|^2/2) dz \\ &= (1 - 2c\sigma^2\gamma)^{-d/2} \exp\{c(1 - 2c\sigma^2\gamma)^{-1}\|\bar{\mathbb{T}}_\gamma(x)\|^2\}. \end{aligned}$$

If $\|x\| \geq M$, we get by (4.46), $(1 - 2c\sigma^2\gamma)^{-1} \leq 8c\sigma^2\gamma$ and $(1 - m\gamma)^2 \leq 1 - m\gamma$ since $2c\sigma^2\gamma \leq 1/2$ and $\gamma \leq 1/m$, by (4.47) and definition of $\bar{\gamma}_1$,

$$\begin{aligned} R_\gamma \bar{V}_{c,\gamma}(x) &\leq (1 - 2c\gamma\sigma^2)^{-d/2} \exp\left[c(1 + 8c\sigma^2\gamma)\{(1 - m\gamma)\|x\| + c_\infty\}^2\right] \\ &\leq (1 - 2c\gamma\sigma^2)^{-d/2} \exp\left[c\{(1 - m\gamma/2)\|x\| + (1 + 8c\sigma^2\gamma)c_\infty\}^2\right]. \end{aligned}$$

Therefore, we get $\liminf_{\|x\| \rightarrow +\infty} [R_\gamma \bar{V}_{c,\gamma}(x)/\bar{V}_{c,\gamma}(x)] = 0$ which completes the proof of (4.48).

4.5.1.3 Proof of Proposition 4.6

The proof is based on this technical lemma.

Lemma 23. For any $g \in \mathbb{R}$, $u \in [0, 1]$ and $\gamma \in (0, \bar{\gamma}]$, $a, b \in [0, +\infty)$, $a \leq b$,

$$\mathcal{H}_\gamma(a, g, u) \leq \mathcal{H}_\gamma(b, g, u) \quad (4.49)$$

Proof. Let $u \in [0, 1]$, $g \in \mathbb{R}$ and $a, b \in [0, +\infty)$, $a \leq b$. We first prove that for any $c \in \mathbb{R}_+$ such that $c - 2(\sigma^2\gamma)^{1/2}g < 0$,

$$\mathcal{H}_\gamma(c, g, u) = 0, \quad (4.50)$$

which implies that for any $c \in \mathbb{R}_+$, $\mathcal{H}_\gamma(c, g, u) \geq 0$. We need to consider the following two cases.

(a) If $c - (\sigma^2\gamma)^{1/2}g < 0$. Then, using $-(\sigma^2\gamma)^{1/2}g \leq c - (\sigma^2\gamma)^{1/2}g$, $\varphi_{\sigma^2\gamma}((\sigma^2\gamma)^{1/2}g) = \varphi_{\sigma^2\gamma}(-(\sigma^2\gamma)^{1/2}g)$, and $t \mapsto \varphi_{\sigma^2\gamma}(t)$ is decreasing on $[0, +\infty)$, we get $\bar{p}_\gamma(c, g) = 1$. Therefore (4.50) holds.

(b) If $0 \leq c - (\sigma^2\gamma)^{1/2}g \leq (\sigma^2\gamma)^{1/2}g$, we obtain similarly to the first case that $\bar{p}_\gamma(c, g) = 1$ and therefore (4.50) holds.

We now show (4.49). It is straightforward by (4.50) if $0 > a - 2(\sigma^2\gamma)^{1/2}g$. If $0 \leq a - 2(\sigma^2\gamma)^{1/2}g$. By using $t \mapsto \varphi_{\sigma^2\gamma}(t)$ is decreasing on $[0, +\infty)$, we obtain

$$\bar{p}_{\sigma^2\gamma}(a, g) \geq \bar{p}_{\sigma^2\gamma}(b, g). \quad (4.51)$$

Then, (4.49) follows from (4.9) and (4.51). \square

Proof of Proposition 4.6. Let $k \in \mathbb{N}$. By **H1**, **H2** and the triangle inequality, for any $x, \tilde{x} \in \mathbb{R}^d$,

$$\|\mathbb{E}(x, \tilde{x})\| \leq \tau_\gamma(\|x - \tilde{x}\|) + \gamma c_\infty. \quad (4.52)$$

By using (4.6), and (4.9) we have,

$$\begin{aligned} \|X_{k+1} - \tilde{X}_{k+1}\| &= B_{k+1}\| - \mathbb{E}(X_k, \tilde{X}_k) + 2(\sigma^2\gamma)^{1/2}e(X_k, \tilde{X}_k)e(X_k, \tilde{X}_k)^T Z_{k+1}\| \\ &= B_{k+1}\| - \|\mathbb{E}(X_k, \tilde{X}_k)\|e(X_k, \tilde{X}_k) + 2(\sigma^2\gamma)^{1/2}G_{k+1}e(X_k, \tilde{X}_k)\| \\ &= B_{k+1}\left| \|\mathbb{E}(X_k, \tilde{X}_k)\| - 2(\sigma^2\gamma)^{1/2}G_{k+1} \right| \\ &= \mathcal{H}_\gamma(\|\mathbb{E}(X_k, \tilde{X}_k)\|, G_{k+1}, U_{k+1}). \end{aligned}$$

This gives (4.11) when combined with (4.52). Finally, the last statement follows from Lemma 23 and **H1** ensuring that τ_γ is non-decreasing on $[0, +\infty)$. \square

4.5.1.4 Proof of Proposition 4.7

The proof is an easy consequence of this technical lemma.

Lemma 24. *Assume **H1** and **H2** hold. Then, for any $w \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$, we have that*

$$Q_\gamma \mathcal{V}_1^*(w) = \tau_\gamma(w) + \gamma c_\infty.$$

4.5. Postponed proofs

Proof. For any $w \in \mathbb{R}$, we have,

$$\begin{aligned} Q_\gamma \mathcal{V}_1^*(w) &= \int_{\mathbb{R}} (1 - \bar{p}_{\sigma^2\gamma}(w, g)) (\tau_\gamma(w) + \gamma c_\infty - 2(\sigma^2\gamma)^{1/2}g) \varphi(g) dg \\ &= \int_{\mathbb{R}} (\tau_\gamma(w) + \gamma c_\infty - 2g) \left(\varphi_{(\sigma^2\gamma)^{1/2}}(g) - \varphi_{(\sigma^2\gamma)^{1/2}}(g) \wedge \varphi_{(\sigma^2\gamma)^{1/2}}(\tau_\gamma(w) + \gamma c_\infty - g) \right) dg \\ &= \int_{-\infty}^{(\tau_\gamma(w) + \gamma c_\infty)/2} (\tau_\gamma(w) + \gamma c_\infty - 2g) \left(\varphi_{(\sigma^2\gamma)^{1/2}}(g) - \varphi_{(\sigma^2\gamma)^{1/2}}(\tau_\gamma(w) + \gamma c_\infty - g) \right) dg. \end{aligned}$$

By using change of variable $g \rightarrow a - g$ we have,

$$\begin{aligned} &\int_{-\infty}^{(\tau_\gamma(w) + \gamma c_\infty)/2} (\tau_\gamma(w) + \gamma c_\infty - 2g) \left(\varphi_{(\sigma^2\gamma)^{1/2}}(g) - \varphi_{(\sigma^2\gamma)^{1/2}}(\tau_\gamma(w) + \gamma c_\infty - g) \right) dg \\ &= \frac{1}{2} \int_{\mathbb{R}} (\tau_\gamma(w) + \gamma c_\infty - 2g) \left(\varphi_{(\sigma^2\gamma)^{1/2}}(g) - \varphi_{(\sigma^2\gamma)^{1/2}}(\tau_\gamma(w) + \gamma c_\infty - g) \right) dg \\ &= \tau_\gamma(w) + \gamma c_\infty. \end{aligned}$$

□

Proof of Proposition 4.7. By Lemma 24 and H1, for any $w \in [0, +\infty)$,

$$\begin{aligned} Q_\gamma \mathcal{V}_1^*(w) &= \tau_\gamma(w) + \gamma c_\infty \\ &\leq (1 - \gamma \mathfrak{m}) \mathcal{V}_1^*(w) \mathbb{1}_{(R_1, \infty)}(w) + (1 + \gamma \mathfrak{L}) \mathcal{V}_1^*(w) \mathbb{1}_{(0, R_1]}(w) + \gamma c_\infty. \end{aligned}$$

This completes the proof. □

4.5.1.5 Proof of Proposition 4.8

We first establish that Q_γ admits a unique invariant probability measure μ_γ and is geometrically ergodic. To that end, we show that Q_γ is (a) irreducible and aperiodic, (c) any compact set of $[0, +\infty)$ is small and (d) there exists $\lambda > 0$ and $b \geq 0$ such that $Q_\gamma \mathcal{V}(w) \leq \lambda \mathcal{V}(w) + b$ for any $w \in [0, +\infty)$ with $\mathcal{V}(w) = w + 1$. The proof then follows from [37, Corollary 14.1.6, Theorem 15.2.4].

(a) Let K be a compact set. Then for any $w \in K$ we have

$$Q_\gamma(w, \{0\}) \geq \int_{[-1, 1]} \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g) \varphi(g) dg \geq \eta_K, \quad (4.53)$$

where using H1

$$\begin{aligned} \eta_K &= \inf_{(r, g) \in K \times [-1, 1]} \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g) \int_{[-1, 1]} \varphi(g) dg \\ &\geq \inf_{(a, g) \in [0, M] \times [-1, 1]} \bar{p}_{\sigma^2\gamma}(a, g) \int_{[-1, 1]} \varphi(g) dg, \end{aligned}$$

and $M = (1 + \gamma \mathfrak{L}) \sup(K) + \gamma c_\infty$. Note that since $(a, g) \rightarrow \bar{p}_{\sigma^2\gamma}(a, g)$ is a continuous positive function, and $[0, M] \times [-1, 1]$ is compact, $\eta_K > 0$. Therefore $\{0\}$ is an accessible $(1, \delta_0)$ -small set and Q_γ is irreducible. In addition, $Q_\gamma(0, \{0\}) > 0$ which implies that Q_γ is strongly aperiodic.

- (b) Let now C be a compact set, we show that C is small. By (4.53), for $A \in \mathcal{B}([0, +\infty))$ and $w \in [0, +\infty)$,

$$Q_\gamma^2(w, A) \geq \int_{\mathbb{R}} \mathbb{1}_{\{0\}}(\tilde{w}) Q_\gamma(\tilde{w}, A) Q_\gamma(w, d\tilde{w}) \geq \eta_C Q_\gamma(0, A).$$

Therefore C is a $(2, Q_\gamma(0, \cdot))$ -small set.

- (c) In addition by **H1** and (4.13) we have, for any $w \in [0, +\infty)$,

$$\begin{aligned} & \int_{[0, +\infty)} \mathcal{V}(\tilde{w}) Q_\gamma(w, d\tilde{w}) \\ & \leq 1 + \int_{\mathbb{R}} (1 - \bar{p}_{\sigma^2 \gamma}(\tau_\gamma(w) + \gamma c_\infty, z)) \left(\tau_\gamma(w) + \gamma c_\infty + 2(\sigma^2 \gamma)^{1/2}(g)_- \right) \varphi(g) dg \\ & \leq (1 - \gamma \mathfrak{m}) \mathcal{V}(w) + \gamma R_1(\mathfrak{m} + L) + \gamma c_\infty + 2(\sigma^2 \gamma)^{1/2} / \sqrt{2\pi}. \end{aligned}$$

The proof of the first part of the proposition is complete.

We now establish the second part. Let $A \in \mathcal{B}(\mathbb{R})$ such that $(\delta_0 + \text{Leb})(A) = 0$. Then $0 \notin A$ and $\text{Leb}(A) = 0$ therefore for any $w \in \mathbb{R}$,

$$Q_\gamma(w, A) \leq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbb{1}_A \left(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\gamma^{1/2}g \right) dg = 0.$$

It follows that $\mu_\gamma(A) = \mu_\gamma Q_\gamma(A) = 0$ and $\mu_\gamma \ll (\delta_0 + \text{Leb})$.

Since for any $w \in [0, +\infty)$, $Q_\gamma(w, \{0\}) > 0$, $\delta_{\{0\}}$ is an irreducibility measure, and by [37, Theorem 9.2.15], μ_γ is a maximal irreducibility measure for Q_γ , $\delta_0 \ll \mu_\gamma$ implying that $\mu_\gamma(\{0\}) > 0$.

In the case $c_\infty \neq 0$, by (4.13) and (4.10), for any $A \in \mathcal{B}([0, +\infty))$, $\text{Leb}(A) > 0$, $Q_\gamma(w, A) > 0$ for any $w \in [0, +\infty)$ and therefore Leb is an irreducibility measure. Applying [37, Theorem 9.2.15] again, we get that $\delta_0 + \text{Leb} \ll \mu_\gamma$. This completes the proof since we have already shown that $\mu_\gamma \ll (\delta_0 + \text{Leb})$.

4.5.1.6 Proof of Theorem 4.9

Lemma 25. *Assume **H1** and **H2** hold. For any $w \in \mathbb{R}$,*

$$Q_\gamma(w, \{0\}) = 2\Phi \left(-\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}} \right),$$

where Q_γ is defined by (4.13) and Φ is the cumulative distribution of the one-dimensional Gaussian distribution with mean 0 and variance 1.

Proof. Let $w \in \mathbb{R}$. By (4.12) and the change of variable $g \rightarrow \sigma\sqrt{\gamma}g$, we get

$$\begin{aligned} Q_\gamma(r, \{0\}) &= \int_{\mathbb{R}} \left(1 \wedge \frac{\varphi_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty - \sigma\sqrt{\gamma}g)}{\varphi_{\sigma^2\gamma}(\sigma\sqrt{\gamma}g)} \right) \varphi(g) dg \\ &= \int_{\mathbb{R}} \varphi_{\sigma^2\gamma}(g) \wedge \varphi_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty - g) dg = \int_{\mathbb{R}} \varphi_{\sigma^2\gamma}(g) \wedge \varphi_{\sigma^2\gamma}(g - \tau_\gamma(w) + \gamma c_\infty) dg \\ &= \int_{-\infty}^{(\tau_\gamma(w) + \gamma c_\infty)/2} \varphi_{\sigma^2\gamma}(g - \tau_\gamma(w) + \gamma c_\infty) dg + \int_{(\tau_\gamma(w) + \gamma c_\infty)/2}^{+\infty} \varphi_{\sigma^2\gamma}(g) dg \\ &= \int_{-\infty}^{-(\tau_\gamma(w) + \gamma c_\infty)/2} \varphi_{\sigma^2\gamma}(g) dg + \int_{(\tau_\gamma(w) + \gamma c_\infty)/2}^{+\infty} \varphi_{\sigma^2\gamma}(g) dg = 2\Phi \left(-\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}} \right), \end{aligned}$$

and the lemma follows. \square

Lemma 26. *Let $\sigma^2, \gamma > 0$. For any $t, a \geq 0$, we have*

$$\int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{t - 2\sigma\gamma^{1/2}g}{2a} \right) \right] \bar{p}_{\sigma^2\gamma}(t, g) \varphi(g) dg = 0,$$

where $\bar{p}_{\sigma^2\gamma}$ is defined by (4.10), φ and Φ are the density and the cumulative distribution function of the one-dimensional Gaussian distribution with mean 0 and variance 1 respectively.

Proof. Using the changes of variable $g \mapsto \sigma\gamma^{1/2}g$, $g \mapsto g - t$ and $g \mapsto -g$, we obtain

$$\begin{aligned} &\int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{t - 2\sigma\gamma^{1/2}g}{2a} \right) \right] \bar{p}_{\sigma^2\gamma}(t, g) \varphi(g) dg \\ &= \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{t - 2\sigma\gamma^{1/2}g}{2a} \right) \right] \left\{ 1 \wedge \frac{\varphi_{\sigma^2\gamma}(a - \sigma\sqrt{\gamma}g)}{\varphi_{\sigma^2\gamma}(\sigma\sqrt{\gamma}g)} \right\} \varphi(g) dg \\ &= \int_{-\infty}^{t/2} \left[1 - 2\Phi \left(-\frac{t - 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(t - g) dg + \int_{t/2}^{+\infty} \left[1 - 2\Phi \left(-\frac{t - 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg \\ &= \int_{-\infty}^{-t/2} \left[1 - 2\Phi \left(-\frac{-t - 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg + \int_{t/2}^{+\infty} \left[1 - 2\Phi \left(-\frac{t - 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg \\ &= \int_{t/2}^{+\infty} \left[1 - 2\Phi \left(-\frac{-t + 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg + \int_{t/2}^{+\infty} \left[1 - 2\Phi \left(-\frac{t - 2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg. \end{aligned}$$

Using that $s \in \mathbb{R}$, $1 - 2\Phi(s) = -[1 - 2\Phi(-s)]$ completes the proof. \square

Lemma 27. *Let $\sigma^2, \gamma > 0$. For any $t, s \geq 0$ and $a > 0$,*

$$\begin{aligned} &\int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{t + s - 2\sigma\sqrt{\gamma}g}{2a} \right) \right] \bar{p}_{\sigma^2\gamma}(t, g) \varphi(g) dg \\ &= 2\mathbb{P} \left(\sigma\sqrt{\gamma}G \geq t/2, -s - t \leq 2a\tilde{G} - 2\sigma\sqrt{\gamma}G \leq s - t \right), \end{aligned}$$

where G, \tilde{G} are two independent one-dimensional standard Gaussian random variables, $\bar{p}_{\sigma^2\gamma}$ is defined by (4.10), φ and Φ are the density and the cumulative distribution function of the one-dimensional Gaussian distribution with mean 0 and variance 1 respectively.

