Weak over-damped asymptotic and variance reduction

Yushun Xu

UNIVERSITÉ PARIS-EST

École Doctorale Mathématiques et STIC

Laboratoire d'Analyse et de Mathématiques Appliquées

Thèse

Présentée pour l'obtention du grade de DOCTEUR
DE L'UNIVERSITE PARIS-EST
par

Yushun Xu

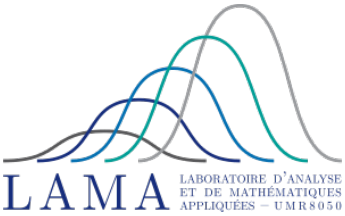
Asymptotique suramortie de la dynamique de Langevin et réduction de variance par repondération

---

Spécialité : Probabilité

Soutenue le devant un jury composé de :

Directeur de thèse	**Mathias Rousset**	(Inria Rennes & Université Rennes 1)
Directeur de thèse	**Pierre-André Zitt**	(Université Marne-la-Vallée)
Rapporteur	**Grigorios A.Pavliotis**	(Imperial College London)
Rapporteur	**Jérémie Bigot**	(Université de Bordeaux)
Examinateur	**Bernard Lapeyre**	(Ecole des Ponts ParisTech)
Examinateur	**Sophie Laruelle**	(Université Paris-Est Créteil)
Examinateur	**Maud Thomas**	(Sorbonne Université)



Thèse effectuée au sein du **Laboratoire d'Analyse et de Mathématiques Appliquées**

de l'Université Paris-Est
Cité Descartes
5, boulevard Descartes
77454 Marne-la-Vallée Cedex 2
France

Résumé

Cette thèse est consacrée à l'étude de deux problèmes différents : l'asymptotique suramortie de la dynamique de Langevin d'une part, et l'étude d'une technique de réduction de variance dans une méthode de Monte Carlo par une repondération optimale des échantillons, d'autre part.

Dans le premier problème, on montre la convergence en distribution de processus de Langevin dans l'asymptotique sur-amortie. La preuve repose sur la méthode classique des "fonctions test perturbées", qui est utilisée pour montrer la tension dans l'espace des chemins, puis pour identifier la limite comme solution d'un problème de martingale. L'originalité du résultat tient aux hypothèses très faibles faites sur la régularité de l'énergie potentielle.

Dans le deuxième problème, nous concevons des méthodes de réduction de la variance pour l'estimation de Monte Carlo d'une espérance de type $\mathbb{E}[\phi(X, Y)]$, lorsque la distribution de X est exactement connue. L'idée générale est de donner à chaque échantillon un poids, de sorte que la distribution empirique pondérée qui en résulte ait une marginale par rapport à la variable X aussi proche que possible de sa cible. Nous prouvons plusieurs résultats théoriques sur la méthode, en identifiant des régimes où la réduction de la variance est garantie. Nous montrons l'efficacité de la méthode en pratique, par des tests numériques qui comparent diverses variantes de notre méthode avec la méthode naïve et des techniques de variable de contrôle. La méthode est également illustrée pour une simulation d'équation différentielle stochastique de Langevin.

Mots-clé:

Dynamiques de Langevin Asymptotique suramortie Convergence faible Probléme de martingale Méthode de Monte-Carlo Réduction de variance Distance de Wasserstein

Weak over-damped asymptotic and variance reduction

Abstract

This dissertation is devoted to studying two different problems: the over-damped asymptotics of Langevin dynamics and a new variance reduction technique based on an optimal reweighting of samples.

In the first problem, the convergence in distribution of Langevin processes in the over-damped asymptotic is proven. The proof relies on the classical perturbed test function (or corrector) method, which is used (i) to show tightness in path space, and (ii) to identify the extracted limit with a martingale problem. The result holds assuming the continuity of the gradient of the potential energy, and a mild control of the initial kinetic energy.

In the second problem, we devise methods of variance reduction for the Monte Carlo estimation of an expectation of the type $\mathbb{E}[\phi(X, Y)]$, when the distribution of X is exactly known. The key general idea is to give each individual sample a weight, so that the resulting weighted empirical distribution has a marginal with respect to the variable X as close as possible to its target. We prove several theoretical results on the method, identifying settings where the variance reduction is guaranteed, and also illustrate the use of the weighting method in Langevin stochastic differential equation. We perform numerical tests comparing the methods and demonstrating their efficiency.

Keywords:

Langevin dynamics Overdamped asymptotics Weak convergence Martingale problem Monte Carlo method Variance reduction reweighting of samples Wasserstein distance.

Acknowledgements

First and foremost I want to thank my advisors Mathias Rousset and Pierre-André Zitt. It has been an honor be to their first Ph.D. student. You have been a tremendous mentor for me. I would like to thank you for encouraging my research and your advice on both research as well as on my career have been invaluable.

Besides my advisor, I would like to thank the rest of my thesis committee: Grigorios A.Pavliotis, Jérémie Bigot, Bernard Lapeyre, Sophie Laruelle, and Maud Thomas, for their insightful comments and encouragement, but also for the hard question which invented me to widen my research from various perspectives.

My sincere thanks also goes to Prof. Changxing Miao and Prof. Marco Cannone , who provided me an opportunity to come to France with the support of Bézout program. Without their precious support and encouragement, it would not be possible to conduct this research.

The members in the laboratories LAMA and Cermics have contributed immensely to my personal and professional time at UPE. The laboratories have been a source of friendships as well as good advice and collaboration. I would like to acknowledge honorary my fellow labmates: Karol Cascavita, Krisztián Benvó, Huong Nguyen, Revekka Kyriakoglou, for all the fun we have had in the last three years. In particular, I would like to thank Lingling Cao for her modification of the abstract in french version.

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. Words can not express how grateful I am to my father and mother for all of the sacrifices that you have made on my behalf. I also would like to thank my brother Dr. Yuxiang Xu and my sister Yilin Guo, thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience. And most of all for my loving, supportive, encouraging, and patient wife Yue Wu whose faithful support during the final period of this Ph.D. is so appreciated. Thank you.

Contents

Part I

General introduction of main results

Chapter 1

Weak over-damped asymptotic of Langevin processes

The first chapter in this thesis focuses on the over-damped asymptotic of Langevin dynamics.

1.1 Langevin and over-damped Langevin process

We study in this section the movements of physical particles, characterized by a position and a momentum. A generic element of the position space $\mathbb{T}^d$ ($\mathbb{T}^d$ is the torus in dimension d) will be denoted by $(q_1, ..., q_d)$ and a generic element of the momentum space $\mathbb{R}^d$ by $(p_1, ..., p_d)$. The total energy of the molecular system is given by the Hamiltonian

$$H(q,p) = E_{\text{kin}}(p) + V(q). \tag{1.1.1}$$

In the above expression, the kinetic energy is $E_{\text{kin}}(p) = \frac{1}{2}p^T M^{-1} p$ (M is the mass), and V is the potential energy experienced by one particle.

Let us now make precise the way the trajectories $(q(t), p(t))_{t\geq 0}$ are computed in practice. We denote by Φ the flow of the Hamiltonian dynamics, i.e. $\Phi(q^0, p^0)$ is the solution at time t of the Hamiltonian equation

$$\begin{cases} \frac{dq(t)}{dt} = \nabla_p H\left(q(t), p(t)\right) = M^{-1}p(t), \\ \frac{dp(t)}{dt} - \nabla_q H\left(q(t), p(t)\right) = -\nabla V\left(q(t)\right), \end{cases} \tag{1.1.2}$$

with initial conditions $(q(0), p(0)) = (q^0, p^0)$.

After a general presentation of Langevin dynamics and Over-damped Langevin dynamics in its usual form in Section 1.1.1 and Section 1.1.2, the over-damped limit of Langevin

dynamics is proposed in Section 1.1.3

1.1.1 Langevin dynamics

The Langevin Stochastic Differential Equation (SDE) describes the dynamics of a classical mechanical system perturbed by a stochastic thermostat. In some case studies, this phenomenological model can be derived in some limiting regime [KSTT02], relying on the Mori-Zwanzig formalism [Zwa73]. Historically, the model was introduced by the botanist R.Brown to describe the movement of particles in a fluid, which were undergoing many collisions. From a numerical viewpoint, several studies advocate the use of Langevin dynamics rather than over-damped Langevin dynamics, e.g. Scemama [SLS+06] and Cancès [CLS07]. The system state at time $t \geq 0$ is encoded by its position Q_t and its momentum P_t.