Proof. Using the changes of variable $g \mapsto \sigma\gamma^{1/2}g$, $g \mapsto g - t$, we get

$$\begin{aligned} & \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{t+s-2\sigma\sqrt{\gamma}g}{2a} \right) \right] \bar{p}_{\sigma^2\gamma}(t, g) \varphi(g) dg \\ &= \int_{-\infty}^{-t/2} \left[1 - 2\Phi \left(-\frac{-t+s-2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg \\ & \quad + \int_{t/2}^{+\infty} \left[1 - 2\Phi \left(-\frac{t+s-2g}{2a} \right) \right] \varphi_{\sigma^2\gamma}(g) dg \\ &= 2[\mathbb{P}(\sigma\gamma^{1/2}G \geq t/2) - A - B] \end{aligned} \quad (4.54)$$

$$A = \int_{-\infty}^{-t/2} \Phi \left(-\frac{-t+s-2g}{2a} \right) \varphi_{\sigma^2\gamma}(g) dg, \quad B = \int_{t/2}^{+\infty} \Phi \left(-\frac{t+s-2g}{2a} \right) \varphi_{\sigma^2\gamma}(g) dg.$$

In addition, we have since $(-G, -\tilde{G})$ has the same distribution than (G, \tilde{G}) ,

$$\begin{aligned} A &= \mathbb{P} \left(\sigma\sqrt{\gamma}G \leq -\frac{t}{2}, \tilde{G} \leq -\frac{-t+s-2\sigma\sqrt{\gamma}G}{2a} \right) = \mathbb{P} \left(\sigma\sqrt{\gamma}G \geq \frac{t}{2}, \tilde{G} \geq \frac{-t+s+2\sigma\sqrt{\gamma}G}{2a} \right) \\ B &= \mathbb{P} \left(\sigma\sqrt{\gamma}G \geq \frac{t}{2}, \tilde{G} \leq -\frac{t+s-2\sigma\sqrt{\gamma}G}{2a} \right) \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} & A + B \\ &= \mathbb{P}(\sigma\sqrt{\gamma}G \geq t/2) - \mathbb{P} \left(\sigma\sqrt{\gamma}G \geq t/2, -\frac{t+s-2\sigma\sqrt{\gamma}G}{2a} \leq \tilde{G} \leq \frac{-t+s+2\sigma\sqrt{\gamma}G}{2a} \right) \\ &= \mathbb{P}(\sigma\sqrt{\gamma}G \geq t/2) - \mathbb{P} \left(\sigma\sqrt{\gamma}G \geq t/2, -s-t \leq 2a\tilde{G} - 2\sigma\sqrt{\gamma}G \leq s-t \right) \end{aligned}$$

Plugging this expression in (4.54) concludes the proof. \square

Lemma 28. For any $a \in \mathbb{R}$, $b \in [0, 1]$, it holds

$$\Phi(a+b) - \Phi(a-b) \geq 1 - 2\Phi(-b) - a^2b \exp(-b^2/2)/\sqrt{2\pi},$$

where Φ are the density and the cumulative distribution function of the one-dimensional Gaussian distribution with mean 0 and variance 1.

Proof. Define $\psi : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}$ for any $a \in \mathbb{R}$, $b \in [0, 1]$ by

$$\psi(a, b) = \Phi(a+b) - \Phi(a-b) - 1 + 2\Phi(-b) + a^2b \exp(-b^2/2)/\sqrt{2\pi}.$$

We show that $\psi(a, b) \geq 0$ for any $a \in \mathbb{R}$ and $b \in [0, 1]$. Using that $1 - \Phi(-t) = \Phi(t)$ for any $t \in \mathbb{R}$, we get that $\psi(a, b) = \psi(-a, b)$ and therefore we only need to consider

4.5. Postponed proofs

the case $a \leq 0$ and $b \in [0, 1]$. In addition, for any $b \in [0, 1]$, $\psi(0, b) = 0$ and thus, it is sufficient to establish that for any $b \in [0, 1]$, $a \mapsto \psi(a, b)$ is non-increasing on \mathbb{R}_- .

For any $a \leq 0$ and $b \in (0, 1)$, we have using that $\sinh(t) = \int_0^t \cosh(s) ds \leq t \cosh(t)$ and $e^{-t^2/2} \cosh(t) \leq 1$ for any $t \in [0, +\infty)$,

$$\begin{aligned} \sqrt{2\pi} \exp(b^2/2) \frac{\partial \psi}{\partial a}(a, b) &= 2 \exp(-a^2/2) \sinh(-ab) + 2ab \\ &< -2ab[\exp(-a^2b^2/2) \cosh(ab) - 1] \leq 0. \end{aligned}$$

By continuity, it also holds for $a \leq 0$ and $b \in [0, 1]$ which concludes the proof. \square

Lemma 29. *Assume **H1** and **H2** hold. For any $w \in [0, +\infty)$ and $\alpha, \beta \in [0, +\infty)$, such that $\alpha/(2\beta) \leq 1$,*

$$\begin{aligned} \int_{[0, +\infty)} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tilde{w}) + \alpha}{2\beta} \right) \right] Q_\gamma(w, d\tilde{w}) \\ \leq 1 - 2\Phi \left(-\frac{\tau_\gamma(w) + \gamma c_\infty + \alpha/(1 + \gamma L)}{2\sqrt{\sigma^2\gamma + \beta/(1 + \gamma L)^2}} \right) + \zeta \frac{\gamma\alpha}{\beta^3}, \end{aligned}$$

where Φ is the density and the cumulative distribution function of the one-dimensional Gaussian distribution with mean 0 and variance 1,

$$\zeta = (1 + \bar{\gamma}L)^2 \sigma^2 (2\sqrt{2\pi})^{-1} [\sup_{t \geq 0} \{t^2 \Phi(-t)\} + 1/8]. \quad (4.55)$$

Proof. Let $\alpha, \beta \geq 0$ such that $\alpha/(2\beta) \leq 1$. By (4.13), we have

$$\begin{aligned} &\int_{[0, +\infty)} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tilde{w}) + \alpha}{2\beta} \right) \right] Q_\gamma(w, d\tilde{w}) \\ &= \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g)) \varphi(g) dg \\ &\quad + \left[1 - 2\Phi \left(-\frac{\tau_\gamma(0) + \alpha}{2\beta} \right) \right] \int_{\mathbb{R}} \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g) \varphi(g) dg \\ &= \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g)) \varphi(g) dg \\ &\quad + \left[1 - 2\Phi \left(-\frac{\tau_\gamma(0) + \alpha}{2\beta} \right) \right] 2\Phi \left(-\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}} \right). \end{aligned} \quad (4.56)$$

By **H1**, $t \mapsto 1 - 2\Phi(-t)$ is increasing, we have setting $\psi_\gamma(w) = \tau_\gamma(w) + \gamma c_\infty + \alpha/(1 + \gamma L)$,

$$1 - 2\Phi \left(-\frac{\tau_\gamma(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \leq 1 - 2\Phi \left(-\frac{\psi_\gamma(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma L)} \right).$$

Using [44, Lemma 20], Lemma 26 and Lemma 27, we get

$$\begin{aligned}
 & \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\tau_{\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty} - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty}, g)) \varphi(g) dg \\
 & \leq \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\psi_{\gamma}(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma\mathbf{L})} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\psi_{\gamma}(w), g)) \varphi(g) dg \\
 & \quad - \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\psi_{\gamma}(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma\mathbf{L})} \right) \right] \bar{p}_{\sigma^2\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty}, g) \varphi(g) dg \\
 & = 1 - 2\Phi \left(-\frac{\psi_{\gamma}(w)}{2\sqrt{\sigma^2\gamma + \beta/(1 + \gamma\mathbf{L})^2}} \right) - 2\mathbb{P}(\{2\sigma\sqrt{\gamma}G \geq \tau_{\gamma}(w) + \gamma c_{\infty}\} \cap \mathbf{A}), \quad (4.57)
 \end{aligned}$$

where

$$\mathbf{A} = \left\{ -\tau_{\gamma}(w) - \gamma c_{\infty} - \frac{\alpha}{1 + \gamma\mathbf{L}} \leq \frac{2\beta\tilde{G}}{1 + \gamma\mathbf{L}} - 2\sigma\sqrt{\gamma}G \leq -\tau_{\gamma}(w) - \gamma c_{\infty} + \frac{\alpha}{1 + \gamma\mathbf{L}} \right\},$$

and G, \tilde{G} are two independent one-dimensional standard Gaussian random variables.

Define $\theta_{\gamma} : [0, +\infty) \times \mathbb{R} \rightarrow \mathbb{R}$ for any $w \in [0, +\infty)$ and $g \in \mathbb{R}$ by $\theta_{\gamma}(w, g) = (2\beta)^{-1}(1 + \gamma\mathbf{L})[-\tau_{\gamma}(w) - \gamma c_{\infty} + \sigma\sqrt{\gamma}g]$. Then, using that

$$\mathbf{A} = \left\{ \theta_{\gamma}(w, G) - \alpha/(2\beta) \leq \tilde{G} \leq \theta_{\gamma}(w, G) + \alpha/(2\beta) \right\},$$

we have

$$\begin{aligned}
 & \mathbb{P}(\{2\sigma\sqrt{\gamma}G \geq \tau_{\gamma}(w) + \gamma c_{\infty}\} \cap \mathbf{A}) \\
 & = \int_{\mathbb{R}} \mathbb{1}_{[0, +\infty)} \left(g - \frac{\tau_{\gamma}(w) + \gamma c_{\infty}}{2\sigma\sqrt{\gamma}} \right) \left[\Phi \left(\theta_{\gamma}(w, g) + \frac{\alpha}{2\beta} \right) - \Phi \left(\theta_{\gamma}(w, g) - \frac{\alpha}{2\beta} \right) \right] \varphi(g) dg.
 \end{aligned}$$

Since $\alpha/(2\beta) \leq 1$ by Lemma 28 we have, for any $a \in \mathbb{R}$,

$$\Phi \left(a + \frac{\alpha}{2\beta} \right) - \Phi \left(a - \frac{\alpha}{2\beta} \right) \geq 1 - 2\Phi \left(-\frac{\alpha}{2\beta} \right) - \frac{a^2\alpha}{2\sqrt{2\pi}\beta} e^{-a^2/(8\beta^2)},$$

which implies

$$\begin{aligned}
 \mathbb{P}(\{2\sigma\sqrt{\gamma}G \geq \tau_{\gamma}(w) + \gamma c_{\infty}\} \cap \mathbf{A}) & \geq \Phi \left(-\frac{\tau_{\gamma}(w) + \gamma c_{\infty}}{2\sigma\sqrt{\gamma}} \right) \left(1 - 2\Phi \left(-\frac{\alpha}{2\beta} \right) \right) \\
 & \quad - \int_{\mathbb{R}} \mathbb{1}_{[0, +\infty)} \left(g - \frac{\tau_{\gamma}(w) + \gamma c_{\infty}}{2\sigma\sqrt{\gamma}} \right) \frac{\theta_{\gamma}^2(g, w)\alpha}{2\sqrt{2\pi}\beta} e^{-\alpha^2/(8\beta^2)} \varphi(g) dg.
 \end{aligned}$$

4.5. Postponed proofs

Therefore, we obtain using that $\mathbb{E}[\mathbb{1}_{[0,+\infty)}(G)G^2] = 1/2$,

$$\begin{aligned}
& \left(1 - 2\Phi\left(-\frac{\alpha}{2\beta}\right)\right) \Phi\left(-\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}}\right) - \mathbb{P}(\{2\sigma\sqrt{\gamma}G \geq \tau_\gamma(w) + \gamma c_\infty\} \cap A) \\
& \leq \int_{\mathbb{R}} \mathbb{1}_{[0,+\infty)}\left(g - \frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}}\right) \frac{\theta_\gamma^2(g, w)\alpha}{2\sqrt{2\pi}\beta} e^{-\alpha^2/(8\beta^2)} \varphi(g) dg \\
& \leq \frac{\alpha\gamma(1 + \gamma L)^2 \sigma^2}{\beta^3} e^{-\alpha^2/(8\beta^2)} \left[\left(\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}}\right)^2 \Phi\left(-\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}}\right) \right. \\
& \quad \left. - \frac{1}{\sqrt{2\pi}} \frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}} \exp\left(-\left(\frac{\tau_\gamma(w) + \gamma c_\infty}{2\sigma\sqrt{\gamma}}\right)^2 / 2\right) + 1/8 \right].
\end{aligned}$$

Combining this inequality and (4.57) in (4.56) concludes the proof. \square

Under **H1** and **H2**, define $(\alpha_k)_{k \geq 1}$, $(\beta_k)_{k \geq 1}$ for any $k \geq 1$ by

$$\alpha_k = \gamma c_\infty \sum_{i=0}^{k-1} (1 + \gamma L)^{-i}, \quad \beta_k^2 = \gamma \sigma^2 \sum_{i=0}^{k-1} (1 + \gamma L)^{-2i}. \quad (4.58)$$

Lemma 30. *Assume **H1** and **H2** hold. For any $\gamma > 0$, $k \geq 1$, we have*

$$\begin{aligned}
k\gamma c_\infty e^{-k\gamma L} &\leq \alpha_k \leq k\gamma c_\infty, & (k\gamma)^{1/2} \sigma e^{-k\gamma L} &\leq \beta_k \leq (k\gamma)^{1/2} \sigma, \\
[(k\gamma)^{1/2} c_\infty / \sigma] e^{-k\gamma L} &\leq \alpha_k / \beta_k \leq [(k\gamma)^{1/2} c_\infty / \sigma] e^{k\gamma L} \\
[c_\infty \gamma^{1/2} / (\sigma^3 k^{1/2})] e^{-k\gamma L} &\leq \gamma \alpha_k / \beta_k^3 \leq [c_\infty \gamma^{1/2} / (\sigma^3 k^{1/2})] e^{3k\gamma L}
\end{aligned} \quad (4.59)$$

$$\gamma \sum_{i=1}^{k-1} \{\alpha_k / \beta_k^3\} \leq [2c_\infty (k\gamma)^{1/2} / \sigma^3] e^{3k\gamma L}, \quad (4.60)$$

where $(\alpha_k)_{k \geq 1}$, $(\beta_k)_{k \geq 1}$ are defined in (4.58).

Proof. Let $k \geq 1$. Using for any $i \in \mathbb{N}$, $e^{-i\gamma L} \leq (1 + \gamma L)^{-i} \leq 1$, we have

$$k\gamma c_\infty e^{-k\gamma L} \leq \gamma c_\infty \sum_{i=0}^{k-1} (1 + \gamma L)^{-i} \leq k\gamma c_\infty. \quad (4.61)$$

In the same way, using for any $i \in \mathbb{N}$, $e^{-2i\gamma L} \leq (1 + \gamma L)^{-2i} \leq 1$, we obtain

$$k\gamma \sigma^2 e^{-2k\gamma L} \leq \gamma \sigma^2 \sum_{i=0}^{k-1} (1 + \gamma L)^{-2i} \leq k\gamma \sigma^2. \quad (4.62)$$

Combining (4.61) and (4.62) completes the proof of the first four inequalities. Then, (4.60) is a simple consequence of (4.59) and a comparison test. \square

Lemma 31. *Assume **H 1** and **H 2** hold. Let $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$, with the convention $1/0 = +\infty$. For any $\gamma \in (0, \bar{\gamma}]$, $n \in \{0, \dots, n_\gamma\}$, $n_\gamma = \lfloor \bar{\delta}/\gamma \rfloor$, and $w \in [0, +\infty)$, it holds*

$$\int_{[0, +\infty)} \mathbb{1}_{(0, +\infty)}(\tilde{w}) Q_\gamma^{n+1}(w, d\tilde{w}) \leq 1 - 2\Phi\left(-\frac{\tau_\gamma(w) + \alpha_{n+1}}{2\beta_{n+1}}\right) + \zeta \sum_{k=1}^n \frac{\gamma \alpha_k}{\beta_k^3}. \quad (4.63)$$

where $(\alpha_k)_{k \geq 1}, (\beta_k)_{k \geq 1}$ and ζ are defined in (4.58) and (4.55) respectively and Φ is the cumulative distribution of the one-dimensional Gaussian distribution with mean 0 and variance 1.

Proof. Let $\gamma \in (0, \bar{\gamma}]$, $w \in [0, +\infty)$. Note that for any $n \in \{0, \dots, n_\gamma\}$, by (4.58) and Lemma 30,

$$\alpha_n/(2\beta_n) \leq 1. \quad (4.64)$$

Then, by Lemma 25, (4.63) holds for $n = 0$. Assume it holds for $n \in \{0, \dots, n_\gamma - 1\}$. Then, we get

$$\begin{aligned} \int_{[0, +\infty)} \mathbb{1}_{(0, +\infty)}(\tilde{w}) Q_\gamma^{n+1}(w, d\tilde{w}) &\leq \int_{[0, +\infty)} \left[1 - 2\Phi\left(-\frac{\tau_\gamma(w) + \alpha_n}{2\beta_n}\right)\right] Q_\gamma(w, \tilde{w}) \\ &\quad + \zeta \sum_{k=1}^{n-1} \frac{\gamma \alpha_k}{\beta_k^3}. \end{aligned}$$

The proof is then concluded by a straightforward induction using Lemma 29 and (4.64). \square

Theorem 4.23. *Assume **H 1** and **H 2** hold. Let $R \geq 0$. For any $\gamma \in (0, \bar{\gamma}]$ and $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$, $\mu_\gamma((0, R)) \leq \eta_R c_\infty$, where*

$$\eta_R = [\bar{\delta} + \bar{\gamma}]^{1/2} \left[\frac{\zeta e^{3(\bar{\delta} + \bar{\gamma})\mathbf{L}}}{\sigma^3} + \frac{e^{(\bar{\delta} + \bar{\gamma})\mathbf{L}}}{2\sqrt{2\pi}\sigma} \right] / \Phi\left(-\frac{(1 + \bar{\gamma}\mathbf{L})R + (\bar{\delta} + \bar{\gamma})c_\infty}{2\bar{\delta}^{1/2}\sigma e^{-(\bar{\delta} + \bar{\gamma})\mathbf{L}}}\right). \quad (4.65)$$

Proof. Let $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$ and $\gamma \in (0, \bar{\gamma}]$. Set $n_\gamma = \lfloor \bar{\delta}/\gamma \rfloor$. Note that $\bar{\delta} \leq \gamma(n_\gamma + 1) \leq \bar{\delta} + \bar{\gamma}$. By Lemma 31, Proposition 4.8, integrating (4.63) with respect

to μ_γ and using that $\tau_\gamma(0) = 0$, $\Phi(-t) \leq 1/2$ for any $t \geq 0$, gives

$$\begin{aligned}
 \mu_\gamma((0, +\infty)) &\leq \int_{\mathbb{R}} \left\{ 1 - 2\Phi \left(-\frac{\tau_\gamma(w) + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) \right\} d\mu_\gamma(w) + \zeta \sum_{k=1}^{n_\gamma} \frac{\gamma\alpha_k}{\beta_k^3} \quad (4.66) \\
 &\leq 1 - 2\Phi \left(-\frac{\alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) + \zeta \sum_{k=1}^{n_\gamma} \frac{\gamma\alpha_k}{\beta_k^3} \\
 &\quad + 2 \int_{(0, +\infty)} \left\{ \Phi \left(-\frac{\alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) - \Phi \left(-\frac{\tau_\gamma(w) + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) \right\} d\mu_\gamma(w) \\
 &\leq 1 - 2\Phi \left(-\frac{\alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) + \mu_\gamma((0, +\infty)) + \zeta \sum_{k=1}^{n_\gamma} \frac{\gamma\alpha_k}{\beta_k^3} \\
 &\quad - 2 \int_{(0, R_1)} \Phi \left(-\frac{\tau_\gamma(w) + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) d\mu_\gamma(w).
 \end{aligned}$$

Rearranging the terms yields

$$2 \int_{(0, R_1)} \Phi \left(-\frac{\tau_\gamma(w) + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) d\mu_\gamma(w) \leq 1 - 2\Phi \left(-\frac{\alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) + \zeta \sum_{k=1}^{n_\gamma} \frac{\gamma\alpha_k}{\beta_k^3}. \quad (4.67)$$

In addition by **H1** using Lemma 30 and $t \mapsto \Phi(-t)$ is decreasing on \mathbb{R} , we have

$$\begin{aligned}
 \int_{(0, R_1)} \Phi \left(-\frac{\tau_\gamma(w) + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) d\mu_\gamma(w) &\geq \Phi \left(-\frac{(1 + \gamma L)R_1 + \alpha_{n_\gamma+1}}{2\beta_{n_\gamma+1}} \right) \mu_\gamma((0, R_1)) \\
 &\geq \Phi \left(-\frac{(1 + \bar{\gamma}L)R_1 + (\bar{\delta} + \bar{\gamma})c_\infty}{2\bar{\delta}^{1/2}\sigma e^{-(\bar{\delta} + \bar{\gamma})L}} \right) \mu_\gamma((0, R_1)), \quad (4.68)
 \end{aligned}$$

Using that $t \mapsto 1 - 2\Phi(-t)$ is $\sqrt{2/\pi}$ -Lipschitz and combining (4.67), (4.68) and Lemma 30 we get that

$$\begin{aligned}
 \Phi \left(-\frac{(1 + \bar{\gamma}L)R_1 + (\bar{\delta} + \bar{\gamma})c_\infty}{2\bar{\delta}^{1/2}\sigma e^{-(\bar{\delta} + \bar{\gamma})L}} \right) \mu_\gamma((0, R_1)) &\leq \alpha_{n_\gamma+1}/(2\sqrt{2\pi}\beta_{n_\gamma+1}) + \zeta\gamma \sum_{k=1}^{n_\gamma} \{\alpha_k/\beta_k^3\} \\
 &\leq c_\infty[\bar{\delta} + \bar{\gamma}]^{1/2} \left[\frac{\zeta e^{3(\bar{\delta} + \bar{\gamma})L}}{\sigma^3} + \frac{e^{(\bar{\delta} + \bar{\gamma})L}}{2\sqrt{2\pi}\sigma} \right],
 \end{aligned}$$

which implies that $\mu_\gamma((0, R_1)) \leq \eta c_\infty$ and completes the proof. \square

Proof of Theorem 4.9. Let $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$ and $\gamma \in (0, \bar{\gamma}]$. By Proposition 4.7 and using μ_γ is invariant for Q_γ , we obtain

$$\int_{\mathbb{R}} w d\mu_\gamma(w) \leq (1 - \gamma\mathbf{m}) \int_{[R_1, +\infty)} w d\mu_\gamma(w) + (1 + \gamma L) \int_{(0, R_1)} w d\mu_\gamma(w) + \gamma c_\infty$$

Then, rearranging the terms in this inequality yields

$$\begin{aligned} \int_{[R_1, +\infty)} w \, d\mu_\gamma(w) &\leq R_1 \mu_\gamma((0, R_1)) L/m + c_\infty/m \\ \int_{[0, +\infty)} w \, d\mu_\gamma(w) &\leq R_1 \mu_\gamma((0, R_1)) (1 + L/m) + c_\infty/m, \end{aligned}$$

which, combined with Theorem 4.23 applied to $R \leftarrow R_1$, concludes the proof of the first inequality in (4.17). Finally, by (4.66), using that $t \mapsto 1 - 2\Phi(-t)$ is $\sqrt{2/\pi}$ -Lipschitz, we have

$$\begin{aligned} \mu_\gamma((0, +\infty)) &\leq (\sqrt{2\pi}\beta_{n_\gamma+1})^{-1} \int_{[0, +\infty)} \{(1 + \bar{\gamma}L)w + \alpha_{n_\gamma+1}\} d\mu_\gamma(w) + \zeta\gamma \sum_{k=1}^{n_\gamma} \{\alpha_k/\beta_k^3\} \\ &\leq (c_\infty c_1(1 + \bar{\gamma}L) + \alpha_{n_\gamma+1})/(\sqrt{2\pi}\beta_{n_\gamma+1}) + \zeta\gamma \sum_{k=1}^{n_\gamma} \{\alpha_k/\beta_k^3\}. \end{aligned}$$

This finishes the proof using $n_\gamma = \lfloor \bar{\delta}/\gamma \rfloor$, $\bar{\delta} \leq \gamma(n_\gamma + 1) \leq \bar{\delta} + \bar{\gamma}$ and Lemma 30. \square

4.5.1.7 Proof of Theorem 4.10

For any $a > 0$, define $\mathcal{V}_a^* : \mathbb{R} \rightarrow [0, +\infty)$ for any $w \in \mathbb{R}$ by

$$\mathcal{V}_a^*(w) = \mathbb{1}_{(0, +\infty)}(w) \exp(aw). \quad (4.69)$$

Proposition 4.24. *Assume **H1** and **H2** hold. Let $a > 0$. Then, for any $w \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$,*

$$Q_\gamma \mathcal{V}_a^*(w) / \mathcal{V}_a^*(w) \leq \lambda_a^\gamma \mathbb{1}_{[R_a, +\infty)}(w) + \exp\{\gamma a(Lw + 2\sigma^2 a + c_\infty)\} \mathbb{1}_{[0, R_a)}(w), \quad (4.70)$$

where

$$R_a = 1 \vee R_1 \vee [(4a\sigma^2 + 2c_\infty)/m], \quad \lambda_a = \exp(-amR_a/2). \quad (4.71)$$

Proof. Set for any $w \in \mathbb{R}$, $\tau_\gamma^\infty = \tau_\gamma(w) + \gamma c_\infty$. By definition (4.13) and using the change of variable $g \mapsto \sigma\gamma^{1/2}g$ twice, we have

$$\begin{aligned} Q_\gamma \mathcal{V}^*(w) &= \int_{\mathbb{R}} \exp(a\{\tau_\gamma^\infty(w) - 2\sigma\gamma^{1/2}g\}) \{1 - \bar{p}_{\sigma^2\gamma}(\tau_\gamma^\infty(w), g)\} \varphi(g) dg \\ &= \int_{-\infty}^{\tau_\gamma^\infty(w)/2} \exp(a\{\tau_\gamma^\infty(w) - 2g\}) \{\varphi_{\sigma^2\gamma}(g) - \varphi_{\sigma^2\gamma}(\tau_\gamma(w) - g)\} \\ &\leq (2\pi)^{-1/2} \int_{-\infty}^{\sigma\gamma^{1/2}\tau_\gamma^\infty(w)/2} \exp(a\{\tau_\gamma^\infty(w) - 2\sigma\gamma^{1/2}g\} - g^2/2) dg \\ &= \exp(a\tau_\gamma^\infty(w) + 2a^2\sigma^2\gamma) (2\pi)^{-1/2} \int_{-\infty}^{\sigma\gamma^{1/2}\tau_\gamma^\infty(w)/2} \exp(-\{g + 2a\sigma\gamma^{1/2}\}^2/2) dg \\ &\leq \exp(a\tau_\gamma^\infty(w) + 2a^2\sigma^2\gamma) \Phi(\sigma\gamma^{1/2}\tau_\gamma^\infty(w)/2 + 2a\sigma\gamma^{1/2}) \leq e^{a\tau_\gamma^\infty(w) + 2a^2\sigma^2\gamma}. \end{aligned}$$

This concludes the proof of (4.70) for $w \in [0, R_a)$ using **H1**. In the case $w \in [R_a, +\infty)$, we only need to use that by **H1** and definition of R_a (4.71), $a\tau_\gamma^\infty(w) + 2a^2\sigma^2\gamma \leq a(1 - m\gamma)w + ac_\infty + 2a^2\sigma^2\gamma \leq aw - a\gamma mR_a/2$. \square

4.5. Postponed proofs

Proof of Theorem 4.10. Let $\bar{\delta} \in (0, \{\mathbf{L}^{-1} \wedge (\sigma e^{-1}/c_\infty)^2\}]$ and $\gamma \in (0, \bar{\gamma}]$. We show that (4.17) holds with

$$c_3 = (B_a/\lambda_a)^{\bar{\gamma}} a(\mathbf{L}R_a + 2\sigma^2 a + c_\infty) \eta_{R_a} \exp(aR_a) / |\log(\lambda_a)|,$$

where $B_a = \exp\{a(\mathbf{L}R_a + 2\sigma^2 a + c_\infty)\}$, λ_a, R_a are defined in (4.71) and η_R in (4.65).

By Proposition 4.24 and since $\mathcal{V}^*(0) = 0$ and μ_γ is invariant for Q_γ , we have

$$\int_{(0,+\infty)} \mathcal{V}_a^*(w) d\mu_\gamma(w) \leq \lambda_a^\gamma \int_{[R_a,+\infty)} \mathcal{V}_a^*(w) d\mu_\gamma(w) + B_a^\gamma \int_{(0,R_a)} \mathcal{V}_a^*(w) d\mu_\gamma(w),$$

setting $B_a = \exp\{a(\mathbf{L}R_a + 2\sigma^2 a + c_\infty)\}$. Rearranging terms yields

$$\begin{aligned} \int_{(0,+\infty)} \mathcal{V}_a^*(w) d\mu_\gamma(w) &\leq \{B_a^\gamma - \lambda_a^\gamma\} / \{1 - \lambda_a^\gamma\} \int_{(0,R_a)} \mathcal{V}^*(w) d\mu_\gamma(w) \\ &\leq \{B_a^\gamma \lambda_a^{-\gamma} - 1\} / \{\lambda_a^{-\gamma} - 1\} \exp(aR_a) c_\infty \eta_{R_a}, \end{aligned}$$

where we have used Theorem 4.23 applied to $R \leftarrow R_a$ in the last inequality. The proof is then completed upon using that for any $t \geq 0$, $t \leq e^t - 1 \leq te^t$. \square

4.5.1.8 Proof of Theorem 4.11

Lemma 32. *Assume H1 and H2 hold. For any $w \in [0, +\infty)$ and $\alpha, \beta \in [0, +\infty)$, $\beta > 0$,*

$$\int_{(0,+\infty)} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tilde{w}) + \alpha}{2\beta} \right) \right] Q_\gamma(w, d\tilde{w}) \leq 1 - 2\Phi \left(-\frac{\tau_\gamma(w) + \gamma c_\infty + \alpha/(1 + \gamma\mathbf{L})}{2\sqrt{\sigma^2\gamma + \beta/(1 + \gamma\mathbf{L})^2}} \right),$$

where Φ are the density and the cumulative distribution function of the one-dimensional Gaussian distribution with mean 0 and variance 1.

Proof. Let $\alpha, \beta \geq 0$, $\beta > 0$. By (4.13), we have

$$\begin{aligned} &\int_{(0,+\infty)} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tilde{w}) + \alpha}{2\beta} \right) \right] Q_\gamma(w, d\tilde{w}) \\ &= \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\tau_\gamma(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\tau_\gamma(w) + \gamma c_\infty, g)) \varphi(g) dg. \end{aligned}$$

By H1, $t \mapsto 1 - 2\Phi(-t)$ is increasing, we have setting $\psi_\gamma(w) = \tau_\gamma(w) + \gamma c_\infty + \alpha/(1 + \gamma\mathbf{L})$,

$$1 - 2\Phi \left(-\frac{\tau_\gamma(\tau_\gamma(w) + \gamma c_\infty - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \leq 1 - 2\Phi \left(-\frac{\psi_\gamma(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma\mathbf{L})} \right).$$

Using [44, Lemma 20], Lemma 26 and Lemma 27, we get

$$\begin{aligned}
 & \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\tau_{\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty} - 2\sigma\sqrt{\gamma}g) + \alpha}{2\beta} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty}, g)) \varphi(g) dg \\
 & \leq \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\psi_{\gamma}(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma L)} \right) \right] (1 - \bar{p}_{\sigma^2\gamma}(\psi_{\gamma}(w), g)) \varphi(g) dg \\
 & \quad - \int_{\mathbb{R}} \left[1 - 2\Phi \left(-\frac{\psi_{\gamma}(w) - 2\sigma\sqrt{\gamma}g}{2\beta/(1 + \gamma L)} \right) \right] \bar{p}_{\sigma^2\gamma}(\tau_{\gamma}(w) + \gamma c_{\infty}, g) \varphi(g) dg \\
 & \leq 1 - 2\Phi \left(-\frac{\psi_{\gamma}(w)}{2\sqrt{\sigma^2\gamma + \beta/(1 + \gamma L)^2}} \right),
 \end{aligned}$$

which completes the proof. \square

We consider in what follows that $(W_k)_{k \in \mathbb{N}}$ is the canonical process on $([0, +\infty)^{\mathbb{N}}, \mathcal{B}([0, +\infty))^{\otimes \mathbb{N}})$ and for any $w \in [0, +\infty)$, \mathbb{P}_w and \mathbb{E}_w correspond to the probability and expectation respectively, associated with Q_{γ} and the initial condition δ_w on this space.