The paradigm of Langevin dynamics is to introduce in the Newton equation of motion (1.1.2) some fictitious Brownian forces modelling fluctuations, balanced by viscous damping forces modelling dissipation. More precisely, the equations of motion read:

$$\begin{cases} dQ_t &= M^{-1}P_t dt, \\ dP_t &= -\nabla V(Q_t)dt - \gamma M^{-1}P_t dt + \sigma dW_t, \end{cases} \tag{1.1.3}$$

where in the above, (Q_t, P_t) take values in $\mathbb{T}^d \times \mathbb{R}^d$, the function $V : \mathbb{T}^d \to \mathbb{R}$ is the particles' potential energy, and $t \mapsto W_t \in \mathbb{R}^d$ is a standard d-dimensional Brownian motion.

The term σdW_t is a fluctuation term bringing energy into the system, while this energy is dissipated through the friction term $-\gamma M^{-1}P_t dt$; the sum of these two terms forming the so-called thermostat part.

The Langevin dynamics can be considered as a perturbation of the Newton dynamics (for which $\gamma = 0$ and $\sigma = 0$). Denote by $\beta = 1/\kappa_B T$ (T denotes the temperature and κ_B the Boltzmann constant), the magnitudes σ and γ of the random forces σdW_t and of the drag term $-\gamma M^{-1}P_t dt$ are related through the fluctuation-dissipation formula

$$\sigma^2 = \frac{2\gamma}{\beta}. \tag{1.1.4}$$

The generator L' associated with (1.1.3) acts on smooth test functions f of the variable (q, p) and is given formally by:

$$L'f(q,p) := p \cdot (\nabla_q f - \nabla_p f) + \frac{1}{\beta}\Delta_p g - \nabla_q V \cdot \nabla_p f. \tag{1.1.5}$$

1.1.2 Over-damped Langevin dynamics

Over-damped processes are stochastic dynamics on the system positions $q \in \mathbb{T}^d$ only. It is defined by the dynamics:

$$dQ_t = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t, \tag{1.1.6}$$

where $t \mapsto B_t \in \mathbb{R}^d$ is a standard d-dimensional Wiener process. The generator L associated with (1.1.6) acts on smooth test functions f of the variable q as follows:

$$Lf(q) := -\nabla_q V \cdot \nabla_q f + \frac{1}{\beta}\Delta_q f. \tag{1.1.7}$$

1.1.3 Over-damped limit of the Langevin dynamics

Now we will follow [LRS10, Section 2.2.4] to discuss the limit of the stochastic processes. Over-damped processes can be derived from Langevin processes in the so-called "over-damped regime". Precisely, let us consider for the ease of notation the case when the mass tensor is a scalar times identity, and the diffusion tensor (and thus the friction tensor) is also a scalar times identity which does not depend on position:

$$M = mId \quad \text{and} \quad \gamma \quad \text{and} \quad \sigma \quad \text{are constant and scalar.} \tag{1.1.8}$$

We firstly introduce three units: a unit of time t_0, a unit of length l_0 and a unit of mass m_0. Let us introduce the variables associated to these characteristic quantities:

$$\begin{aligned} \bar{t} = \frac{t}{t_0}, \quad \bar{W}_{\bar{t}} = \frac{1}{\sqrt{t_0}} W_{t_0\bar{t}}, \\ \bar{Q}_{\bar{t}} = \frac{Q_t}{l_0} = \frac{Q_{t_0\bar{t}}}{l_0}, \quad \bar{P}_{\bar{t}} = \frac{P_t}{m_0 l_0 t_0^{-1}} = \frac{P_{t_0\bar{t}}}{m_0 l_0 t_0^{-1}}, \\ \bar{V}(\bar{q}) = \beta V(q) = \beta V(l_0\bar{q}). \end{aligned}$$

By a change of variable, the Langevin equation (1.1.3) then writes:

$$\begin{cases} d\bar{Q}_{\bar{t}} = m_0 m^{-1} \bar{P}_{\bar{t}} d\bar{t}, \\ d\bar{P}_{\bar{t}} = -m_0^{-1} l_0^{-2} \beta^{-1} t_0^2 \nabla_{\bar{q}} \bar{V}(\bar{Q}_{\bar{t}}) d\bar{t} - \gamma m^{-1} t_0 \bar{P}_{\bar{t}} d\bar{t} + \sqrt{2\beta^{-1}\gamma t_0^3 m_0^{-2} l_0^{-2}} d\bar{W}_{\bar{t}}. \end{cases}$$

Using the following non-dimensional numbers:

$$\alpha_1 = \frac{m}{m_0} \quad \alpha_2 = \frac{\gamma t_0}{m_0}, \quad \alpha_3 = \frac{\beta m_0 l_0^2}{t_0^2},$$

the equation can be rewritten as

$$\begin{cases} d\bar{Q}_{\bar{t}} = \bar{v}_{\bar{t}} d\bar{t}, \\ \alpha_1 d\bar{v}_{\bar{t}} = -\frac{1}{\alpha_3}\nabla_{\bar{q}}\bar{V}(\bar{Q}_{\bar{t}})d\bar{t} - \alpha_2 \bar{v}_{\bar{t}} d\bar{t} + \sqrt{\frac{2\alpha_2}{\alpha_3}} d\bar{W}_{\bar{t}}, \end{cases}$$

where we introduced the velocity $v_t = m^{-1}P_t$, which can be written as $\bar{v}_{\bar{t}} = m_0 m^{-1}\bar{P}_{\bar{t}}$.

Consider now the following scaling for a small parameter $\eta > 0$:

$$\frac{1}{\alpha_3} = \alpha_2 = \sqrt{\frac{\alpha_2}{\alpha_3}} = \frac{\alpha_1}{\eta}.$$

Dropping the bar for the ease of notation, we get:

$$\begin{cases} dQ_t = m^{-1}P_t dt, \\ \eta dP_t = -\nabla V(Q_t)dt - m^{-1}P_t dt + \sqrt{2\beta^{-1}} dW_t, \end{cases} \tag{1.1.9}$$

then we can rewrite it as

$$\begin{cases} dQ_t^\varepsilon = \frac{1}{\varepsilon}P_t^\varepsilon dt, \\ dP_t^\varepsilon = -\frac{1}{\varepsilon}\nabla V(Q_t^\varepsilon)dt - \frac{1}{\varepsilon^2}P_t^\varepsilon dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}} dW_t, \end{cases} \tag{1.1.10}$$

where $\varepsilon = \sqrt{\eta}$, $Q_t^\varepsilon = Q_t$, and $P_t^\varepsilon = \varepsilon P_t$.

Note that we allow the potential $V_\varepsilon \in C^1(\mathbb{T}^d)$ to depend on ε and will only suppose that it converges to a limit V; see below for a precise statement:

$$\begin{cases} dQ_t^\varepsilon = \frac{1}{\varepsilon}P_t^\varepsilon dt, \\ dP_t^\varepsilon = -\frac{1}{\varepsilon}\nabla V_\varepsilon(Q_t^\varepsilon)dt - \frac{1}{\varepsilon^2}P_t^\varepsilon dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}} dW_t. \end{cases} \tag{1.1.11}$$

The generator L_ε associated with (1.1.11) is given by

$$L_\varepsilon f(q,p) := \frac{1}{\varepsilon^2}\underbrace{\left(\frac{1}{\beta}\Delta_p f - p\cdot\nabla_p f\right)}_{\text{Thermostat}} + \frac{1}{\varepsilon}\underbrace{(p\cdot\nabla_q f - \nabla_q V_\varepsilon\cdot\nabla_p f)}_{\text{Hamiltonian dynamics}}, \tag{1.1.12}$$

where f denotes any smooth test function of the variables $(q,p) \in \mathbb{T}^d \times \mathbb{R}^d$.

The case we consider here is the so-called over-damped asymptotic ($\varepsilon \to 0$), where the time scale of the large damping due to friction is much smaller than the time scale of the Hamiltonian dynamics, so that the momentum becomes a fast variable compared to the slow position variable.

Our main result is the proof of the convergence in distribution of the Langevin position process $(Q_t^\varepsilon)_{t\geq 0}$ towards its overdamped counterpart $(Q_t)_{t\geq 0}$, assuming the uniform convergence of the continuous gradient potential $\|\nabla V_\varepsilon - \nabla V\|_\infty \to 0$, as well as a control of

moments of the initial kinetic energy.