Lemma 33. *Assume H1 and H2 hold. For any $k \in \mathbb{N}$ and $w \in (0, +\infty)$,*

$$\mathbb{P}_w \left[\min_{i \in \{0, \dots, k+1\}} W_i > 0 \right] \leq 1 - 2\Phi \left[-\frac{\tau_{\gamma}(w) + \alpha_{k+1}}{2\beta_{k+1}} \right],$$

where $\alpha_{k+1}, \beta_{k+1}$ are given in (4.58).

Proof. The proof is by induction on $k \in \mathbb{N}$. The proof for $k = 0$ follows from Lemma 25. Assume that the result holds for $k - 1 \in \mathbb{N}$ and for any $w \in (0, +\infty)$. Then, by the Markov property and the assumption hypothesis, for any $w \in (0, +\infty)$,

$$\begin{aligned}
 \mathbb{P}_w \left[\min_{i \in \{0, \dots, k\}} W_i > 0 \right] &= \mathbb{E}_w \left[\mathbb{1}_{(0, +\infty)}(W_1) \mathbb{P}_{W_1} \left[\min_{i \in \{0, \dots, k-1\}} W_i > 0 \right] \right] \\
 &\leq \mathbb{E}_x \left[\mathbb{1}_{(0, +\infty)}(W_1) \left\{ 1 - 2\Phi \left(-\frac{\tau_{\gamma}(W_1) + \alpha_k}{2\beta_k} \right) \right\} \right].
 \end{aligned}$$

The proof is then completed upon using Lemma 32. \square

Lemma 34. *Assume H1 and H2 hold. Then, for any $k \in \mathbb{N}$, $w, \tilde{w} \in [0, +\infty)$,*

$$\|\delta_w Q_{\gamma}^{k+1} - \delta_{\tilde{w}} Q_{\gamma}^{k+1}\|_{\text{TV}} \leq 1 - 2\Phi \left[-\frac{\tau_{\gamma}(w \vee \tilde{w}) + \alpha_{k+1}}{2\beta_{k+1}} \right],$$

where $\alpha_{k+1}, \beta_{k+1}$ are given in (4.58).

4.5. Postponed proofs

Proof. We consider again $(G_k)_{k \geq 1}$ and $(U_k)_{k \geq 1}$ two independent sequences of i.i.d. standard Gaussian and $[0, 1]$ -uniform random variables respectively. Define the Markov chains $(W_k)_{k \in \mathbb{N}}$ and $(\tilde{W}_k)_{k \in \mathbb{N}}$ starting from $w \in [0, +\infty)$ and $\tilde{w} \in [0, +\infty)$ respectively, for any $k \in \mathbb{N}$, $W_{k+1} = \mathcal{G}_\gamma(W_k, G_{k+1}, U_{k+1})$ and $\tilde{W}_{k+1} = \mathcal{G}_\gamma(\tilde{W}_k, G_{k+1}, U_{k+1})$. Note that the case $w = \tilde{w}$ is trivial so we only consider the converse and assume that $w < \tilde{w}$, $\tilde{w} > 0$. Then, we obtain by Proposition 4.6 that almost surely $W_k \leq \tilde{W}_k$ for any $k \in \mathbb{N}$, which implies that

$$\|\delta_w Q_\gamma^{k+1} - \delta_{\tilde{w}} Q_\gamma^{k+1}\|_{\text{TV}} \leq \mathbb{P}(W_{k+1} \neq \tilde{W}_{k+1}) \leq \mathbb{P}\left(\min_{i \in \{0, \dots, k+1\}} \tilde{W}_i > 0\right).$$

Indeed, if $\min_{i \in \{0, \dots, k+1\}} \tilde{W}_i = 0$, then there exists $i \in \{0, \dots, k+1\}$, $\tilde{W}_i = 0$ which implies since $\tilde{W}_i \geq W_i \geq 0$ that $\tilde{W}_i = W_i$ and therefore $\tilde{W}_{k+1} = W_{k+1}$ by definition of the two processes. The proof is then completed by Lemma 33. \square

Proposition 4.25. *Assume **H1** and **H2** hold. Let $t > 0$. Then, for any $w, \tilde{w} \in [0, +\infty)$,*

$$\|\delta_w Q_\gamma^{\lceil t_0/\gamma \rceil} - \delta_{\tilde{w}} Q_\gamma^{\lceil t_0/\gamma \rceil}\|_{\text{TV}} \leq 1 - 2\Phi\left[-L^{1/2} \frac{w \vee \tilde{w} + t_0 c_\infty}{2\sigma\{1 - e^{-2L(t_0 + \bar{\gamma})}\}^{1/2}}\right].$$

Proof. Note that by (4.58) for any $k \in \mathbb{N}$, $\alpha_{k+1} \leq k\gamma c_\infty$ and $\beta_{k+1}^2 = (\sigma^2(1 + \gamma L)/L)\{1 - (1 + \gamma L)^{-2k}\} \geq (\sigma^2(1 + \gamma L)/L)\{1 - e^{-2k\gamma L}\}$ using that $(1+t) \leq e^t$ for any $t \in \mathbb{R}$. The proof is then completed using Lemma 34, **H1** and the previous bounds for $k \leftarrow \lceil t_0/\gamma \rceil$. \square

Lemma 35. *Assume **H1** and **H2** hold. Let $a > 0$. Then, for any $w \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$,*

$$Q_\gamma \mathcal{V}_1^*(w) \leq (1 - \gamma \mathfrak{m}) \mathcal{V}_1^*(w) + \gamma \{\mathfrak{L} + \mathfrak{m}\} R_1 + \gamma c_\infty, \quad (4.72)$$

$$Q_\gamma \mathcal{V}_a^*(w) \leq \lambda_a^\gamma \mathcal{V}_a^*(w) + \gamma \alpha_a \mathbb{1}_{[0, R_1]}(w), \quad (4.73)$$

where Q_γ is defined by (4.13), for any $w \in \mathbb{R}$, $\mathcal{V}_1^*(w) = |w|$, \mathcal{V}_a^* by (4.69), λ_a, R_a are defined by (4.71) and

$$\alpha_a = a \exp\{a R_a + \bar{\gamma} a (L R_a + 2\sigma^2 a + c_\infty)\} [L R_a + 2\sigma^2 a + c_\infty]. \quad (4.74)$$

Proof. (4.72) is a simple consequence of Proposition 4.7. Note that for $w \in \mathbb{R}$, $w \geq R_a$, (4.73) holds by Proposition 4.24. In the case $w < R_a$, using Proposition 4.24 and $e^t - 1 \leq te^t$ for any $t \geq 0$, we obtain $Q_\gamma \mathcal{V}^*(w) \leq \lambda_a^\gamma \mathcal{V}^*(w) + \mathcal{V}^*(w) [\lambda_a^{-\gamma} \exp\{\gamma a (Lw + 2\sigma^2 a + c_\infty)\} - 1] \leq \lambda_a^\gamma \mathcal{V}^*(w) + a \mathcal{V}^*(w) \exp\{\gamma a (Lw + 2\sigma^2 a + c_\infty)\} [Lw + 2\sigma^2 a + c_\infty]$, which completes the proof. \square

Define $\mathcal{V}_1, \mathcal{V}_a : \mathbb{R} \rightarrow [0, +\infty)$ for any $w \in \mathbb{R}$ by

$$\mathcal{V}_1(w) = 1 + \Lambda_1 |w|, \quad \mathcal{V}_a(w) = 1 + \Lambda_a \mathbb{1}_{(0, +\infty)}(w) \exp(aw). \quad (4.75)$$

Proposition 4.26. *Assume **H1** and **H2** hold. Let $a > 0$ and $t_0 > 0$. Then, for any $w \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$,*

$$Q_\gamma^{\lceil t_0/\gamma \rceil} \mathcal{V}_1(w) \leq \lambda_1 \mathcal{V}_1^*(w) + \beta_1, \quad Q_\gamma^{\lceil t_0/\gamma \rceil} \mathcal{V}_a(w) \leq \lambda_a \mathcal{V}_a^*(w) + \beta_a,$$

where Q_γ is defined by (4.13), $\mathcal{V}_1, \mathcal{V}_a$ by (4.75), λ_a by (4.71), α_a by (4.74) and

$$\lambda_1 = e^{-m}, \quad \beta_1 = (t_0 + \bar{\gamma})[\{L + m\}R_1 + c_\infty] + 1, \quad \beta_a = (t_0 + \bar{\gamma})\alpha_a + 1.$$

Proof of Theorem 4.11. By Lemma 35 and an easy induction, and using that $1 - t \leq e^{-t}$ for any $t \geq 0$, we have for any $k \in \mathbb{N}$,

$$Q_\gamma^k \mathcal{V}_1^*(w) \leq \lambda_1^{k\gamma} \mathcal{V}_1^*(w) + k\gamma[\{L + m\}R_1 c_\infty], \quad Q_\gamma^k \mathcal{V}_a^*(w) \leq \lambda_a^{k\gamma} \mathcal{V}_a^*(w) + k\gamma\alpha_a.$$

Using that $Q_\gamma^k 1 \equiv 1$ for any $k \in \mathbb{N}$ completes the proof. \square

Proof of Theorem 4.11. Let $t_0 > 0$. We only show the result for $\mathcal{V} = \mathcal{V}_1$. The result for $\mathcal{V} = \mathcal{V}_a$, $a > 0$ is similar upon replacing λ_1 and β_1 by λ_a and β_a given in Proposition 4.26 respectively.

Define $\delta_1 = 4\beta_1/(1 - \lambda_1) - 1$ and $M_1 = \sup\{w \in [0, +\infty) : \mathcal{V}_1(w) \leq \delta_1\}$ which is well defined since $\lim_{w \rightarrow +\infty} \mathcal{V}_1(w) = +\infty$. Define in addition,

$$\varepsilon_1 = 2\Phi \left[-L^{1/2} \frac{M_1 + t_0 c_\infty}{2\sigma \{1 - e^{-2L(t_0 + \bar{\gamma})}\}^{1/2}} \right] < 1.$$

Then, $\{\mathcal{V}_1 \leq \delta_1\}$ is a $(1, \varepsilon)$ -Doebelin set for Q_γ and $\lambda_1 + 2\beta_1/(1 + \delta_1) < 1$. Therefore, [37, Theorem 19.4.1] implies that for any $k \in \mathbb{N}$,

$$\left\| \delta_w Q_\gamma^{\lceil t_0/\gamma \rceil k} - \mu_\gamma \right\|^V \leq C \rho^k \{\mathcal{V}_1(w) + \mu_\gamma(\mathcal{V}_1)\},$$

where a bound on $\mu_\gamma(\mathcal{V}_1)$ is provided by Theorem 4.9 and

$$\begin{aligned} \log(\rho) &= \log(1 - \varepsilon_1) \log(\bar{\lambda}_1) / \{\log(1 - \varepsilon_1) + \log(\bar{\lambda}_1) - \log(\bar{\beta}_1)\} \\ \bar{\lambda}_1 &= \lambda_1 + 2\beta_1/(1 + \delta_1), \quad \bar{\beta}_1 = \lambda_1 \beta_1 + \delta_1 \\ C &= \{\lambda_1 + 1\} / [1 + \bar{\beta}_1 / \{(1 - \varepsilon_1)(1 - \bar{\lambda}_1)\}]. \end{aligned}$$

\square

4.5.2 Postponed proofs of Section 4.3

4.5.2.1 Proof of Proposition 4.14

The proof is an easy consequence of Lemma 38 below and the definition of $(\mathbf{W}_t^{(n)})_{t \geq 0}$. Before stating and proving Lemma 38, we need the following technical results.

Lemma 36. *Assume **A 1**. Then, for any $q \in [1, +\infty)$, we have for any $\gamma \in (0, \bar{\gamma}]$,*

$$\mathbb{E}[|W_1 - \tau_\gamma(W_0) - \gamma c_\infty|^q] \leq (4\sigma^2\gamma)^{q/2} \{ \mathbf{m}_q + 2 \sup_{u \geq 0} [u^q \Phi(-u)] \},$$

where W_1 is defined by (4.12), \mathbf{m}_q is the q -th moment of the standard Gaussian distribution and Φ is its cumulative distribution function.

Proof. Let $w_0 \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$. By definition (4.13) and (4.10), we have setting $\bar{\tau}_\gamma^\infty(w_0) = \{\tau_\gamma(w_0) + \gamma c_\infty\}/(2\sqrt{\sigma^2\gamma})$,

$$\begin{aligned} \int_{\mathbb{R}_+} |w_1 - \tau_\gamma(w_0) - \gamma c_\infty|^q Q_\gamma(w_0, dw_1) &= (4\sigma^2\gamma)^{q/2} \int_{-\infty}^{\bar{\tau}_\gamma^\infty(w_0)} |g|^q \varphi(g) dg \\ &\quad - (4\sigma^2\gamma)^{q/2} \int_{\bar{\tau}_\gamma^\infty(w_0)}^{+\infty} |g|^q \varphi(g) dg + \int_{\mathbb{R}} |\tau_\gamma(w_0) + \gamma c_\infty|^q \varphi(2\bar{\tau}_\gamma^\infty(w_0) - g) \wedge \varphi(g) dg \\ &\leq (4\sigma^2\gamma)^{q/2} \mathbf{m}_q + 2^{p+1} \sigma^p \gamma^{p/2} [\bar{\tau}_\gamma^\infty(w_0)]^q \Phi\{-\bar{\tau}_\gamma^\infty(w_0)\}, \end{aligned}$$

which completes the proof. \square

Lemma 37. *Assume **A 1**. Then, there exists $C \geq 0$ such that for any $k \in \mathbb{N}$ and $\gamma \in (0, \bar{\gamma}]$, $\mathbb{E}[W_k^4] \leq e^{Ck\gamma} \{\mathbb{E}[W_0^4] + 1\}$, where $(W_k)_{k \in \mathbb{N}}$ is defined by (4.12).*

Proof. By **A 1**, (4.12) and (4.13) we have that for any $w_0 \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$, setting $W_0 = w_0$ and $\kappa_\infty(w_0) = \kappa(w_0) + c_\infty$,

$$\begin{aligned} \int_{\mathbb{R}_+} w_1^4 Q_\gamma(w_0, dw_1) &= \mathbb{E}[W_1^4] \leq \mathbb{E}[(w_0 + \gamma \kappa_\infty(w_0) - 2\sqrt{\sigma^2\gamma} G_1)^4] \\ &= \{w_0 + \gamma \kappa_\infty(w_0)\}^4 + 6\sigma^2\gamma \{w_0 + \gamma \kappa_\infty(w_0)\}^2 + 3\sigma^4\gamma^2. \end{aligned}$$

By **A 1**, for any $\ell \in \{2, 4\}$, we have that for any $w_0 \in [0, +\infty)$, $\gamma \in (0, \bar{\gamma}]$,

$$\{w_0 + \gamma \kappa_\infty(w_0)\}^\ell \leq w_0^\ell + 2^{\ell-1} \ell \gamma (1 + L_\kappa)^\ell [|w_0|^\ell + c_\infty].$$

Therefore, we obtain that there exists some constant $C_1, C_2 \geq 0$, such that for any $w_0 \in [0, +\infty)$, $\gamma \in (0, \bar{\gamma}]$,

$$\int_{\mathbb{R}_+} w_1^4 Q_\gamma(w_0, dw_1) \leq w_0^4 + C_1 \gamma \{1 + w_0^2 + w_0^4\} \leq (1 + \gamma C_2) w_0^4 + \gamma C_2.$$

By an easy induction, we get then that for any $w_0 \in [0, +\infty)$, $\gamma \in (0, \bar{\gamma}]$ and $k \in \mathbb{N}$,

$$\int_{\mathbb{R}_+} w_1^4 Q_\gamma^k(w_0, dw_1) \leq (1 + C_2\gamma)^k w_0^4 + C_2\gamma \sum_{i=0}^{k-1} (1 + C_2\gamma)^i \leq e^{k\gamma C_2} [w_0^4 + C_2],$$

which completes the proof by the Markov property. \square

Lemma 38. *Assume **A 1**. Then, there exists $C \geq 0$ such that for any $k \in \mathbb{N}$ and $\gamma \in (0, \bar{\gamma}]$, $\mathbb{E}[\max_{\ell \in \{0, \dots, k\}} [W_\ell - W_0]^4] \leq C(k\gamma)^2 e^{Ck\gamma} \{\mathbb{E}[W_0^4] + 1\}$, where $(W_k)_{k \in \mathbb{N}}$ is defined by (4.12).*

Proof. Assume that $\mathbb{E}[W_0^4] < +\infty$, otherwise the results holds. Denote by $(\mathcal{F}_k)_{k \in \mathbb{N}}$ the filtration associated with $(W_k)_{k \in \mathbb{N}}$. We consider the following decomposition for any $\ell \in \mathbb{N}$,

$$W_\ell - W_0 = A_\ell + B_\ell, \quad A_\ell = \sum_{i=0}^{\ell-1} \Delta M_i, \quad B_\ell = \sum_{i=0}^{\ell-1} H_i,$$

where using that $\mathbb{E}[W_{i+1} | \mathcal{F}_i] = \tau_\gamma^\infty(W_i)$ by Lemma 24 and the Markov property,

$$\Delta M_i = W_{i+1} - \mathbb{E}[W_{i+1} | \mathcal{F}_i] = W_{i+1} - \tau_\gamma^\infty(W_i), \quad H_i = \tau_\gamma^\infty(W_i) - W_i. \quad (4.76)$$

Then, using Young's inequality, we get for any $\gamma \in (0, \bar{\gamma})$ and $k \in \mathbb{N}$,

$$\max_{\ell \in \{0, \dots, k\}} [W_\ell - W_0]^4 \leq 2^3 \{ \max_{\ell \in \{0, \dots, k\}} A_\ell^4 + \max_{\ell \in \{0, \dots, k\}} B_\ell^4 \}. \quad (4.77)$$

We now bound the two last terms in the right hand side of this equation. First, by A1 and Young's inequality, we get for any $\gamma \in (0, \bar{\gamma}]$ and $k \in \mathbb{N}$,

$$\mathbb{E}[\max_{\ell \in \{0, \dots, k\}} B_\ell^4] \leq k^3 \sum_{i=0}^{k-1} H_i^4 \leq 2^3 (k\gamma)^4 (1 + L_\kappa)^4 \{ \max_{i \in \{0, \dots, k-1\}} \mathbb{E}[W_i^4] + c_\infty^4 \}. \quad (4.78)$$

In addition, by definition (4.76), $(\Delta M_i)_{i \in \mathbb{N}}$ are $(\mathcal{F}_i)_{i \in \mathbb{N}}$ -martingale increments. It follows by Burkholder inequality [22, Theorem 3.2] and Young's inequality that there exists $C_4 \geq 0$ satisfying for any $k \in \mathbb{N}$ and $\gamma \in (0, \bar{\gamma}]$,

$$\mathbb{E}[\max_{\ell \in \{0, \dots, k\}} A_\ell^4] \leq C_4 \mathbb{E}[\{\sum_{i=0}^{k-1} \Delta M_i^2\}^2] \leq C_4 k \sum_{i=0}^{k-1} \mathbb{E}[\Delta M_i^4].$$

Therefore by Lemma 36, we get that

$$\mathbb{E}[\max_{\ell \in \{0, \dots, k\}} A_\ell^4] \leq C_4 \mathbb{E}[\{\sum_{i=0}^{k-1} \Delta M_i^2\}^2] \leq C_4 (4\sigma^2 k\gamma)^2 \{ \mathbf{m}_4 + 2 \sup_{u \geq 0} [u^4 \Phi(-u)] \},$$

where \mathbf{m}_4 is the fourth moment of the standard Gaussian distribution. Combining this result with (4.78) and using Lemma 37 in (4.77) concludes the proof. \square

4.5.2.2 Proof of Proposition 4.15

To show this result, we use the Komolgorov criteria [73, Corollary 14.9]: for any $T \geq 0$, there exist $C_T \geq 0$ such that for any $n \in \mathbb{N}$ and $s, t \in [0, +\infty)$, $s \leq t$,

$$\mathbb{E} \left[\left| \mathbf{W}_t^{(n)} - \mathbf{W}_s^{(n)} \right|^4 \right] \leq C_T (t - s)^2.$$

Note that denoting $k_1^{(n)} = \lceil s/\gamma_n \rceil$ and $k_2^{(n)} = \lfloor t/\gamma_n \rfloor$, we have by (4.19)

$$\begin{aligned} & \mathbb{E} \left[\left| \mathbf{W}_t^{(n)} - \mathbf{W}_s^{(n)} \right|^4 \right] \\ & \leq \begin{cases} (t - s) \gamma_n^{-1} \{W_{k_1^{(n)}+1} - W_{k_1^{(n)}}\}^4 & \text{if } k_1^{(n)} < k_2^{(n)} \\ 3^3 [\{W_{k_2^{(n)}+1} - W_{k_1^{(n)}}\}^4 + \{W_{k_2^{(n)}} - W_{k_1^{(n)}}\}^4 + \{W_{k_1^{(n)}} - W_{k_1^{(n)}-1}\}^4] & \text{otherwise.} \end{cases} \end{aligned}$$

Lemma 38, Lemma 37 and the Markov property complete the proof.

4.5.2.3 Proof of Proposition 4.17

We preface the proof by the following technical lemma.

Lemma 39. *Assume A1. Then, for any $q \in [1, +\infty)$, we have*

(a) *for any $\gamma \in (0, \bar{\gamma}]$ and $w_0 \in [0, +\infty)$,*

$$\begin{aligned} & -2(4\sigma^2\gamma)^{q/2} \int_{\bar{\tau}_\gamma^\infty(w_0)}^{+\infty} |g|^q \varphi(g) dg \\ & \leq \int_{\mathbb{R}_+} |w_1 - \tau_\gamma(w_0) - \gamma c_\infty|^q Q_\gamma(w_0, dw_1) - (4\sigma^2\gamma)^{q/2} \mathbf{m}_q \leq 0, \end{aligned} \quad (4.79)$$

where $\bar{\tau}_\gamma^\infty(w_0) = \{\tau_\gamma(w_0) + \gamma c_\infty\}/(2\sqrt{\sigma^2\gamma})$, Q_γ is defined by (4.13), \mathbf{m}_q is the q -th moment of the standard Gaussian distribution. and φ is its probability density function;

(b) *for any $\gamma \in (0, \bar{\gamma}]$,*

$$\left| \int_{\mathbb{R}_+} |w_1 - \gamma c_\infty|^q Q_\gamma(0, dw_1) \right| \leq 3(\gamma c_\infty)^q.$$

Proof. (a) Let $w_0 \in [0, +\infty)$ and $\gamma \in (0, \bar{\gamma}]$. By definition (4.13) and (4.10), we have setting $\bar{\tau}_\gamma^\infty(w_0) = \{\tau_\gamma(w_0) + \gamma c_\infty\}/(2\sqrt{\sigma^2\gamma})$,

$$\begin{aligned} & \int_{\mathbb{R}_+} |w_1 - \tau_\gamma(w_0) - \gamma c_\infty|^q Q_\gamma(w_0, dw_1) = (4\sigma^2\gamma)^{q/2} \int_{-\infty}^{\bar{\tau}_\gamma^\infty(w_0)} |g|^q \varphi(g) dg \\ & - (4\sigma^2\gamma)^{q/2} \int_{\bar{\tau}_\gamma^\infty(w_0)}^{+\infty} |g|^q \varphi(g) dg + \int_{\mathbb{R}} |\tau_\gamma(w_0) + \gamma c_\infty|^q \varphi(2\bar{\tau}_\gamma^\infty(w_0) - g) \wedge \varphi(g) dg. \end{aligned} \quad (4.80)$$

Therefore, we obtain that

$$\begin{aligned} & \int_{\mathbb{R}_+} |w_1 - \tau_\gamma(w_0) - \gamma c_\infty|^q Q_\gamma(w_0, dw_1) - (4\sigma^2\gamma)^{q/2} \mathbf{m}_q \\ & = -2(4\sigma^2\gamma)^{q/2} \int_{\bar{\tau}_\gamma^\infty(w_0)}^{+\infty} |g|^q \varphi(g) dg + 2(4\sigma^2\gamma)^{q/2} [\{\bar{\tau}_\gamma^\infty(w_0)\}^q \Phi(-\bar{\tau}_\gamma^\infty(w_0))]. \end{aligned}$$

Using that $\int_{\bar{\tau}_\gamma^\infty(w_0)}^{+\infty} |g|^q \varphi(g) dg \geq \{\bar{\tau}_\gamma^\infty(w_0)\}^q \Phi(-\bar{\tau}_\gamma^\infty(w_0))$ completes the proof.

(b) By (4.80) and since $\tau_\gamma(0) = 0$ and $\bar{\tau}_\gamma^\infty(0) = \gamma c_\infty/(2\sqrt{\sigma^2\gamma})$,

$$\left| \int_{\mathbb{R}_+} |w_1 - \tau_\gamma(0) - \gamma c_\infty|^q Q_\gamma(0, dw_1) \right| \leq 2(4\sigma^2\gamma)^{q/2} \int_0^{\frac{\gamma c_\infty}{2\sqrt{\sigma^2\gamma}}} |g| \varphi(g) dg + (\gamma c_\infty)^q.$$

Using that $(4\sigma^2\gamma)^{q/2} \int_0^{\gamma c_\infty/(2\sqrt{\sigma^2\gamma})} |g| \varphi(g) dg \leq (\gamma c_\infty)^q$ completes the proof. \square

Proof of Proposition 4.17. Proof of (4.24). Let $\varphi \in C^\infty(\mathbb{R}^d)$ satisfying (4.23), $N \in \mathbb{N}$, $(t_1, \dots, t_N, s, t) \in [0, +\infty)^{N+2}$, $0 \leq t_1 \leq \dots \leq t_N \leq s < t$, $\psi : \mathbb{R}_+^N \rightarrow \mathbb{R}$, continuous and bounded. Note that we only need to show that

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left| \mathbb{E} \left[\varphi(\mathbf{W}_t^{(n)}) - \varphi(\mathbf{W}_s^{(n)}) - \int_s^t \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du \middle| \mathcal{G}_s^{(n)} \right] \right| \right],$$

setting for any $n \in \mathbb{N}$, $u \in [0, +\infty)$, $\mathcal{G}_u^{(n)} = \mathcal{F}_{\lceil u/\gamma_n \rceil}^{(n)}$, where $(\mathcal{F}_k^{(n)})_{k \in \mathbb{N}}$ is the filtration corresponding to $(W_k^{(n)})_{k \in \mathbb{N}}$.

Define $k_1^{(n)} = \lceil t/\gamma_n \rceil$ and $k_2^{(n)} = \lceil s/\gamma_n \rceil$ and consider the following decomposition

$$\mathbb{E} \left[\varphi(\mathbf{W}_t^{(n)}) - \varphi(\mathbf{W}_s^{(n)}) - \int_s^t \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du \middle| \mathcal{G}_s^{(n)} \right] = \mathbb{E} \left[A_1^{(n)} + A_2^{(n)} + A_3^{(n)} \middle| \mathcal{G}_s^{(n)} \right] \quad (4.81)$$

$$A_1^{(n)} = \varphi(\mathbf{W}_t^{(n)}) - \varphi(W_{k_1^{(n)}}^{(n)}) - \{\varphi(\mathbf{W}_s^{(n)}) - \varphi(W_{k_2^{(n)}}^{(n)})\}$$

$$A_2^{(n)} = - \int_s^t \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du + \gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathcal{A}\varphi(W_k^{(n)})$$

$$A_3^{(n)} = \varphi(W_{k_1^{(n)}}^{(n)}) - \varphi(W_{k_2^{(n)}}^{(n)}) - \gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathcal{A}\varphi(W_k^{(n)}).$$

We deal with these three terms separately.

First since φ satisfies (4.23), by the fundamental theorem of calculus, there exists $C \geq 0$ such that for any $w_0, w_1 \in \mathbb{R}$, $|\varphi(w_1) - \varphi(w_0)| \leq C \max(|w_0|, |w_1|) |w_0 - w_1|$. By (4.19), we get that there exists $C \geq 0$ such that for any $n \in \mathbb{N}$, almost surely,

$$|A_1^{(n)}| \leq C\gamma_n \{ \max_{i \in \{k_1^{(n)}, k_1^{(n)}+1, k_2^{(n)}, k_2^{(n)}+1\}} |W_i^{(n)}|^2 + 1 \}.$$

This implies by Lemma 36 and the Lebesgue convergence theorem that

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|A_1^{(n)}|] = 0. \quad (4.82)$$

Regarding $A_2^{(n)}$, we consider the decomposition,

$$A_2^{(n)} = A_{2,1}^{(n)} + A_{2,2}^{(n)},$$

$$A_{2,1}^{(n)} = \int_s^{k_2^{(n)}\gamma_n} \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du + \int_t^{k_1^{(n)}\gamma_n} \mathcal{A}\varphi(\mathbf{W}_u^{(n)}) du$$

$$A_{2,2}^{(n)} = - \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \int_{k\gamma_n}^{(k+1)\gamma_n} \{\mathcal{A}\varphi(\mathbf{W}_u^{(n)}) - \mathcal{A}\varphi(W_k^{(n)})\} du.$$

Since φ satisfies (4.23), by Lemma 36 and the Lebesgue dominated convergence theorem, we get that $\lim_{n \rightarrow +\infty} \mathbb{E}[|A_{2,1}^{(n)}|] = 0$. In addition, we have by definition of \mathcal{A} (4.22) that for any $n \in \mathbb{N}$, $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$, $u \in (k\gamma_n, (k+1)\gamma_n)$,

$$|\mathcal{A}\varphi(\mathbf{W}_u^{(n)}) - \mathcal{A}\varphi(W_k^{(n)})| \leq B_{u,k} + 2\mathbb{1}_{A_k^c} \sigma^2 \sup_{\mathbb{R}} |\varphi''|$$

$$B_{u,k} = (|\kappa(W_k^{(n)})| + c_\infty) |\varphi'(\mathbf{W}_u^{(n)}) - \varphi'(W_k^{(n)})| \\ + |\varphi'(\mathbf{W}_u^{(n)})| |\kappa(\mathbf{W}_u^{(n)}) - \kappa(W_k^{(n)})| + 2\mathbb{1}_{A_k} \sigma^2 |\varphi''(\mathbf{W}_u^{(n)}) - \varphi''(W_k^{(n)})|,$$

4.5. Postponed proofs

where $A_k = \{W_k^{(n)} = 0, W_{k+1}^{(n)} \neq 0\} \cup \{W_{k+1}^{(n)} = 0, W_k^{(n)} \neq 0\}$. Note that using that φ' and φ'' are Lipschitz and $\sup_{\tilde{w} \in [0, +\infty)} |\varphi'|(\tilde{w})/(1 + |\tilde{w}|) < +\infty$ by (4.23), **A 1** and (4.19), we get that there exists $C \geq 0$ such that for any $n \in \mathbb{N}$, $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$, $u \in (k\gamma_n, (k+1)\gamma_n)$,

$$|B_{u,k}^{(n)}| \leq C\gamma_n \{|W_k^{(n)}|^2 + |W_{k+1}^{(n)}|^2 + 1\},$$

which implies by Lemma 36 that

$$\lim_{n \rightarrow +\infty} \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \int_{k\gamma_n}^{(k+1)\gamma_n} \mathbb{E}[|B_{u,k}^{(n)}|] du = 0.$$

To conclude that $\lim_{n \rightarrow +\infty} \mathbb{E}[|A_2^{(n)}|] = 0$, it remains to show that

$$\lim_{n \rightarrow +\infty} \gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathbb{E}[\mathbb{1}_{A_k^c}] = 0. \quad (4.83)$$

Note that using that by definition, $(W_k^{(n)})_{k \in \mathbb{N}}$ is a Markov chain with Markov kernel Q_{γ_n} (4.13), the Markov property implies that for any $n \in \mathbb{N}$ and $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$,

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{A_k^c}] &\leq \mathbb{P}(W_k^{(n)} = 0, W_{k+1}^{(n)} \neq 0) + \mathbb{P}(W_k^{(n)} \neq 0, W_{k+1}^{(n)} = 0) \\ &= 1 - 2\Phi[-c_\infty \sqrt{\gamma_n}/(2\sigma)] + 2\mathbb{E}[\mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)}) \Phi[-\tau_{\gamma_n}^\infty(W_k^{(n)})/\{2(\sigma^2 \gamma_n)^{1/2}\}]], \end{aligned} \quad (4.84)$$

where $\tau_{\gamma_n}^\infty(w) = \tau_{\gamma_n}(w) + \gamma c_\infty$, for any $w \in [0, +\infty)$. Since $1 - 2\Phi(-u) \leq u$ for any $u \in [0, +\infty)$, we get that there exists $C \geq 0$ such that for any $n \in \mathbb{N}$, $\gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} 1 - 2\Phi[-c_\infty \sqrt{\gamma_n}/(2\sigma)] \leq C((t-s) + \gamma_n)\gamma_n^{1/2}$ and therefore

$$\lim_{n \rightarrow +\infty} \gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \{1 - 2\Phi[-c_\infty \sqrt{\gamma_n}/(2\sigma)]\} = 0. \quad (4.85)$$