We will start by discussing intuitive ideas, and proofs under strong assumptions.

By noticing that the second equation in (1.1.10) can be reformulated as

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

it is intuitively clear that in the limit $\varepsilon \to 0$, the left hand side should disappear, yielding an autonomous process on the position Q_t, the limiting process on positions is the over-damped Langevin process presented in (1.1.6):

$$dQ_t^0 = -\nabla V(Q_t^0)dt + \sqrt{2\beta^{-1}}dW_t.$$

In the present case, the momentum variable is averaged out with the diffusion approximation, so that the problem may be labeled as "diffusion approximation with averaging". Broadly speaking, the problem can be approached using strong or weak convergence techniques. For an example of the strong convergence approach, the results in [SSMD82] rely on estimating the dynamics of Q_t^ε and its limit using a Gronwall argument; which requires the Lipschitz continuity of ∇V_ε uniformly in ε.

Proposition 1.1.1 (Strong convergence approach)**.** *For any $\varepsilon > 0$, let $(Q_t^\varepsilon, P_t^\varepsilon)_{t\geq 0} \in \mathbb{T}^d \times \mathbb{R}^d$ be the solution to the SDE (1.1.11), with a given initial condition that $(Q_0^\varepsilon, P_0^\varepsilon) = (Q_{init}, P_0)$, and assume that ∇V_ε is a Lipschitz function and converges to ∇V in the sense that $\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow[\varepsilon\to 0]{} 0$. Then, the following path-wise convergence holds: for any time $t > 0$,*

$$\lim_{\varepsilon\to 0} \sup_{0\leq s\leq t} \|Q_s^\varepsilon - Q_s^0\| = 0 \quad a.s.,$$

where $(Q_t^0)_{t\geq 0}$ is the solution to the following over-damped SDE:

$$dQ_t^0 = -\nabla V(Q_t^0)dt + \sqrt{2\beta^{-1}}dW_t, \tag{1.1.13}$$

with the same Brownian motion in SDE (1.1.11) and the initial condition $Q_0^0 = Q_{init}$.

Proof. It is easily seen from (1.1.11) that

$$P_t^\varepsilon = P_0^\varepsilon e^{-t/\varepsilon^2} - \frac{1}{\varepsilon}\int_0^t e^{-(t-s)/\varepsilon^2}\nabla V_\varepsilon(Q_s^\varepsilon)ds \quad + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\int_0^t e^{-(t-s)/\varepsilon^2}dW_s, \tag{1.1.14}$$

also,

$$Q_t^\varepsilon = Q_0^\varepsilon + \frac{1}{\varepsilon}\int_0^t P_s^\varepsilon ds \tag{1.1.15}$$

Plug (1.1.14) in (1.1.15) to get

$$Q_t^\varepsilon = Q_0^\varepsilon + \varepsilon \cdot P_0^\varepsilon(1 - e^{-t/\varepsilon^2}) - \int_0^t (1 - e^{-(t-r)/\varepsilon^2})\nabla V_\varepsilon(Q_r^\varepsilon)dr + \sqrt{2\beta^{-1}} \int_0^t (1 - e^{-(t-r)/\varepsilon^2})dW_r.$$

From (1.1.13),

$$Q_t^0 = Q_0^0 - \int_0^t \nabla V(Q_s^0)ds + \sqrt{2\beta^{-1}} \int_0^t dW_s,$$

thus finally,

$$\begin{aligned} Q_t^\varepsilon - Q_t^0 = &- \int_0^t \left[(1 - e^{-(t-s)/\varepsilon^2})(\nabla V_\varepsilon(Q_s^\varepsilon) - \nabla V(Q_s^0))\right]ds \\ &+ \varepsilon \cdot P_0^\varepsilon(1 - e^{-t/\varepsilon^2}) + \int_0^t e^{-(t-s)/\varepsilon^2}\nabla V(Q_s^0)ds \\ &- \sqrt{2\beta^{-1}} \int_0^t e^{-(t-s)/\varepsilon^2}dW_s. \end{aligned} \tag{1.1.16}$$

The the first term is bounded by $Kt \sup_{s\le t} |Q_s^\varepsilon - Q_s^0| + t\|\nabla V_\varepsilon - \nabla V\|_\infty$, where K is the Lipschitz constant of ∇V_ε. As $\varepsilon \to 0$, the second term on the right-hand side converges to zero uniformly on the compact time intervals. For the third term,

$$\left| \int_0^t e^{-(t-s)/\varepsilon^2}\nabla V(Q_s^0)ds \right| \le \max_{0\le s\le t} |\nabla V(Q_s^0)|\varepsilon^2(1 - e^{-t/\varepsilon^2}),$$

then the integral also converges to zero uniformly on compact time intervals. For the last term, an integration by parts gives:

$$\int_0^t e^{-(t-s)/\varepsilon^2}dW_s = \frac{1}{\varepsilon^2} \int_0^t e^{-(t-s)/\varepsilon^2}(W_t - W_s)ds + W_t e^{-t/\varepsilon^2}.$$

By the continuity of paths of Brownian motion (thus uniform continuity on compact intervals), the first term above goes to zero uniformly on compact time intervals, while the second one converges to zero uniformly on compact time intervals. Finally, for a fixed time t_0, for any time $t \le t_0$,

$$\sup_{s\le t} |Q_s^\varepsilon - Q_s^0| \le K \int_0^t \sup_{r\le s} |Q_r^\varepsilon - Q_r^0|ds + R_{t_0}(\varepsilon),$$

where $R_{t_0}(\varepsilon) \to 0$ when $\varepsilon \to 0$. Then an application of Gronwall's lemma yields the result. □

In some applications, ∇V_ε does not satisfy such a strong Lipschitz condition. Our goal is to weaken this hypotheses and allow for a continuous $\nabla V_\varepsilon \in C(\mathbb{T}^d)$.

The remainder of this chapter is structured as follows. Firstly recall some known classical

tools for weak convergence in Section 1.2. In Section 1.3, we state and prove the weak convergence for a simple toy model. The method is then applied in Section 1.4 to an over-damped limit for irregular potentials.

1.2 Classical tools for weak convergence

Let (E, d) be a Polish space, that is, a topological space which is metric, complete and separable. Denote $C(E)$ the Banach space of all continuous functions and $C_b(E)$ the Banach space of all bounded continuous functions. We denote by $\mathcal{P}(E)$ the space of probability measures on the Borel σ-field $\mathcal{B}(E)$. The notation $\mathcal{F}_t^X$ means the natural filtration of càd-làg processes $(X_t)_{t\geq 0}$, that is $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$. For any $(s,t) \in \mathbb{R} \times \mathbb{R}$, we denote by $s \wedge t$ the minimum of s and t, and $s \vee t$ the maximum of s and t.

1.2.1 The Skorokhod space

Throughout the remaining section of this thesis, (E, r) denotes a metric space, and l denotes the metric $r \wedge 1$.

A càd-làg (French "continu à droite, limité à gauche", also called RCLL for "right continuous with left limits") function is a function defined on $\mathbb{R}_+$ that is everywhere right-continuous and has left limits everywhere. The collection of càd-làg functions on a given domain is known as the Skorokhod space. We denote $\mathbb{D}_E$ the space of càd-làg functions with values in a Polish space E.

Lemma 1.2.1. *If $x \in \mathbb{D}_E[0, +\infty)$, then x has at most countably many points of discontinuity.*

Proposition 1.2.2. *Let $\{x_n\} \subset \mathbb{D}_E[0, +\infty)$ and $x \in \mathbb{D}_E[0, +\infty)$. Then the following are equivalent:*

1. $\lim_{n\to\infty} d(x_n, x) = 0$.
2. *For each $T > 0$, there exists $\{\lambda_n\} \subset \Lambda'$ (possibly depending on T) such that*

$$\begin{cases} \lim_{n\to\infty} \sup_{0\leq t\leq T} |\lambda_n(t) - t| = 0, \\ \lim_{n\to\infty} \sup_{0\leq t\leq T} r\left(x_n(t), x(\lambda_n(t))\right) = 0, \end{cases} \tag{1.2.1}$$

hold.

The following result will be useful in the proof of Theorem 1.4.7.