Regarding the second term in (4.84), consider the sequence of measurable functions defined for any $n \in \mathbb{N}$, $\omega \in \Omega$, $k \in \mathbb{N}$ by

$$f_n(\omega, k) = \gamma_n \mathbb{1}_{\{k_2^{(n)}, \dots, k_1^{(n)} - 1\}}(k) \mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)}) \Phi[-\tau_{\gamma_n}^\infty(W_k^{(n)})/\{2(\sigma^2 \gamma_n)^{1/2}\}],$$

on the measure space $(\Omega \times \mathbb{N}, \mathcal{F} \otimes 2^{\mathbb{N}}, \mathbb{P} \otimes \nu_c)$, where $2^{\mathbb{N}}$ is the power set of \mathbb{N} and ν_c is the counting measure on \mathbb{N} . Note that $\mathbb{P} \otimes \nu_c$ almost everywhere, $\lim_{n \rightarrow +\infty} f_n(\omega, k) = 0$ and in addition, $\sum_{k \in \mathbb{N}} \int_\Omega \tilde{f}_n(\omega, k) d\mathbb{P}(\omega) \leq (t-s) + \gamma_n$. Therefore by the Lebesgue dominated

convergence theorem, we obtain that $\lim_{n \rightarrow +\infty} \sum_{k \in \mathbb{N}} \int_{\Omega} \tilde{f}_n(\omega, k) d\mathbb{P}(\omega) = 0$ which implies by definition,

$$\lim_{n \rightarrow +\infty} 2\gamma_n \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathbb{E} \left[\mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)}) \Phi[-\tau_{\gamma_n}^{\infty}(W_k^{(n)}) / \{2(\sigma^2 \gamma_n)^{1/2}\}] \right] = 0.$$

This result combined with (4.85) and (4.84) in (4.83) shows that

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[|A_2^{(n)}| \right] = 0. \quad (4.86)$$

Finally we deal with $A_3^{(n)}$ from the decomposition

$$A_3^{(n)} = \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \varphi(W_{k+1}^{(n)}) - \varphi(W_k^{(n)}) - \gamma_n \mathcal{A}\varphi(W_k^{(n)}).$$

Set for any $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$, $\Delta W_k^{(n)} = W_{k+1}^{(n)} - W_k^{(n)}$. Using that φ is three times continuously differentiable, we get by Taylor's theorem with Lagrange remainder, that for any $n \in \mathbb{N}$, $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$, there exists $u_k \in [0, 1]$ satisfying

$$\begin{aligned} \varphi(W_{k+1}^{(n)}) - \varphi(W_k^{(n)}) - \gamma_n \mathcal{A}\varphi(W_k^{(n)}) &= \varphi'(W_k^{(n)}) \Delta W_k^{(n)} \\ &+ (\varphi''(W_k^{(n)})/2) \{\Delta W_k^{(n)}\}^2 + 6^{-1} \varphi^{(3)}(u_k W_{k+1}^{(n)} + (1 - u_k) W_k^{(n)}) \{\Delta W_k^{(n)}\}^3. \end{aligned}$$

It follows from the definition (4.12), **A 1** and Young's inequality, setting $\tau_{\gamma_n}^{\infty}(w) = \tau_{\gamma_n}(w) + c_{\infty}$ and $\kappa^{\infty}(w) = \kappa(w) + c_{\infty}$ that for any $n \in \mathbb{N}$, $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$,

$$|\Delta W_k^{(n)}|^3 \leq 4\{\gamma_n^3 |\kappa^{\infty}(W_k^{(n)})|^3 + |W_{k+1}^{(n)} - \tau_{\gamma_n}^{\infty}(W_k^{(n)})|^3\}.$$

It follows then using the definition of \mathcal{A} (4.22), (4.23), Lemma 38 and Lemma 36 that

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|A_3^{(n)} - A_4^{(n)}|] = 0, \quad \text{where } A_4^{(n)} = A_{4,1}^{(n)} + A_{4,2}^{(n)}, \quad (4.87)$$

$$\begin{aligned} A_{4,1}^{(n)} &= \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} [\varphi'(W_k^{(n)}) \Delta W_{k+1}^{(n)} - \gamma_n \varphi'(W_k^{(n)}) \kappa^{\infty}(W_k^{(n)})] \\ A_{4,2}^{(n)} &= \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} [(\varphi''(W_k^{(n)})/2) \{\Delta W_{k+1}^{(n)}\}^2 - 2\gamma_n \sigma^2 \varphi''(W_k^{(n)}) \mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)})]. \end{aligned}$$

Note that by Lemma 24 and the Markov property, we have that for any $n \in \mathbb{N}$ and $k \in \{k_2^{(n)}, \dots, k_1^{(n)} - 1\}$, $\varphi'(W_k^{(n)}) \mathbb{E}[\Delta W_{k+1}^{(n)} | \mathcal{F}_k^{(n)}] - \gamma_n \varphi'(W_k^{(n)}) \kappa^{\infty}(W_k^{(n)}) = 0$, which implies that

$$\begin{aligned} \mathbb{E}[A_{4,1}^{(n)} | \mathcal{G}_s^{(n)}] &= \mathbb{E}[A_{4,1}^{(n)} | \mathcal{F}_{k_2^{(n)}}^{(n)}] \\ &= \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathbb{E}[\varphi'(W_k^{(n)}) \mathbb{E}[\Delta W_{k+1}^{(n)} | \mathcal{F}_k^{(n)}] - \gamma_n \varphi'(W_k^{(n)}) \kappa^{\infty}(W_k^{(n)}) | \mathcal{F}_{k_2^{(n)}}^{(n)}] = 0. \end{aligned} \quad (4.88)$$

4.5. Postponed proofs

We now show that $\lim_{n \rightarrow +\infty} \mathbb{E}[|\mathbb{E}[A_{4,2}^{(n)} | \mathcal{G}_s^{(n)}]|] = 0$ using the decomposition

$$\begin{aligned} A_{4,2}^{(n)} &= A_{4,2,1}^{(n)} + A_{4,2,2}^{(n)}, \\ A_{4,2,1}^{(n)} &= \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)}) (\varphi''(W_k^{(n)})/2) [\{\Delta W_{k+1}^{(n)}\}^2 - 4\sigma^2 \gamma_n] \\ A_{4,2,2}^{(n)} &= \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \mathbb{1}_{\{0\}}(W_k^{(n)}) (\varphi''(0)/2) \{W_{k+1}^{(n)}\}^2. \end{aligned} \quad (4.89)$$

Then, by Lemma 39-(a), we have using the Markov property that

$$|\mathbb{E}[A_{4,2,1}^{(n)} | \mathcal{G}_s^{(n)}]| = |\mathbb{E}[A_{4,2,1}^{(n)} | \mathcal{F}_{k_2^{(n)}}^{(n)}]| \leq 8\sigma^2 \sup_{\mathbb{R}} \{|\varphi''|\} \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} \tilde{f}_n(\omega, k), \quad (4.90)$$

$$\text{where } \tilde{f}_n(\omega, k) = \gamma_n \mathbb{1}_{\{k_2^{(n)}, \dots, k_1^{(n)}-1\}}(k) \mathbb{1}_{\mathbb{R}_+^*}(W_k^{(n)}(\omega)) \Upsilon(\bar{\tau}_\gamma^\infty(W_k^{(n)}(\omega))),$$

$\Upsilon(u) = \int_u^{+\infty} |g|^q \varphi(g) dg$, $\bar{\tau}_\gamma^\infty(W_k^{(n)}(\omega)) = \tau_{\gamma_n}^\infty(W_k^{(n)}(\omega)) / \{2(\sigma^2 \gamma_n)^{1/2}\}$. $(\tilde{f}_n)_{n \in \mathbb{N}}$ is a sequence of measurable functions defined $(\Omega \times \mathbb{N}, \mathcal{F} \otimes 2^{\mathbb{N}}, \mathbb{P} \otimes \nu_c)$. Note that $\mathbb{P} \otimes \nu_c$ almost everywhere, $\lim_{n \rightarrow +\infty} \tilde{f}_n(\omega, k) = 0$ and in addition,

$$\sum_{k \in \mathbb{N}} \int_{\Omega} \tilde{f}_n(\omega, k) d\mathbb{P}(\omega) \leq [(t-s) + \gamma_n] \int_{\mathbb{R}} |g|^q \varphi(g) dg.$$

Therefore, we obtain that $\lim_{n \rightarrow +\infty} \sum_{k \in \mathbb{N}} \int_{\Omega} \tilde{f}_n(\omega, k) d\mathbb{P}(\omega) = 0$ by the Lebesgue dominated convergence theorem, which implies by (4.90) that

$$\lim_{n \rightarrow +\infty} \mathbb{E}[|\mathbb{E}[A_{4,2,1}^{(n)} | \mathcal{G}_s^{(n)}]|] = 0. \quad (4.91)$$

We now consider $A_{4,2,2}$. By Lemma 39-(b) and the Markov property, it follows that

$$|\mathbb{E}[A_{4,2,2}^{(n)} | \mathcal{G}_s^{(n)}]| = |\mathbb{E}[A_{4,2,2}^{(n)} | \mathcal{F}_{k_2^{(n)}}^{(n)}]| \leq (3\varphi''(0)/2) [(t-s) + \gamma_n] \gamma_n c_\infty^2,$$

showing that $\lim_{n \rightarrow +\infty} \mathbb{E}[|\mathbb{E}[A_{4,2,2}^{(n)} | \mathcal{G}_s^{(n)}]|] = 0$. Combining this result, (4.91)-(4.89)-(4.88)-(4.87), we get that $\lim_{n \rightarrow +\infty} \mathbb{E}[|\mathbb{E}[A_3^{(n)} | \mathcal{G}_s^{(n)}]|] = 0$. Plugging this result and (4.86)-(4.82) in (4.81) completes the proof.

Proof of (4.25). The proof follows exactly the same lines as (4.24) but we use that the only different and non negligible term is $A_{4,2}^{(n)}$ which becomes

$$A_{4,2}^{(n)} = \sum_{k=k_2^{(n)}}^{k_1^{(n)}-1} [(\varphi''(W_k^{(n)})/2) [\{\Delta W_{k+1}^{(n)}\}^2 - 4\gamma_n \sigma^2]].$$

Using (4.79) in Lemma 39, the assumption that $\varphi''(w) \geq 0$ for any $w \in \mathbb{R}$, and the Markov property, we get that for any $n \in \mathbb{N}$, $\mathbb{E}[A_{4,2}^{(n)} | \mathcal{G}_s] \leq 0$, which concludes the proof. \square

4.5.2.4 Proof of Theorem 4.16

Proposition 4.27. *Assume A1. Let μ_∞ be a limit point of $(\mu_n)_{n \in \mathbb{N}}$. Then, μ_∞ -almost everywhere, $\inf_{t \in [0, +\infty)} \mathbb{W}_t \geq 0$.*

Proof. Without loss of generality, we assume that $(\mu_n)_{n \in \mathbb{N}}$ converges to μ_∞ . Since $\omega \mapsto \inf_{t \in [0, +\infty)} \omega_t$ is continuous, $F = \{\omega \in \mathbb{W} : \inf_{t \in [0, +\infty)} \omega_t \geq 0\}$ is closed. Therefore, by the Portmanteau theorem [74, Theorem 13.16], we obtain that $\mu_\infty(F) \geq \limsup_{n \rightarrow +\infty} \mu_n(F) = 1$. \square

Proof of Theorem 4.16. Recall that we denote by $(\mu_n)_{n \in \mathbb{N}}$ the sequence of distribution on \mathbb{W} associated with $\{(\mathbf{W}_t^{(n)})_{t \geq 0} : n \in \mathbb{N}\}$. Let μ_∞ be a limit point of this sequence for the convergence in distribution. Without loss of generality, we assume that $(\mu_n)_{n \in \mathbb{N}}$ converges in distribution to μ_∞ . Note that by Proposition 4.14, for any continuous function $F : \mathbb{W} \rightarrow \mathbb{R}$ such that $|F|(\omega) \leq C_T \{1 + \sup_{t \in [0, T]} |\omega_t|^{\delta_c}\}$ for $\delta_c \in [0, 4)$, $T, C_T \geq 0$, then F is uniformly integrable for $(\mu_n)_{n \in \mathbb{N}}$ and therefore (see e.g. [3, Lemma 5.1.7.])

$$\lim_{n \rightarrow +\infty} \int_{\mathbb{W}} F d\mu_n = \int_{\mathbb{W}} F d\mu_\infty. \quad (4.92)$$

We divide then the proof into two parts. **First part:** we first show that under μ_∞ , $(M_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale. Considering

$$F_1 : \omega \mapsto \left(\varphi_1(\omega_t) - \varphi_1(\omega_s) - \int_s^t \mathcal{A}\varphi_1(\omega_u) du \right) \psi(\omega_{t_1}, \dots, \omega_{t_N}),$$

and applying Proposition 4.17-(4.24) to $\varphi_1(w) = w$ for any $w \in \mathbb{R}$, since $\mathcal{A}\varphi_1$ is continuous under A1, for any $N \in \mathbb{N}$, $(t_1, \dots, t_N, s, t) \in [0, +\infty)^{N+2}$, $0 \leq t_1 \leq \dots \leq t_N \leq s < t$, $\psi : \mathbb{R}_+^N \rightarrow \mathbb{R}$, continuous and bounded,

$$\mathbb{E}^{\mu_\infty} [(M_t - M_s) \psi(\mathbf{W}_{t_1}, \dots, \mathbf{W}_{t_N})] = 0,$$

where $\mathbb{E}^{\mu_\infty} [\cdot]$ is the expectation under μ_∞ on $(\mathbb{W}, \mathcal{W})$. We obtain by the monotone class theorem and [115, Theorem 2.3, Chapter 0] that the first part of the result holds, *i.e.* $(M_t)_{t \geq 0}$ defined by (4.21) is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale on $(\mathbb{W}, \mathcal{W}, (\mathcal{W}_t)_{t \geq 0}, \mu_\infty)$.

Second part: It remains to show that under μ_∞ , $(N_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale. We first establish setting $\varphi_2(w) = w^2$ for $w \in \mathbb{R}$, that

$$\tilde{N}_t = \varphi_2(\mathbf{W}_t) - \varphi_2(\mathbf{W}_0) - \int_0^t \mathcal{A}\varphi_2(\mathbf{W}_u) du,$$

is a $(\mathcal{W}_t)_{t \geq 0}$ -submartingale, which easily implies that $(N_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -submartingale. Let $N \in \mathbb{N}$, $(t_1, \dots, t_N, s, t) \in [0, +\infty)^{N+2}$, $0 \leq t_1 \leq \dots \leq t_N \leq s < t$, $\psi : \mathbb{R}_+^N \rightarrow \mathbb{R}$, continuous, nonnegative and bounded. Then, consider $F_2^+ = F_{2,1}^+ + F_{2,2}^+$ on \mathbb{W} with :

$$F_{2,1}^+ : \omega \mapsto \left\{ \varphi_2(\omega_t) - \varphi_2(\omega_s) - 2 \int_s^t \omega_u (\kappa(\omega_u) + c_\infty) du \right\} \psi(\omega_{t_1}, \dots, \omega_{t_N})$$

$$F_{2,2}^+ : \omega \mapsto -4\sigma^2 \left\{ \int_s^t \mathbb{1}_{\mathbb{R}_+^*}(\omega_u) du \right\} \psi(\omega_{t_1}, \dots, \omega_{t_N})$$

Note that it is easy to check that $F_{2,1}^+$ is continuous and $F_{2,2}^+$ is bounded lower semi-continuous on \mathbb{W} , *i.e.* for any $(\omega^n)_{n \in \mathbb{N}}$ converging to ω^∞ in \mathbb{W} endowed with the uniform convergence on compact set, $\liminf_{n \rightarrow +\infty} F_{2,2}^+(\omega^n) \geq F_{2,2}^+(\omega^\infty)$. Therefore, we obtain by the Portmanteau theorem [74, Theorem 13.16] and (4.92) that

$$\int_{\mathbb{W}} F_{2,1}^+ d\mu = \lim_{n \rightarrow \infty} \int_{\mathbb{W}} F_{2,1}^+ d\mu_n, \text{ and } \int_{\mathbb{W}} F_{2,2}^+ d\mu \leq \liminf_{n \rightarrow +\infty} \int_{\mathbb{W}} F_{2,2}^+ d\mu_n.$$

Therefore, Proposition 4.17-(4.24) applied with $\varphi \leftarrow \varphi_2$ implies that

$$0 = \limsup_{n \rightarrow +\infty} \int_{\mathbb{W}} F_2^+ d\mu_n \leq \int_{\mathbb{W}} F_2^+ d\mu.$$

Using the same arguments as before, we obtain that under μ_∞ , $(\tilde{N}_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -submartingale. Then, it is easy to verify that $(N_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -submartingale. We complete then the proof by showing that $(N_t)_{t \geq 0}$ is also a $(\mathcal{W}_t)_{t \geq 0}$ -supermartingale under μ_∞ . To do so, we need the following lemma.