Lemma 1.2.3. *Integration with respect to time is continuous with respect to the Skorokhod topology: if $(q_t^\varepsilon)_{t\geq 0}$ converges to $(q_t^0)_{t\geq 0}$ in $\mathbb{D}_E$, and $\psi : E \to \mathbb{R}$ is bounded and continuous, then for each $T > 0$,*

$$\int_0^T \psi(q_t^\varepsilon)dt \xrightarrow[\varepsilon\to 0]{} \int_0^T \psi(q_t^0)dt.$$

Proof. Let us denote by $J_T := \{t \in [0,T], q_{t^-}^0 \neq q_t^0\}$ the countable set of jump times in $[0,T]$ of q^0. By definition of convergence in the Skorokhod space,

$$\lim_{\varepsilon\to 0} q_s^\varepsilon = q_s^0 \qquad \forall s \in [0,T] \setminus J_T.$$

Since J_T has Lebesgue measure 0 and ψ is continuous and bounded, dominated convergence yields the result. □

1.2.2 Martingale problems

Let us first recall some basics on martingales and stochastic calculus. Let $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t\geq 0})$ a filtered probability space. A càd-làg real-valued process $(X_t)_{t\geq 0}$ is said to be adapted if X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$, and is called a $(\mathcal{F}_t)_{t\geq 0}$-martingale if $\mathbb{E}(|X_t|\,|\mathcal{F}_s) < +\infty$ and $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for any $0 \leq s \leq t$.

For continuous martingales we have the following important inequality due to Doob: (See e.g. Stroock and Varadhan [SV07], Th. 1.2.3)

Lemma 1.2.4 (Doob's martingale inequality)**.** *If M_t is a martingale such that $t \to M_t(w)$ is continuous a.s., then for all $p \geq 1$, $T \geq 0$ and all $\lambda > 0$*

$$\mathbb{P}\left[\sup_{0\leq t\leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p} \cdot \mathbb{E}\left[|M_t|^p\right].$$

We will often need the technical tool of localization by stopping times, to deal with the unboundedness of the momentum variable. We follow here the presentation of [EK86, Chapter 4].

Definition 1.2.5 (Local martingale)**.** *A càd-làg real-valued process $(X_t)_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, \boldsymbol{P}, (\mathcal{F}_t)_{t\geq 0})$ is called a* local martingale *with respect to $(\mathcal{F}_t)_{t\geq 0}$ if there exists a non-decreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of $(\mathcal{F}_t)_{t\geq 0}$-stopping times such that $\tau_n \to \infty$ $\boldsymbol{P}$-almost surely, and for every $n \in \mathbb{N}$, $(X_{t\wedge\tau_n})_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$-martingale.*

Firstly, based on the standard Itô's calculation (see Def. 3.1.4 and Th. 3.2.1 in [Øks03]), we observe the following:

Lemma 1.2.6. *Let F_t be adapted, continuous function and B_t be Brownian motion. If for any $t > 0$, $\mathbb{E}(\int_0^t F_s^2 ds) < +\infty$, then $\int_0^t F_s dB_s$ is a square integrable martingale and*

$\mathbb{E}(\int_0^t F_s dB_s) = 0$.

Let us now state precisely what it means for a process to solve a martingale problem.

Definition 1.2.7 (Martingale problem)**.** *Let E be a Polish space. Let L be a linear operator mapping a given space $\mathcal{D} \subset C_b(E)$ into bounded measurable functions. Let μ be a probability distribution on E. A càd-làg process $(X_t)_{t\geq 0}$ with values in E solves the* martingale problem *for the generator L on the space $\mathcal{D}$ with initial measure μ - in short, X solves $\boldsymbol{MP}(L, D(L), \mu)$ - if $Law(X_0) = \mu$ and if, for any $\varphi \in \mathcal{D}$,*

$$t \mapsto M_t(\varphi) := \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s)ds \tag{1.2.2}$$

is a martingale with respect to the natural filtration $\left(\mathcal{F}_t^X = \sigma\left(X_s,\, 0 \leq s \leq t\right)\right)_{t\geq 0}$. Moreover, the martingale problem $\boldsymbol{MP}(L, \mathcal{D}, \mu)$ is said to be well-posed *if:*

- *There exists a probability space and a càd-làg process defined on it that solves the martingale problem (existence);*
- *whenever two processes solve $\boldsymbol{MP}(L, \mathcal{D}, \mu)$, then they have the same distribution on $\mathbb{D}_E$ (uniqueness).*

1.2.3 Itô Calculus

The most important tools of integration are change of variable and integration by parts, which are proved on the basis of the formula for differentiating superpositions. The formula for the stochastic differential of a superposition is called *Itô's formula*. We will follow [Øks03], *Chapter 4*, to which we refer for additional material.

Lemma 1.2.8 (The Itô isometry)**.** *Suppose that X_t be an Itô process, then for all $T > 0$*

$$\mathbb{E}\left[\left(\int_0^T X_t dB_t\right)^2\right] = \mathbb{E}\left[\int_0^T X_t^2 dB_t\right].$$

Lemma 1.2.9 (The 1-dimensional Itô's formula)**.** *Let $(\Omega, \mathcal{F}, \boldsymbol{P})$ be a complete probability space. Suppose X_t be an Itô process given by*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t,$$

where B_t is a 1-dimensional Brownian motion. Let $g(t, x) \in C^2([0, \infty) \times \mathbb{R})$ (i.e. g is twice continuously differentiable on $C^2([0, \infty) \times \mathbb{R})$. Then

$$Y_t = g(t, X_t)$$

is again an Itô process, and

$$dY_t = \frac{\partial g}{\partial t}(t, X_t)dt + \frac{\partial g}{\partial x}(t, X_t)dX_t + \frac{1}{2}\sigma^2\frac{\partial^2 g}{\partial x^2}(t, X_t)dt.$$

Lemma 1.2.10 (Integration by parts)**.** *Suppose $f(s, w) = f(s)$ only depend on s and that f is continuous and of bounded variation in $[0, t]$. Then*

$$\int_0^t f(s)dB_s = f(t)B_t - \int_0^t B_s df_s.$$

We not turn to the situation in higher dimensions.

Lemma 1.2.11 (The general Itô's formula)**.** *Let X_t be an d-dimensional Itô process given by*

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t,$$

where B_t denotes m-dimensional Brownian motion, b is a d-vector and σ is a $d \times m$ matrix. Let $g(t, x) = (g_1(t, x), ..., g_p(t, x)) \in C^2([0, \infty) \times \mathbb{R}^d)$ be a twice continuously differentiable function. Then the process $Y(t, w) = g(t, X(t))$ is again an Itô process, whose component number k $(k = 1, ..., p)$, Y_k is given by

$$dY_k = \frac{\partial g_k}{\partial t}(t, X_t)dt + \sum_{i=1}^{d}\frac{\partial g_k}{\partial x_i}(t, X_t)dX_{i,t} + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2 g_k}{\partial x_i \partial x_j}(t, X_t)dX_{i,t}dX_{j,t},$$

where $dX_{i,t}dX_{j,t}$ is computed using the rules $dtdt = dtdB_i = dB_i dt = 0$, $dB_i dB_j = 0$ for all $i \neq j$ and $(dB_i)^2 = dt$.

1.2.4 Weak solutions of SDEs

Let $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d \times n}$ be locally bounded. Consider a stochastic differential equation in $\mathbb{R}^d$ of the form:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{1.2.3}$$

with an initial condition $\text{Law}(X_0) = \mu_0$. Let L be the formal generator

$$L := \sum_{i=1}^{d} b_i \partial_i + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\partial_i \partial_j, \tag{1.2.4}$$

where $a = \sigma\sigma^T$.

Definition 1.2.12 (Weak solution of the SDE)**.** *A continuous process $(X_t)_{t \geq 0}$ is a* weak solution *of (1.2.3) if there exists a filtered probability space $(\Omega, \mathcal{F}, \boldsymbol{P}, (\mathcal{F}_t)_{t \geq 0})$ such that:*

- *$t \mapsto W_t$ is a $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion, that is, an $(\mathcal{F})_{t\geq 0}$-adapted process such that $Law(W_{t+h} - W_t|\mathcal{F}_t) = \mathcal{N}(0,h)$.*
- *X is a continuous, $(\mathcal{F}_t)_{t\geq 0}$-adapted process and satisfies the stochastic integral equation*

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW_s \quad a.s.$$

We now quote two results from [EK86] concerning existence and uniqueness of solutions to SDEs and martingale problems.

The first is an existence result, and can be found in [EK86, Section 5.3] (Corollary 3.4 and Theorem 3.10).

Theorem 1.2.13. *Assume that b, σ are continuous. If there exists a constant K such that for any $t \geq 0$, $x \in \mathbb{R}^d$:*

$$|\sigma|^2 \leq K(1+|x|^2); \tag{1.2.5}$$

$$x \cdot b(x) \leq K(1+|x|^2), \tag{1.2.6}$$

then there exists a weak solution of the stochastic differential equation (1.2.3) corresponding to (σ, b, μ), which is also solution of the martingale problem $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$, $C_c^\infty(\mathbb{R}^d)$ being the set of smooth functions with compact support.

Remark 1.2.14. *For the Langevin equation (1.1.11) we first remark that the latter can be set in $\mathbb{R}^d \times \mathbb{R}^d$ using the $\mathbb{Z}^d$-periodic extension of V_ε. Then $b(q,p) = \left(\frac{1}{\varepsilon}p, -\frac{1}{\varepsilon}\nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p\right)$ and $\sigma = (0, \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\mathrm{Id}_{\mathbb{R}^d})$ are continuous since $V_\varepsilon \in C^1(\mathbb{R}^d)$. Moreover, $|\sigma|^2 = \sigma\sigma^\top = (0, \frac{2}{\beta\varepsilon^2}\mathrm{Id}_{\mathbb{R}^d})$, and on the other hand*

$$(q,p)\cdot b(q,p) = \frac{1}{\varepsilon}pq - \frac{1}{\varepsilon}p\nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p^2 \leq \frac{1}{2\varepsilon}(1+\|\nabla V_\varepsilon\|_\infty)(1+|p|^2+|q|^2),$$

which implies the existence of weak solution of (1.1.11) in $\mathbb{R}^d$. One then obtains existence of a weak solution in $\mathbb{T}^d$ of the original (1.1.11) using the canonical continuous mapping $\mathbb{R}^d \to \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$.

The next result follows from [EK86] (Theorem 1.7 in Section 8.1) and [SV07] (Theorem 10.2.2 and the discussion following their Corollary 10.1.2) .

Theorem 1.2.15. *Assume that the bounds (1.2.5) and (1.2.6) hold. Suppose that $a := \sigma\sigma^\top$ is continuous and uniformly elliptic:*

$$\exists C_a > 0, \forall \xi \in \mathbb{R}^d, \forall x \in \mathbb{R}^d, \quad \xi^\top a(x)\xi \geq C_a|\xi|^2.$$

Then for any initial condition μ, there is a unique weak solution of the stochastic differential equation (1.2.3); which is also the unique solution of the martingale problem

$\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$.

Remark 1.2.16. *For the over-damped Langevin equation (1.1.6), we remark again that the latter can be set in $\mathbb{R}^d$ using the $\mathbb{Z}^d$-periodic extension of V_ε. One then obtains well-posedness of the martingale problem $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ in $\mathbb{R}^d$ since ∇V is bounded and continuous by assumption. This solution obviously solves $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ in $\mathbb{T}^d$. The fact that uniqueness of $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ implies uniqueness of $\boldsymbol{MP}(L, C_c^\infty(\mathbb{T}^d), \mu)$ is technically less obvious. It can be treated using the localization technique of Theorem A.1.1 stated in appendix. More precisely, using the notation of Theorem A.1.1, one can defines the covering of $\mathbb{R}^d$ by the open sets*

$$U_k := \Big\{(x_1, \dots, x_d) \in \mathbb{R}^d | \, |x_i - k_i/8| \leq 1/4 \ \forall i = 1 \dots d\Big\}$$

where $k \in \mathbb{Z}^d$ and then remarks that by partition of unity for smooth functions, any $\varphi \in C_c^\infty(\mathbb{R}^d)$ can be written as a finite sum of smooth functions with compact support in each given U_k, $k \in \mathbb{Z}^d$.

1.2.5 Convergence in distribution

Let us briefly recall several key results that will be used later in weak convergence.

For completeness, we start by recalling the very classical Prohorov theorem, characterizing relative compactness by tightness (see for example Section 2 in[EK86, Chapter 3]).

Theorem 1.2.17 (Prohorov theorem)**.** *Let $(\mu_\varepsilon)_\varepsilon$ be a family of probability measures on a Polish space E. Then the following are equivalent:*

1. *$(\mu_\varepsilon)_\varepsilon$ is relatively compact for the topology of convergence in distribution.*
2. *$(\mu_\varepsilon)_\varepsilon$ is tight, that is to say, for any $\delta > 0$, there is a compact set K_δ such that*

$$\inf_\varepsilon \mu_\varepsilon(K_\delta) \geq 1 - \delta.$$

Over the years several relative compactness criteria in Skorokhod space have been developed. We will use the following one [EK86, Theorem 8.6, Chapter 4].

Theorem 1.2.18 (Kurtz-Aldous tightness criterion)**.** *Consider a family of stochastic processes $((X_t^\varepsilon)_{t\geq 0})_\varepsilon$ in $\mathbb{D}_\mathbb{R}$. Assume that $\big(Law(X_0^\varepsilon)\big)_\varepsilon$ is tight. $\forall \delta \in (0,1)$ and $T > 0$, there exists a family of nonnegative random variable $\Gamma_{\varepsilon,\delta}$, such that: $\forall\ 0 \leq t \leq t+h \leq t+\delta \leq T$*

$$\mathbb{E}\Big(|X_{t+h}^\varepsilon - X_t^\varepsilon|^2 | \mathcal{F}_t^{X^\varepsilon}\Big) \leq \mathbb{E}(\Gamma_{\varepsilon,\delta} | \mathcal{F}_t^{X^\varepsilon}); \tag{1.2.7}$$

with

$$\lim_{\delta\to 0}\sup_{\varepsilon}\mathbb{E}(\Gamma_{\varepsilon,\delta})=0. \tag{1.2.8}$$

Then the family of distributions $(Law((X_t^\varepsilon)_{t\geq 0}))_\varepsilon$ *is tight.*

Remark 1.2.19 (On using sequences)**.** *If* $\varepsilon>0$ *is a real number and that instead of* (1.2.8)*, one considers the condition* $\lim_{\delta\to 0}\limsup_{\varepsilon\to 0^+}\mathbb{E}(\Gamma_{\varepsilon,\delta})=0$*, then the conclusion becomes the following:* $(Law((X_t^{\varepsilon_n})_{t\geq 0}))_{\varepsilon_n}$ *is tight for any* $(\varepsilon_n)_{n\geq 1}$*-sequence such that* $\varepsilon_n>0$ *and* $\lim_{n\to+\infty}\varepsilon_n=0$*. This version will be the one used in this thesis.*

If the processes, say $(Q_t^\varepsilon)_{t\geq 0}$, is defined in a general state space E, it is natural to consider the image processes $(f(Q_t^\varepsilon))_{t\geq 0}$ for various observables, or test functions, f. The following result enables us to recover the tightness for the original process from the tightness of the observed processes (Corollary 9.3 Chapter 3 in [EK86]).

Theorem 1.2.20 (Tightness from observables)**.** *Let* E *be a compact Polish space and* $((Q_t^\varepsilon)_{t\geq 0})_{\varepsilon>0}$ *be a family of stochastic processes in* $\mathbb{D}_E$*. Assume that there is an algebra of test functions* $\mathcal{D}\subset C_b(E)$*, dense for the uniform convergence, such that for any* $f\in\mathcal{D}$*,* $((f(Q_t^\varepsilon))_{t\geq 0})_{\varepsilon>0}$ *is tight in* $\mathbb{D}_{\mathbb{R}}$*. Then* $(Law(Q_t^\varepsilon)_{t\geq 0})_{\varepsilon>0}$ *is tight in* $\mathbb{D}_E$*.*

Remark 1.2.21. *Again, the above theorem will be used for families indexed by sequences* $(\varepsilon_n)_{n\geq 1}$ *such that* $\varepsilon_n>0$ *and* $\lim_{n\to+\infty}\varepsilon_n=0$*.*

Finally, the following two lemmas will be useful when we considering martingale problems. The first one states that the distribution of jumps of càd-làg processes have atoms in a countable set (see Lemma 7.7 Chapter 3 in [EK86]).