Lemma 40. *Assume A1. Then, for any limit point μ_∞ of $(\mu_n)_{n \in \mathbb{N}}$, μ_∞ -almost everywhere, $t \mapsto \langle M \rangle_t - 4\sigma^4 t$ is nonincreasing, where $(\langle M \rangle_t)_{t \geq 0}$ is the quadratic variation of $(M_t)_{t \geq 0}$.*

Proof. Let $N \in \mathbb{N}$, $(t_1, \dots, t_N, s, t) \in [0, +\infty)^{N+2}$, $0 \leq t_1 \leq \dots \leq t_N \leq s < t$, $\psi : \mathbb{R}_+^N \rightarrow \mathbb{R}$, continuous, nonnegative and bounded. Consider now the continuous map

$$F_2^- : \omega \mapsto \left\{ \varphi_2(\omega_t) - \varphi_2(\omega_0) - \int_0^t \tilde{\mathcal{A}}\varphi_2(\omega_u) du \right\} \psi(\omega_{t_1}, \dots, \omega_{t_N}).$$

Then, by (4.92) and Proposition 4.17-(4.25), we get $\lim_{n \rightarrow \infty} \int_{\mathbb{W}} F_2^- d\mu_n \leq 0$. Using that under μ_∞ $(M_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale, we get that $(M_t^2 - 4\sigma^2 t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ supermartingale. By the Doob-Meyer decomposition [73, Theorem 22.5], under μ_∞ , there exists a unique nondecreasing, locally integrable and predictable process $(C_t)_{t \geq 0}$, such that $(M_t^2 - 4\sigma^2 t + C_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale. In addition, under μ_∞ , by [115, Theorem 1.8, Chapter IV], the quadratic variation $(\langle M \rangle_t)_{t \geq 0}$ of $(M_t)_{t \geq 0}$ is a finite variation process satisfying $(M_t^2 - \langle M \rangle_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale therefore $(M_t^2 - 4\sigma^2 t - (\langle M \rangle_t - 4\sigma^2 t))_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale. Therefore, $(\langle M \rangle_t - 4\sigma^2 t - C_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale and a finite variation process. By [115, Proposition 1.2, Chapter IV], μ_∞ -almost everywhere, for any $t \in [0, +\infty)$, $\langle M \rangle_t - 4\sigma^2 t + C_t = 0$, which completes the proof. \square

By Lemma 40, denoting by $(\langle M \rangle_t)_{t \geq 0}$, the quadratic variation of $(M_t)_{t \geq 0}$, see [115, Theorem 1.8, Chapter IV], μ_∞ -almost everywhere, $t \mapsto \langle M \rangle_t - 4\sigma^2 t$ is nondecreasing and therefore we get that for any $s, t \in [0, +\infty)$, $s \leq t$, μ_∞ -almost everywhere,

$$\int_s^t \mathbb{1}_{\mathbb{R}_+^*}(W_u) d\langle M \rangle_u \leq 4\sigma^2 \int_s^t \mathbb{1}_{\mathbb{R}_+^*}(W_u) du. \quad (4.93)$$

In addition, by the occupation times formula [115, Corollary 1.6, Chapter VI] applied twice and Proposition 4.27, μ_∞ -almost everywhere,

$$\langle M \rangle_t = \int_0^t d \langle M \rangle_u = \int_0^t \mathbb{1}_{\mathbb{R}_+}(W_u) d \langle W \rangle_u = \int_{\mathbb{R}_+} L_t^a da = \int_{\mathbb{R}_+^*} L_t^a da = \int_0^t \mathbb{1}_{\mathbb{R}_+^*}(W_u) d \langle M \rangle_u$$

Using this result and (4.93), we get that $\langle M \rangle_t - \langle M \rangle_s \leq 4\sigma^2 \int_s^t \mathbb{1}_{\mathbb{R}_+^*}(W_u) du$, for any $s, t \in [0, +\infty)$, $s \leq t$. Therefore since $(M_t^2 - \langle M \rangle_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -martingale under μ_∞ , we conclude that $(N_t)_{t \geq 0}$ is a $(\mathcal{W}_t)_{t \geq 0}$ -supermartingale which completes the proof. \square

4.5.3 Postponed proofs of Section 4.4

Proof of Proposition 4.19. From **AO 3** we know that $z_\theta(t) \in \mathbb{K}$ for all $t \in [0, T]$. In particular, using that for all $0 \leq s \leq t \leq T$, $z_\theta(t) - z_\theta(s) = \int_s^t F_\theta(z_\theta(u), u) du$, we get that

$$\|z_\theta(t) - z_\theta(s)\| \leq \mathbf{C}_F(t - s). \quad (4.94)$$

Let $k \in \mathbb{N}$ be such that $\tilde{z}_\theta^h(kh) \in \mathbb{K}$ (this is for instance the case of $k = 0$). Then, for $t \in [kh, (k+1)h]$, using by (4.32) that

$$z_\theta(t) - \tilde{z}_\theta^h(t) = z_\theta(kh) - \tilde{z}_\theta^h(kh) + \int_{kh}^t \left(F_\theta(z_\theta(s), s) - F_\theta(\tilde{z}_\theta^h(kh), kh) \right) ds,$$

and setting $f(t) = \|z_\theta(t) - \tilde{z}_\theta^h(t)\|$, we get by (4.94) and **AO 3** for any $h > 0$ and $t \in [kh, (k+1)h]$,

$$f(t) \leq (1 + L'_F h) f(kh) + L'_F (1 + \mathbf{C}_F) \frac{h^2}{2}.$$

Assuming that $h \leq \bar{h}$ where \bar{h} is sufficiently small so that

$$\frac{1}{2} L'_F (1 + \mathbf{C}_F) \bar{h} T e^{L'_F T} < \delta,$$

we get by a direct induction that, for all $k \in \{0, \dots, \lfloor T/h \rfloor\}$, $\tilde{z}_\theta^h(kh) \in \mathbb{K}$ and for all $t \in [kh, (k+1)h]$, $t \in [0, T]$,

$$f(t) \leq \frac{1}{2} L'_F (1 + \mathbf{C}_F) h^2 (k+1) e^{L'_F h(k+1)}$$

Conclusion follows with

$$\mathbf{C} = \frac{1}{2} L'_F (1 + \mathbf{C}_F) N T e^{L'_F T}.$$

\square

Proof of Proposition 4.21. We have by **AO 2** and (4.37), $T_\gamma(\theta) = \theta - \gamma \nabla \nabla U(\theta)$ with

$$\nabla U(\theta) = \nabla U_0(\theta) + \sum_{i=1}^N \mathbf{s}_i(z_\theta(t_i)) \quad (4.95)$$

where z_θ solves (4.30) and \mathbf{s}_i are defined in (4.37). For all $\theta, \tilde{\theta} \in \mathbb{R}^d$ and $t \in [0, T]$,

$$z_\theta(t) - z_{\tilde{\theta}}(t) = \int_0^t (F_\theta(z_\theta(s), s) - F_{\tilde{\theta}}(z_{\tilde{\theta}}(s), s)) ds$$

and thus, using (4.33), Grönwall's inequality implies that

$$\|z_\theta(t) - z_{\tilde{\theta}}(t)\| \leq \mathbf{L}_F \|\theta - \tilde{\theta}\| t e^{\mathbf{L}'_F t} \quad \forall t \geq 0.$$

In particular, by (4.95) and (4.37),

$$\|\nabla U(\theta) - \nabla U(\tilde{\theta})\| \leq \left(\mathbf{L}_U + \mathbf{L}_s \mathbf{L}_F \sum_{i=1}^N t_i e^{\mathbf{L}'_F t_i} \right) \|\theta - \tilde{\theta}\|.$$

Moreover, similarly, we get if $\|\theta - \tilde{\theta}\| \geq R_U$,

$$\langle \theta - \tilde{\theta}, \nabla U(\theta) - \nabla U(\tilde{\theta}) \rangle \geq m_U \|\theta - \tilde{\theta}\|^2 - N \mathbf{C}_s \|\theta - \tilde{\theta}\|.$$

Combining the last two estimates yields **H1**. Finally, **H2** follows using **AO4**, (4.35) and (4.37) from

$$\|\nabla U(\theta) - \tilde{b}_h(\theta)\| \leq \mathbf{L}_s \sum_{i=1}^N \|z_\theta(t_i) - \Psi_i^h(\theta)\| \leq \mathbf{C}_\Psi \mathbf{L}_s h^\alpha.$$

□

Bibliography

- [1] Sungjin Ahn, Anoop Korattikara et al. “Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring”. In: *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. Ed. by John Langford and Joelle Pineau. ICML '12. Edinburgh, Scotland, GB: Omnipress, July 2012, pp. 1591–1598.
- [2] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. 2nd. Oxford University Press, Inc., 2017.
- [3] L. Ambrosio, N. Gigli et al. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2008.
- [4] Ioan Andricioaei, John E Straub et al. “Smart darting Monte Carlo”. In: *The Journal of Chemical Physics* 114.16 (2001), pp. 6994–7000.
- [5] C. Andrieu and S. Livingstone. “Peskun–Tierney ordering for Markov chain and process Monte Carlo: beyond the reversible scenario”. In: *Accepted for publication in Ann. Statist.* (2019).
- [6] Nawaf B.-R. and M. C. Holmes-Cerfon. “Sticky Brownian motion and its numerical solution”. In: *SIAM Review* 62.1 (2020), pp. 164–195.
- [7] D. Bakry, I. Gentil et al. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552.
- [8] V. Bally and L. Caramellino. “Positivity and lower bounds for the density of Wiener functionals”. In: *Potential Anal.* 39.2 (2013), pp. 141–168.
- [9] V. Bally and D. Talay. “The law of the Euler scheme for stochastic differential equations. I. Convergence rate of the distribution function”. In: *Probability Theory and Related Fields* 104.1 (1996), pp. 43–60.
- [10] David Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012.
- [11] M. Barkhagen, N. H. Chau et al. “On stochastic gradient Langevin dynamics with dependent data streams in the logconcave case”. In: *Bernoulli* 27.1 (2021), pp. 1–33.

-
- [12] M. Betancourt, S. Byrne et al. “The geometric foundations of Hamiltonian Monte Carlo”. English. In: *Bernoulli* 23.4A (NOV 2017), 2257–2298.
- [13] Patrick Billingsley. “Convergence of probability measures”. In: (1999).
- [14] Luke Bornn, Pierre E Jacob et al. “An adaptive interacting Wang–Landau algorithm for automatic density exploration”. In: *Journal of Computational and Graphical Statistics* 22.3 (2013), pp. 749–773.
- [15] Valentin de Bortoli, Alain Durmus et al. “Efficient stochastic optimisation by unadjusted Langevin Monte Carlo. Application to maximum marginal likelihood and empirical Bayesian estimation”. In: *Statistics and Computing* 31 (2021), p. 29.
- [16] N. Bou-Rabee and M. Hairer. “Nonasymptotic mixing of the MALA algorithm”. In: *IMA J. Numer. Anal.* 33 (2013), pp. 80–110.
- [17] N. Bou-Rabee and S.-S. Jesús María. “Geometric Integrators and the Hamiltonian Monte Carlo method”. In: *Acta Numerica* 27 (2018), pp. 113–206.
- [18] N. Bou-Rabee and H. Owhadi. “Long-run accuracy of variational integrators in the stochastic context”. In: *SIAM J. Numer. Anal.* 48 (2010), pp. 278–297.
- [19] S. Boucheron, G. Lugosi et al. *Concentration Inequalities*. Oxford University Press, Oxford, 2013, pp. x+481.
- [20] Anthony E Brockwell and Joseph B Kadane. “Identification of regeneration times in MCMC simulation, with application to adaptive schemes”. In: *Journal of Computational and Graphical Statistics* 14.2 (2005), pp. 436–458.
- [21] R. Bubley, M. Dyer et al. “An elementary analysis of a procedure for sampling points in a convex body”. In: *Random Structures Algorithms* 12.3 (1998), pp. 213–235.
- [22] D. L. Burkholder. “Distribution Function Inequalities for Martingales”. In: *Ann. Probab.* 1.1 (Feb. 1973), pp. 19–42.
- [23] S. Chandrasekhar. “Stochastic problems in physics and astronomy”. In: *Rev. Mod. Phys.* 15 (1943), pp. 1–89.
- [24] Changyou Chen, Nan Ding et al. “On the Convergence of Stochastic Gradient MCMC Algorithms with High-Order Integrators”. In: *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*. Ed. by Corinna Cortes, Neil D. Lawrence et al. 2015, pp. 2278–2286.
- [25] T. Chen, E. Fox et al. “Stochastic Gradient Hamiltonian Monte Carlo”. In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by E. P. Xing and T. Jebara. Vol. 32. Proceedings of Machine Learning Research. Beijing, China: PMLR, 22–24 June 2014, pp. 1683–1691.
- [26] X. Cheng, N. S. Chatterji et al. “Underdamped Langevin MCMC: A non-asymptotic analysis”. In: *Proceedings of the 31st Conference On Learning Theory*. Ed. by S. Bubeck, V. Perchet et al. Vol. 75. Proceedings of Machine Learning Research. PMLR, June 2018, pp. 300–323.
-

- [27] A. S. Cherny and H.-J. Engelbert. *Singular stochastic differential equations*. Vol. 1858. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2005, pp. viii+128.
- [28] S. Chunmei, X. Yu et al. “The convergence and MS stability of exponential Euler method for semilinear stochastic differential equations”. In: *Abstract and Applied Analysis* 2012 (2012), pp. 1–19.
- [29] G. Ciccotti, R. Kapral et al. “Non-Equilibrium Molecular Dynamics”. In: *Handbook of Materials Modeling: Methods*. Ed. by S. Yip. Springer, 2005, pp. 745–761.
- [30] Arnak S Dalalyan and Avetik Karagulyan. “User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient”. In: *Stochastic Processes and their Applications* (2019).
- [31] Arnak S Dalalyan and Lionel Riou-Durand. “On sampling from a log-concave density using kinetic Langevin diffusions”. In: *Bernoulli* 26.3 (2020), pp. 1956–1988.
- [32] Arnak S. Dalalyan. “Theoretical guarantees for approximate sampling from smooth and log-concave densities”. In: *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 79.3 (2017), pp. 651–676.
- [33] V. De Bortoli and A. Durmus. “Convergence of diffusions and their discretizations: from continuous to discrete processes and back”. In: *arXiv preprint* 1904.09808 (2019).
- [34] Persi Diaconis, Susan Holmes et al. “Analysis of a nonreversible Markov chain sampler”. In: *Ann. Appl. Probab.* 10.3 (2000), pp. 726–752.
- [35] Nan Ding, Youhan Fang et al. “Bayesian Sampling Using Stochastic Gradient Thermostats”. In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani, M. Welling et al. Vol. 27. Curran Associates, Inc., 2014, pp. 3203–3211.
- [36] R. Douc, A. Durmus et al. “Boost your favorite Markov Chain Monte Carlo sampler using Kac’s theorem: the Kick-Kac teleportation algorithm”. In: *arXiv preprint arXiv:2201.05002* (2022).
- [37] R. Douc, E. Moulines et al. *Markov Chains*. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018, pp. xviii+757.
- [38] M. Dufflo. *Random Iterative Models*. Vol. 34. Springer Science & Business Media, 2013.
- [39] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis. I. Differentiation*. Vol. 86. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2004, pp. xviii+422.
- [40] A. Durmus, A. Eberle et al. “Discrete sticky couplings of functional autoregressive processes”. In: *arXiv preprint arXiv:2104.06771* (2021).