Lemma 1.2.22. *Let* $(X_t)_{t\geq 0}$ *be a random process in the Skorokhod path space* $\mathbb{D}_E$*. The set of instants where no jump occurs almost surely:*

$$\mathcal{C}_{Law(X)}:=\{t\in\mathbb{R}^+|\mathbb{P}(X_{t^-}=X_t)=1\},$$

has countable complement in $\mathbb{R}^+$*. In particular, it is a dense set.*

The second one is a very useful way to check that whether a process is a martingale or not (see page 174 in Ethier-Kurtz[EK86]).

Lemma 1.2.23 (Martingale equivalent condition)**.** *Let* $(M_t)_{t\geq 0}$ *and* $(X_t)_{t\geq 0}$ *be two càd-làg processes and let* $\mathcal{C}$ *be an arbitrary dense subset of* $\mathbb{R}_+$*. Then* $(M_t)_{t\geq 0}$ *is* $\mathcal{F}_t^X$*-martingale if and only if*

$$\mathbb{E}[(M_{t_{k+1}}-M_{t_k})\varphi_k(X_{t_k})...\varphi_1(X_{t_1})]=0,$$

for any time ladder $t_1\leq ...\leq t_{k+1}\in\mathcal{C}\subset\mathbb{R}_+$*,* $k\geq 1$*, and* $\varphi_1,...,\varphi_k\in C_b(E)$*.*

1.3 Weak convergence approach for a simple example

As we said, we are interested here in proving weak convergence for processes. Weak convergence results rely on the so-called "perturbed" test function or "corrector" approach, that have been developed since Panicolaou-Stroock-Varadhan in [PSV77]. The literature on diffusion approximations is very rich; we refer for instance to Stuart-Pavliotis in [PS08] for a recent pedagogical overview of related issues. Historically, a possible chain of seminal references is given by Stratonovich in [Str63], Khas'minskii in [Kha66], Papanicolaou-Varadhan in [PV73], as well as Papanicolaou-Kohler in [PK74]; complemented with the more modern viewpoint of Ethier-Kurtz in [EK86], Chapter 12 "Random evolutions".

This classical perturbed test function method, which is used both to show tightness in path space and to identify the extracted limit with a martingale problem, can be summed up as follows: for f a function of q, can we perturb it to get f_ε such that both $f_\varepsilon - f$ and $L_\varepsilon f_\varepsilon - Lf$ are small?

To give an idea of the techniques involved, let us work through a simple example: Zig Zag on the circle.

For each ε, we consider a càd-lag process $t \to (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^1 \times \{-1, 1\}$, where the position Q_t^ε follows the momentum P_t^ε with speed $\frac{1}{\varepsilon}$ and P_t^ε jumps to $-P_t^\varepsilon$ at rate $\frac{1}{\varepsilon^2}$. The natural filtration of the full process and the process $(Q_t^\varepsilon)_{t\geq 0}$ are denoted respectively by $\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon} := \sigma\left((Q_s^\varepsilon, P_s^\varepsilon), 0 \leq s \leq t\right)$, and $\mathcal{F}_t^{Q^\varepsilon} := \sigma\left(Q_s^\varepsilon, 0 \leq s \leq t\right)$. Then the Markov generator L_ε associated with this model is given by

$$L_\varepsilon f(q,p) = \frac{1}{\varepsilon} p \cdot \nabla_q f + \frac{1}{\varepsilon^2}\left(f(q,-p) - f(q,p)\right),$$

where f denotes any smooth test function of the variables $(q,p) \in \mathbb{T}^1 \times \{-1,1\}$. We first construct a perturbed test function $f_\varepsilon \in C^\infty\left(\mathbb{T}^1 \times \{-1,1\}\right)$ in the following form (see [PSV77])

$$f_\varepsilon = f(q) + \varepsilon f_1(q,p). \tag{1.3.1}$$

Applying the generator L_ε, using the fact that f does not depend on p, and grouping terms with respect to powers of ε, we get

$$\begin{aligned} L_\varepsilon f_\varepsilon(q,p) &= \frac{1}{\varepsilon} p \cdot \nabla_q f(q) + p \cdot \nabla_q f_1(q,p) + \frac{1}{\varepsilon^2}\left(\varepsilon f_1(q,-p) - \varepsilon f_1(q,p)\right) \\ &= p \cdot \nabla_q f_1(q,p) + \frac{1}{\varepsilon}\left(p \cdot \nabla_q f + f_1(q,-p) - f_1(q,p)\right). \end{aligned}$$

In order for $L_\varepsilon f_\varepsilon$ to converge to Lf, the ε^{-1}-order terms should vanish, and the ε^0-order terms should converge at least formally to $L(f)$. As result, f_1 should solve the following equation

$$0 = p \cdot \nabla_q f + f_1(q,-p) - f_1(q,p). \tag{1.3.2}$$

The function $f_1 = \frac{1}{2} p \cdot \nabla_q f$ clearly solves 1.3.2. Therefore, in view of Eq.(1.3.3), we define the perturbed test function by :

$$f_\varepsilon(q,p) = f(q) + \frac{\varepsilon}{2} p \cdot \nabla_q f. \tag{1.3.3}$$

With this choice since $p^2 = 1$, we get

$$L_\varepsilon f_\varepsilon(q,p) - Lf(q) = \frac{p^2}{2}\Delta_q f - \frac{1}{2}\Delta_q f = 0,$$

where Lf is the generator of Brownian motion on $\mathbb{T}^1$.

Here we state the key assumptions that will imply convergence in distribution of the process $(Q_t^\varepsilon)_{t\geq 0}$ towards the solution of a martingale problem.

Assumption 1.3.1 (Initial condition)**.** *The initial condition* $(Law(Q_0^\varepsilon))_{\varepsilon>0}$ *converge to a limit* μ_0*, when* $\varepsilon \to 0$*.*

Now we will use standard tightness arguments and characterization through martingale problems.

Step one: The proof of tightness. We want to prove that for each sequence $(\varepsilon_n)_{n\geq 1}$ satisfying $\lim_n \varepsilon_n = 0$, $(\mathrm{Law}(Q_t^{\varepsilon_n}))_{n\geq 1}$ is tight. By Theorem 1.2.20, it is enough to prove the tightness of $(\mathrm{Law}\,(f(Q_t^{\varepsilon_n})))_{n\geq 1}$ for all $f \in C^\infty(\mathbb{T})$. The latter fact will follow from Theorem 1.2.18, if we are able to construct, for any function $f \in C^\infty(\mathbb{T})$ and any ε, $\delta > 0$ and any $T > 0$, a random variable $\Gamma_{\varepsilon,\delta}(f)$ such that for all $0 \leq t \leq t+h \leq t+\delta \leq T$, one has

$$\mathbb{E}\left[(f(Q_{t+h}^\varepsilon) - f(Q_t^\varepsilon))^2 \middle| \mathcal{F}_t^{Q^\varepsilon}\right] \leq \mathbb{E}\left[\Gamma_{\varepsilon,\delta}(f) \middle| \mathcal{F}_t^{Q^\varepsilon}\right], \tag{1.3.4}$$

$$\text{where} \quad \lim_{\delta\to 0} \limsup_{\varepsilon\geq 0} \mathbb{E}\left[\Gamma_{\varepsilon,\delta}(f)\right] = 0. \tag{1.3.5}$$

In this introduction, we will only check the following variant, that can be seen to imply (1.3.4), (1.3.5) (see Chapter 3 for details).

Lemma 1.3.2. *For any* $g \in \mathcal{C}^\infty(\mathbb{T}^d)$*, and any* $\delta, \varepsilon, T > 0$*, there exists a random variable* $\Gamma'_{\varepsilon,\delta}(g)$ *such that for all* $0 \leq t \leq t+h \leq t+\delta \leq T$*,*

$$\left|\mathbb{E}\left[g(Q_{t+h}^\varepsilon) - g(Q_t^\varepsilon) \middle| \mathcal{F}_t^{Q^\varepsilon}\right]\right| \leq \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g) \middle| \mathcal{F}_t^{Q^\varepsilon}\right], \tag{1.3.6}$$

$$\textit{where} \quad \lim_{\delta\to 0} \limsup_{\varepsilon\geq 0} \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g)\right] = 0. \tag{1.3.7}$$

Let us now prove the Lemma 1.3.2. Let g be an arbitrary smooth function, and let g_ε be

the perturbed test function given by 1.3.3. An elementary rewriting leads to

$$\begin{aligned} g(Q^\varepsilon_{t+h}) - g(Q^\varepsilon_t) &= \left(g(Q^\varepsilon_{t+h}) - g_\varepsilon(Q^\varepsilon_{t+h}, P^\varepsilon_{t+h})\right) - \left(g(Q^\varepsilon_t) - g_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)\right) \\ &\quad - \int_t^{t+h} \left(Lg(Q^\varepsilon_s) - L_\varepsilon g_\varepsilon(Q^\varepsilon_s, P^\varepsilon_s)\right) ds + \int_t^{t+h} Lg(Q^\varepsilon_s) ds \\ &\quad - M^\varepsilon_t(g_\varepsilon) + M^\varepsilon_{t+h}(g_\varepsilon) \\ &= \frac{\varepsilon}{2} P^\varepsilon_{t+h} \cdot \nabla_q g(Q^\varepsilon_{t+h}) - \frac{\varepsilon}{2} P^\varepsilon_t \cdot \nabla_q g(Q^\varepsilon_t) + \int_t^{t+h} Lg(Q^\varepsilon_s) ds \\ &\quad - M^\varepsilon_t(g_\varepsilon) + M^\varepsilon_{t+h}(g_\varepsilon), \end{aligned} \tag{1.3.8}$$

where $(M^\varepsilon_t(g_\varepsilon))_{t\geq 0}$ is a $\mathcal{F}^{Q^\varepsilon, P^\varepsilon}$-martingale. Taking the conditional expectation with respect to $\mathcal{F}^{Q^\varepsilon}_t$, the martingale terms cancel out, and we get:

$$\begin{aligned} \left|\mathbb{E}\left[g\left(Q^\varepsilon_{t+h}\right) - g(Q^\varepsilon_t)\middle|\mathcal{F}^{Q^\varepsilon}_t\right]\right| &\leq \left|\mathbb{E}\left[\frac{\varepsilon}{2} P^\varepsilon_t \cdot \nabla_q g(Q^\varepsilon_{t+h})\middle|\mathcal{F}^{Q^\varepsilon}_t\right]\right| + \left|\mathbb{E}\left[\frac{\varepsilon}{2} P^\varepsilon_t \cdot \nabla_q g(Q^\varepsilon_t)\middle|\mathcal{F}^{Q^\varepsilon}_t\right]\right| \\ &\quad + h \sup_{q\in\mathbb{T}} |Lg(q)| \\ &\leq \varepsilon \left\|\nabla g\right\|_\infty + \delta \sup_{q\in\mathbb{T}} |Lg(q)| . \end{aligned}$$

Let $\Gamma'_{\varepsilon,\delta}(g) = \varepsilon \left\|\nabla g\right\|_\infty + \delta \sup_{q\in\mathbb{T}} |Lg(q)|$, then the proof of tightness is concluded.

Step 2: identification of the limit. In this step, we suppose that a sequence $Q^n_t = Q^{\varepsilon_n}_t$ converges in distribution to a limit Q^0_t, and we prove that necessarily, Q^0 solves the martingale problem for the generator L.

Let $f \in C^\infty(\mathbb{T}^1)$, we have to check that

$$M_t(Q^0_t) := f(Q^0_t) - f(Q^0_0) - \int_0^t Lf(Q^0_s) ds \tag{1.3.9}$$

is a martingale with respect to $\mathcal{F}^{Q^0}_t = \sigma(Q^0_s, 0 \leq s \leq t)$. Consider a time sequence $0 \leq t_1 \leq \cdots \leq t_p \leq t_{p+1}$ for $p \geq 1$, taken in the continuity set $\mathcal{C}_{\mathrm{Law}(Q)}$ given by Lemma 1.2.22. Recall that $\mathcal{C}_{\mathrm{Law}(Q)}$ is dense in $\mathbb{R}$. Let $\varphi_1, ..., \varphi_p \in C_b(\mathbb{T}^1)$ be p test functions. By the Martingale equivalent condition in Lemma 1.2.23, it is enough to prove that

$$I_0 := \mathbb{E}\left[\left(f(Q^0_{t_{p+1}}) - f(Q^0_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^0_s) ds\right) \varphi_1(Q^0_{t_1}) \cdots \varphi_p(Q^0_{t_p})\right] = 0.$$

Let I_ε be the corresponding quantity for $\varepsilon > 0$, that is,

$$I_\varepsilon := \mathbb{E}\left[\left(f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s) ds\right) \varphi_1(Q^\varepsilon_{t_1}) \cdots \varphi_p(Q^\varepsilon_{t_p})\right].$$

Let us first show that I_ε converges to 0. We first condition on $\mathcal{F}_{t_p}^{Q^\varepsilon}$ to get:

$$|I_\varepsilon| \leq \mathbb{E}\left[\mathbb{E}\left[\left|f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s)ds\right| \middle| \mathcal{F}_{t_p}^{Q^\varepsilon}\right] |\varphi_1(Q^\varepsilon_{t_1})| \cdots \left|\varphi_p(Q^\varepsilon_{t_p})\right|\right]$$
$$\leq \mathbb{E}\left[\mathbb{E}\left[\left|f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s)ds\right| \middle| \mathcal{F}_{t_p}^{Q^\varepsilon}\right]\right] \|\varphi_1\|_\infty \cdots \|\varphi_p\|_\infty.$$

Using again the perturbed test function f_ε and the decomposition in (1.3.8), we get by the same argument as in *Step 1* that

$$|I_\varepsilon| \leq \mathbb{E}\left[\varepsilon \left\|\nabla_q f\right\|_\infty\right] \|\varphi_1\|_\infty ... \|\varphi_p\|_\infty.$$

This implies that $I_\varepsilon \to 0$.

Let us now prove that I_ε converges to I_0. Let $\Phi : \mathbb{D}_\mathbb{T} \to \mathbb{R}$ be the functional

$$\Phi : (q_t)_{t\geq 0} \mapsto \left(f(q_{t_{p+1}}) - f(q_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(q_s)ds\right) \varphi_1(q_{t_1}) \cdots \varphi_p(q_{t_p})$$

so that $I_\varepsilon = \mathbb{E}\left[\Phi((Q^\varepsilon_t)_{t\geq 0})\right]$ and $I_0 = \mathbb{E}\left[\Phi((Q^0_t)_{t\geq 0})\right]$. Let us first check that Φ is a continuous functional of $\mathbb{D}_\mathbb{T}$ at the point $q^0 \in \mathbb{D}_\mathbb{T}$ if $q^0_{t_k^-} = q^0_{t_k}$ for each $1 \leq k \leq p+1$. Indeed: (i) by Lemma 1.2.3, and since Lf is continuous and bounded, the map $(q_t)_{t\geq 0} \mapsto \int_{t_p}^{t_{p+1}} Lf(q_s)ds$ is continuous with respect to Skorokhod topology; and (ii) q^0 is continuous by assumption at the time t_k for each $1 \leq k \leq p+1$, so that the map on $\mathbb{D}_{\mathbb{T}^1}$

$$(q_t)_{t\geq 0} \mapsto \phi_k(q_{t_k})$$

is continuous at $q^0 \in \mathbb{D}_{\mathbb{T}^1}$.

Let now $(\varepsilon_n)_{n\geq 1}$ be any sequence such that $\varepsilon_n \to 0$ and $(Q^{\varepsilon_n}_t)_{t\geq 0}$ converges in distribution to $(Q^0_t)_{t\geq 0}$. The Skorokhod representation theorem (Theorem 1.8 in [EK86, Chapter 3]) ensures that one can construct a probability space where the distribution of $(Q^{\varepsilon_n}_t)_{t\geq 0}$ for each n is unchanged but for which $\lim_{n\to+\infty} Q^{\varepsilon_n} = Q^0$ almost surely in $\mathbb{D}_{\mathbb{T}^1}$. Since $t_k \in \mathcal{C}_{\mathrm{Law}(Q^0)}$ for each $k = 1 \dots p+1$, Ψ is almost surely continuous at Q^0 and we can apply the dominated convergence theorem to obtain $\lim_{n\to+\infty} I_{\varepsilon_n} = I_0$. Since the choice of the vanishing sequence $(\varepsilon_n)_{n\geq 1}$ is arbitrary, we conclude that $\lim_{\varepsilon\to 0} I_\varepsilon = I_0$. The limit process thus solves the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^1), \mu)$.