-
- [41] A. Durmus, A. Enfroy et al. “Uniform minorization condition and convergence bounds for discretizations of kinetic Langevin dynamics”. In: *arXiv preprint arXiv:2107.14542* (2021).
- [42] A. Durmus, S. Majewski et al. “Analysis of Langevin Monte Carlo via Convex Optimization”. In: *Journal of Machine Learning Research* 20.73 (2019), pp. 1–46.
- [43] A. Durmus and É. Moulines. “Nonasymptotic convergence analysis for the unadjusted Langevin algorithm”. In: *Ann. Appl. Probab.* 27.3 (2017), pp. 1551–1587.
- [44] A. Durmus and É. Moulines. “High-dimensional Bayesian inference via the unadjusted Langevin algorithm”. In: *Bernoulli* 25.4A (2019), pp. 2854–2882.
- [45] A. Durmus and E. Moulines. “Supplement to High-dimensional Bayesian inference via the unadjusted Langevin algorithm”. In: *Bernoulli* ().
- [46] A. Durmus, É. Moulines et al. “Irreducibility and geometric ergodicity of Hamiltonian Monte Carlo”. In: *Ann. Statist.* 48.6 (2020), pp. 3545–3564.
- [47] Alain Durmus, Andreas Eberle et al. “Sticky nonlinear SDEs and convergence of McKean-Vlasov equations without confinement”. In: *arXiv preprint arXiv:2201.07652* (2022).
- [48] A. Eberle. “Reflection couplings and contraction rates for diffusions”. English. In: *Probab. Theory Related Fields* (2015), pp. 1–36.
- [49] A. Eberle, A. Guillin et al. “Coupling and quantitative contraction rates for Langevin dynamics”. In: *Ann. Probab.* 47.4 (2019), pp. 1982–2010.
- [50] A. Eberle and M. Majka. “Quantitative contraction rates for Markov chains on general state spaces”. In: *Electronic Journal of Probability* 24 (2019).
- [51] A. Eberle and R. Zimmer. “Sticky couplings of multidimensional diffusions with different drifts”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 55. 4. Institut Henri Poincaré. 2019, pp. 2370–2394.
- [52] 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.
- [53] D. L. Ermak and H. Buckholz. “Numerical integration of the Langevin equation: Monte Carlo simulation”. In: *Journal of Computational Physics* 35.2 (1980), pp. 169–182.
- [54] 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.
- [55] D. J. Evans and G. P. Morriss. *Statistical Mechanics of Nonequilibrium Liquids*. Cambridge University Press, 2008.
- [56] F. Feroz, M. Hobson et al. “MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics”. In: *Monthly Notices of the Royal Astronomical Society* 398.4 (2009), pp. 1601–1614.
-

- [57] James Allen Fill. “An interruptible algorithm for perfect sampling via Markov chains”. In: *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*. 1997, pp. 688–695.
- [58] D. Frenkel and B. Smit. *Understanding Molecular Simulation, From Algorithms to Applications*. 2nd. Academic Press, 2002.
- [59] Andrew Gelman, Hal S Stern et al. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.
- [60] Charles J Geyer. *Markov chain Monte Carlo maximum likelihood*. Interface Foundation of North America, 1991.
- [61] W. K. Hastings. “Monte Carlo sampling methods using Markov chains and their applications”. In: *Biometrika* 57.1 (Apr. 1970), pp. 97–109.
- [62] F. Hiai and D. Petz. *Introduction to Matrix Analysis and Applications*. Springer Science & Business Media, 2014.
- [63] J. Hobert, G. Jones et al. “On the applicability of regenerative simulation in Markov chain Monte Carlo”. In: *Biometrika* 89.4 (2002), pp. 731–743.
- [64] J. P. Hobert and C. P. Robert. “A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling”. In: *Ann. Appl. Probab.* 14.3 (2004), pp. 1295–1305.
- [65] Martina Hofmanová and Jan Seidler. “On weak solutions of stochastic differential equations”. In: *Stochastic analysis and applications* 30.1 (2012), pp. 100–121.
- [66] Martina Hofmanová and Jan Seidler. “On weak solutions of stochastic differential equations II”. In: *Stochastic Analysis and Applications* 31.4 (2013), pp. 663–670.
- [67] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.
- [68] C. J. Howitt. “Stochastic flows and sticky Brownian motion”. PhD thesis. University of Warwick, 2007.
- [69] Alexander T Ihler, John W Fisher et al. “Nonparametric belief propagation for self-localization of sensor networks”. In: *IEEE Journal on Selected Areas in Communications* 23.4 (2005), pp. 809–819.
- [70] Pierre E Jacob, John O’Leary et al. “Unbiased Markov chain Monte Carlo methods with couplings”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82.3 (2020), pp. 543–600.
- [71] J. E. Johndrow, J. C. Mattingly et al. “Approximations of Markov chains and high-dimensional Bayesian inference”. In: *arXiv preprint arXiv:1508.03387* (2015).
- [72] T Kailath, A Segall et al. “Fubini-type theorems for stochastic integrals”. In: *Sankhyā: The Indian Journal of Statistics, Series A* (1978), pp. 138–143.
- [73] Olav Kallenberg. *Foundations of modern probability*. Second. Probability and its Applications (New York). Springer-Verlag, New York, 2002, pp. xx+638.

-
- [74] A. Klenke. *Probability Theory: A Comprehensive Course*. 2nd ed. Universitext. Springer-Verlag London, 2014.
- [75] Y. Komori and K. Burrage. “A stochastic exponential Euler scheme for simulation of stiff biochemical reaction systems”. In: *BIT Numerical Mathematics* 54.4 (2014), pp. 1067–1085.
- [76] V. Konakov, S. Menozzi et al. “Explicit parametrix and local limit theorems for some degenerate diffusion processes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 46.4 (2010), pp. 908–923.
- [77] M. Kopec. “Weak backward error analysis for Langevin process”. In: *BIT Numerical Mathematics* 55.4 (2015), pp. 1057–1103.
- [78] S. Kou, Q. Zhou et al. “Equi-energy sampler with applications in statistical inference and statistical mechanics”. In: *Ann. Statist.* 34.4 (2006), pp. 1581–1619.
- [79] Shiwei Lan, Jeffrey Streets et al. “Wormhole Hamiltonian Monte Carlo”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 28. 1. 2014.
- [80] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, 2016.
- [81] Anthony Lee, Arnaud Doucet et al. “Perfect simulation using atomic regeneration with application to sequential Monte Carlo”. In: *arXiv:1407.5770* (2014).
- [82] B. Leimkuhler and C. Matthews. *Molecular Dynamics: With Deterministic and Stochastic Numerical Methods*. Springer, 2015.
- [83] B. Leimkuhler, C. Matthews et al. “The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics”. In: *IMA J. Numer. Anal.* 36.1 (2016), pp. 13–79.
- [84] B. Leimkuhler and Ch. Matthews. “Rational construction of stochastic numerical methods for molecular sampling”. In: *Appl. Math. Res. Express* (2013), pp. 34–56.
- [85] T. Lelièvre, M. Ramil et al. “A probabilistic study of the kinetic Fokker-Planck equation in cylindrical domains”. In: *arXiv preprint 2010.10157* (2020).
- [86] T. Lelièvre and G. Stoltz. “Partial differential equations and stochastic methods in molecular dynamics”. In: *Acta Numerica* 25 (2016), pp. 681–880.
- [87] Samuel Livingstone, Michael F Faulkner et al. “Kinetic energy choice in Hamiltonian/hybrid Monte Carlo”. In: *Biometrika* 106.2 (2019), pp. 303–319.
- [88] M. Ludkin and C. Sherlock. “Hug and Hop: a discrete-time, non-reversible Markov chain Monte Carlo algorithm”. In: *arXiv:1907.13570* (2019).
- [89] R. B. Lund, S. P. Meyn et al. “Computable exponential convergence rates for stochastically ordered Markov processes”. In: *Ann. Appl. Probab.* 6.1 (Feb. 1996), pp. 218–237.
- [90] M. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2010.
-

- [91] Y-A. Ma, T. Chen et al. “A Complete Recipe for Stochastic Gradient MCMC”. In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes, N. Lawrence et al. Vol. 28. Curran Associates, Inc., 2015, pp. 2917–2925.
- [92] M. Marcus and H. Minc. *A Survey of Matrix Theory and Matrix Inequalities*. Vol. 14. Courier Corporation, 1992.
- [93] Jean-Michel Marin and Christian P. Robert. *Bayesian core: a practical approach to computational Bayesian statistics*. New-York: Springer, 2007.
- [94] George Weiss Marius Tucsnak. *Observation and Control for Operator Semigroups (Birkhäuser Advanced Texts Basler Lehrbücher)*. 1st ed. Birkhäuser Basel, 2009.
- [95] J. C. Mattingly, A. M. Stuart et al. “Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise”. In: *Stochastic Process. Appl.* 101.2 (2002), pp. 185–232.
- [96] R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2020.
- [97] F. Medina-Aguayo, D. Rudolf et al. “Perturbation bounds for Monte Carlo within Metropolis via restricted approximations”. In: *Stochastic Processes and their Applications* 130.4 (2020), pp. 2200–2227.
- [98] N. Metropolis, A. W. Rosenbluth et al. “Equations of state calculations by fast computing machines”. In: *Journal of Chemical Physics* 23.2 (1953), pp. 1087–1092.
- [99] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Second. With a prologue by Peter W. Glynn. Cambridge University Press, Cambridge, 2009, pp. xxviii+594.
- [100] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer, 2012.
- [101] Błażej Miasojedow, Eric Moulines et al. “An adaptive parallel tempering algorithm”. In: *Journal of Computational and Graphical Statistics* 22.3 (2013), pp. 649–664.
- [102] A. Y. Mitrophanov. “Sensitivity and convergence of uniformly ergodic Markov chains”. In: *Journal of Applied Probability* 42.4 (2005), pp. 1003–1014.
- [103] Per Mykland, Luke Tierney et al. “Regeneration in Markov chain samplers”. In: *Journal of the American Statistical Association* 90.429 (1995), pp. 233–241.
- [104] R. M. Neal. “MCMC using Hamiltonian Dynamics”. In: *Handbook of Markov Chain Monte Carlo* (2011), pp. 113–162.
- [105] Radford M Neal. “Sampling from multimodal distributions using tempered transitions”. In: *Statistics and Computing* 6.4 (1996), pp. 353–366.
- [106] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Applied Optimization. Springer, 2004.

-
- [107] G. Parisi. “Correlation functions and computer simulations”. In: *Nuclear Physics B* 180 (1981), pp. 378–384.
- [108] G. A. Pavliotis. *Stochastic Processes and Applications*. Vol. 60. Texts in Applied Mathematics. Springer, New York, 2014.
- [109] M. Pelletier. “Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing”. In: *Ann. Appl. Probab.* 8.1 (1998), pp. 10–44.
- [110] James Gary Propp and David Bruce Wilson. “Exact sampling with coupled Markov chains and applications to statistical mechanics”. In: *Random Structures & Algorithms* 9.1-2 (1996), pp. 223–252.
- [111] Z. Qian and W. Zheng. “Sharp bounds for transition probability densities of a class of diffusions”. In: *C. R. Math. Acad. Sci. Paris* 335.11 (2002), pp. 953–957.
- [112] Z. Qian and W. Zheng. “A representation formula for transition probability densities of diffusions and applications”. In: *Stoch. Proc. Appl.* 111.1 (2004), pp. 57–76.
- [113] Miklos Z. Racz and Mykhaylo Shkolnikov. “Multidimensional sticky Brownian motions as limits of exclusion processes”. In: *Ann. Appl. Probab.* 25.3 (June 2015), pp. 1155–1188.
- [114] M. Raginsky, A. Rakhlin et al. “Non-convex learning via Stochastic Gradient Langevin Dynamics: a nonasymptotic analysis”. In: *Proceedings of the 2017 Conference on Learning Theory*. Ed. by Satyen Kale and Ohad Shamir. Vol. 65. Proceedings of Machine Learning Research. PMLR, July 2017, pp. 1674–1703.
- [115] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. Second. Vol. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1994, pp. xii+560.
- [116] Daniel Revuz and Marc 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.
- [117] L. Rey-Bellet. “Ergodic Properties of Markov Processes”. In: *Open Quantum Systems II*. Ed. by S. Attal, A. Joye et al. Vol. 1881. Lecture Notes in Mathematics. Springer, 2006, pp. 1–39.
- [118] C. Robert. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer Texts in Statistics. Springer New York, 2007.
- [119] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (2nd ed.)* New York: Springer-Verlag, 2004.
- [120] C.P. Robert. *Le choix bayésien: Principes et pratique*. Statistique et probabilités appliquées. Springer Paris, 2006.
- [121] G. O. Roberts and R. L. Tweedie. “Exponential convergence of Langevin distributions and their discrete approximations”. In: *Bernoulli* 2.4 (1996), pp. 341–363.
-

- [122] 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.
- [123] D. Rudolf and N. Schweizer. “Perturbation theory for Markov chains via Wasserstein distance”. In: *Bernoulli* 24.4A (2018), pp. 2610–2639.
- [124] Mark J Schervish. *Theory of Statistics*. 1st ed. Springer Series in Statistics. Springer-Verlag New York, 1995.
- [125] T. Shardlow and A. M. Stuart. “A Perturbation Theory for Ergodic Markov Chains and Application to Numerical Approximations”. In: *SIAM Journal on Numerical Analysis* 37.4 (2000), pp. 1120–1137. eprint: <https://doi.org/10.1137/S0036142998337235>.
- [126] D. Talay. “Stochastic Hamiltonian dissipative systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme”. In: *Markov Proc. Rel. Fields* 8 (2002), pp. 163–198.
- [127] 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).
- [128] L. Tierney. “Markov Chains for Exploring Posterior Distributions”. In: *Ann. Statist.* 22.4 (Dec. 1994), pp. 1701–1728.
- [129] L. Tierney. “A note on Metropolis-Hastings kernels for general state spaces”. In: *Annals of applied probability* 8 (1998), pp. 1–9.
- [130] C. Villani. “Hypocoercivity”. In: *Mem. Amer. Math. Soc.* 202.950 (2009).
- [131] C. Villani. *Optimal transport*. Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973.
- [132] Andi Q Wang, Murray Pollock et al. “Regeneration-enriched Markov processes with application to Monte Carlo”. In: *Ann. Appl. Probab.* 31.2 (2021), pp. 703–735.
- [133] Fugao Wang and DP Landau. “Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram”. In: *Physical Review E* 64.5 (2001), p. 056101.
- [134] S. Watanabe. “On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions”. In: *J. Math. Kyoto Univ.* 11.1 (1971), pp. 169–180.
- [135] S. Watanabe. “On stochastic differential equations for multi-dimensional diffusion processes with boundary conditions II”. In: *J. Math. Kyoto Univ.* 11.3 (1971), pp. 545–551.
- [136] M. Welling and Y. W. Teh. “Bayesian learning via stochastic gradient Langevin dynamics”. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011, pp. 681–688.

-
- [137] L. Wu. “Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems”. In: *Stoch. Proc. Appl.* 91.2 (2001), pp. 205–238.
- [138] E. Zeidler. *Nonlinear functional analysis and its applications. Fixed-point theorems*. Vol. Vol.1. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1986.

Titre : Contributions à la conception, l'étude et la mise en oeuvre de méthodes de Monte Carlo par chaîne de Markov appliquées à l'inférence bayésienne

Mots clés : Monte Carlo par chaîne de Markov, Diffusion de Langevin sous-amortie, Chaînes de Markov irréversibles, Algorithme de Langevin non ajusté, Inférence bayésienne

Résumé : Cette thèse s'intéresse à l'analyse et la conception de méthodes de Monte Carlo par chaîne de Markov (MCMC) utilisées dans l'échantillonnage en grande dimension. Elle est constituée de trois parties.

La première introduit une nouvelle classe de chaînes de Markov et méthodes MCMC. Celles-ci permettent d'améliorer des méthodes MCMC à l'aide d'échantillons visant une restriction de la loi cible originale sur un domaine choisi par l'utilisateur. Cette procédure donne naissance à une nouvelle chaîne qui tire au mieux parti des propriétés de convergences des deux processus qui lui sont sous-jacents. En plus de montrer que cette chaîne vise toujours la mesure cible originale, nous établissons également des propriétés d'ergodicité sous des hypothèses faibles sur les noyaux de Markov mis en jeu.

La seconde partie de ce document s'intéresse aux discrétisations de la diffusion de Langevin sous-amortie. Cette diffusion ne pouvant être calculée ex-

plicitement en général, il est classique de considérer des discrétisations. Cette thèse établit pour une large classe de discrétisations une condition de minoration uniforme en le pas de temps. Avec des hypothèses supplémentaires sur le potentiel, cela permet de montrer que ces discrétisations convergent géométriquement vers leur unique mesure de probabilité invariante en V -norme.

La dernière partie étudie l'algorithme de Langevin non ajusté dans le cas où le gradient du potentiel est connu à une erreur uniformément bornée près. Cette partie fournit des bornes en V -norme et en distance de Wasserstein entre les itérations de l'algorithme avec le gradient exact et celle avec le gradient approché. Pour ce faire il est introduit une chaîne de Markov auxiliaire qui borne la différence. Il est établi que cette chaîne auxiliaire converge en loi vers un processus dit collant déjà étudié dans la littérature pour la version continue de ce problème.

Title : Contributions to the design, study and implementation of Markov chain Monte Carlo methods applied to Bayesian inference

Keywords : Markov chain Monte Carlo, Underdamped Langevin diffusion, Non-reversible Markov chain, Unadjusted Langevin algorithm, Bayesian inference

Abstract : This thesis focuses on the analysis and design of Markov chain Monte Carlo (MCMC) methods used in high-dimensional sampling. It consists of three parts.

The first part introduces a new class of Markov chains and MCMC methods. These methods allow to improve MCMC methods by using samples targeting a restriction of the original target distribution on a domain chosen by the user. This procedure gives rise to a new chain that takes advantage of the convergence properties of the two underlying processes. In addition to showing that this chain always targets the original target measure, we also establish ergodicity properties under weak assumptions on the Markov kernels involved.

The second part of this thesis focuses on discretizations of the underdamped Langevin diffusion. As this diffusion cannot be computed explicitly in general, it is

classical to consider discretizations. This thesis establishes for a large class of discretizations a condition of uniform minimization in the time step. With additional assumptions on the potential, it shows that these discretizations converge geometrically to their unique V -invariant probability measure.

The last part studies the unadjusted Langevin algorithm in the case where the gradient of the potential is known to within a uniformly bounded error. This part provides bounds in V -norm and in Wasserstein distance between the iterations of the algorithm with the exact gradient and the one with the approximated gradient. To do this, an auxiliary Markov chain is introduced that bounds the difference. It is established that this auxiliary chain converges in distribution to sticky process already studied in the literature for the continuous version of this problem.