Conclusion For each sequence $(\varepsilon_n)_{n\geq 1}$ satisfying $\lim_n \varepsilon_n = 0$, we have proven that $(Law(Q^{\varepsilon_n}_t))_{n\geq 1}$ is tight and that any converging subsequence is solution to the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^1), \mu)$. By uniqueness of the latter according to the well-posed of $\mathbf{MP}(L, C^\infty(\mathbb{T}^1), \mu)$, this identifies the limit, showing that $(Law(Q^{\varepsilon_n}_t))_{n\geq 1}$ converges

to the solution of $\mathbf{MP}(L, C^\infty(\mathbb{T}^1), \mu)$. Since the sequence $(\varepsilon_n)_{n\geq 1}$ is arbitrary and convergence in distribution is metrizable, $(Law(Q_t^\varepsilon))_{\varepsilon>0}$ also converges to the solution of $\mathbf{MP}(L, C^\infty(\mathbb{T}^1), \mu)$.

1.4 An over-damped limit for irregular potentials

1.4.1 The main results

Our main result is the proof of the convergence in distribution of the Langevin position $(Q_t^\varepsilon)_{t\geq 0}$ towards its over-damped counterpart $(Q_t)_{t\geq 0}$, assuming the uniform convergence of the gradient potential as well as a control of moments of the initial kinetic energy.

Theorem 1.4.1 (Over-damped limit of the Langevin dynamics)**.** *For any $\varepsilon > 0$, let $(Q_t^\varepsilon, P_t^\varepsilon)_{t\geq 0} \in \mathbb{T}^d \times \mathbb{R}^d$ be a weak solution to the SDE (1.1.11). Assume that the following conditions hold:*

1. *V_ε is $C^1(\mathbb{T}^d)$, and converges to V in the sense that $\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow[\varepsilon\to 0]{} 0$,*
2. *The following moment bound holds true:*

$$\lim_{\varepsilon\to 0} \varepsilon \mathbb{E}(|P_0^\varepsilon|^3) = 0$$

3. *The initial position distribution is converging to some limit: $Law(Q_0^\varepsilon) \xrightarrow[\varepsilon\to 0]{} Law(Q_0)$.*

Then, when $\varepsilon \to 0$, the process $(Q_t^\varepsilon)_{t\geq 0} \in C(\mathbb{R}_+ \to \mathbb{T}^d)$ converges in distribution to the unique weak solution of the over-damped SDE (1.1.6).

Remark 1.4.2. *There are two main differences with respect to the simple case that we need to take into account: the momentum p is no longer bounded, and $L_\varepsilon f_\varepsilon(q,p) - Lf(q)$ will not be identically zero.*

In a series of papers [PV01, PV03, PV05], Pardoux-Veretennikov extend the classical diffusion approximation with averaging to the non-compact state space case. In the latter setting however, the slow variable has a dynamics independent of the fast one, which is not the case in the Langevin case (1.1.11).

In order to prove Theorem 1.4.1, we will establish a more general weak convergence result. We consider a sequence (indexed by a small parameter $\varepsilon > 0$) of Markov processes of the form $t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d$ taking value in the Skorokhod path space $\mathbb{D}_{\mathbb{T}^d\times\mathbb{R}^d}$. Our general convergence result, namely Theorem 1.4.7, gives general conditions under which $(Q_t^\varepsilon)_{t\geq 0}$ converges in distribution to the unique solution of a particular martingale problem. The proof follows the usual pattern as the discussion in Section 1.3.

We now state the key assumptions that will imply convergence in distribution of the process $(Q_t^\varepsilon)_{t\geq 0}$ towards the solution of a martingale problem.

Assumption 1.4.3 (Generator of the process $(Q_t^\varepsilon, P_t^\varepsilon)$)**.** *There exists a linear operator L_ε acting on $C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$ which is the extended Markov generator of $(Q_t^\varepsilon, P_t^\varepsilon)_{t\geq 0}$ in the sense that, for all $f \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$, $L_\varepsilon f$ is locally bounded and*

$$t \mapsto M_t^\varepsilon(f) := f(Q_t^\varepsilon, P_t^\varepsilon) - f(Q_0^\varepsilon, P_0^\varepsilon) - \int_0^t L_\varepsilon f(Q_s^\varepsilon, P_s^\varepsilon) ds$$

is a $(\mathcal{F}_t^{Q^\varepsilon P^\varepsilon})_{t\geq 0}$-local martingale.

Assumption 1.4.4 (The limit process)**.** *There exists a linear operator L mapping $C^\infty(\mathbb{T}^d)$ to $C(\mathbb{T}^d)$ such that the martingale problem $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ is well-posed for any initial condition μ.*

Assumption 1.4.5 (Initial condition)**.** *The initial condition $(Law(Q_0^\varepsilon))_{\varepsilon>0}$ converge to a limit μ_0, when $\varepsilon \to 0$.*

Assumption 1.4.6 (Existence of perturbed test functions)**.** *For all $f \in C^\infty(\mathbb{T}^d)$, there exists a perturbed test function $f_\varepsilon \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$, such that for all T, the rest terms*

$$R_{1,t}^\varepsilon(f) := |f(Q_t^\varepsilon) - f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| \quad \text{and} \quad R_{2,t}^\varepsilon(f) := |Lf(Q_t^\varepsilon) - L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)|$$

satisfy the following bounds:

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\sup_{0\leq t\leq T} R_{1,t}^\varepsilon(f)\right) = 0, \tag{1.4.1}$$

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\int_0^T R_{2,t}^\varepsilon(f) dt\right) = 0. \tag{1.4.2}$$

We are now in position to state our main abstract result.

Theorem 1.4.7 (The general convergence theorem)**.** *Under the Assumptions 1.4.3, 1.4.4, 1.4.5, and 1.4.6, the family $\left(Law(Q_t^\varepsilon)_{t\geq 0}\right)_{\varepsilon>0}$ converges when $\varepsilon \to 0$ to the unique solution of martingale problem $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$.*

We will now discuss the main idea of the proof, following the structure that we have seen in the simple example. We first prove that the processes Q_t^ε are relatively compact in $\mathbb{D}_{\mathbb{T}^d}$; then we show that any possible limit must solve the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$.

Step 1: the proof of tightness. Following the same argument as in simple example, we just need to prove Lemma 1.3.2.

Let g_ε be the perturbed test function given by Assumption 1.4.6 and τ_n be an associated localizing sequence of stopping times. Taking the conditional expectation with respect to $\mathcal{F}_t^{Q^\varepsilon}$ and cancelling the martingale terms, we get

$$\begin{aligned}
&\left|\mathbb{E}\left[g(Q^\varepsilon_{(t+h)\wedge\tau_n}) - g(Q^\varepsilon_{t\wedge\tau_n})\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| \\
&\quad\leq \mathbb{E}\left[R^\varepsilon_{1,(t+h)\wedge\tau_n} + R^\varepsilon_{1,t\wedge\tau_n}\middle|\mathcal{F}_t^{Q^\varepsilon}\right] + \int_t^{t+h}\mathbb{E}\left[R^\varepsilon_{2,s}|\mathcal{F}_t^{Q^\varepsilon}\right]\mathbf{1}_{s\leq\tau_n}ds + \delta\sup_{q\in\mathbb{T}^d}|Lg(q)| \\
&\quad\leq 2\mathbb{E}\left[\sup_{t\in[0,T]}R^\varepsilon_{1,t}\middle|\mathcal{F}_t^{Q^\varepsilon}\right] + \int_0^T\mathbb{E}\left[R^\varepsilon_{2,s}|\mathcal{F}_t^{Q^\varepsilon}\right]ds + \delta\sup_{q\in\mathbb{T}^d}|Lg(q)|\,.
\end{aligned}$$

The right hand side does not depend on n any longer. On the left hand side, we apply dominated convergence for $n\to\infty$ to get

$$\left|\mathbb{E}\left[g(Q^\varepsilon_{(t+h)}) - g(Q^\varepsilon_t)\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| \leq \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g)\right]$$

for $\Gamma'_{\varepsilon,\delta}(g) = 2\sup_{[0,T]}R^\varepsilon_{1,t} + \int_0^T R^\varepsilon_{2,s}ds + \delta\|Lg\|_\infty$. The controls on the rest terms given by Assumption 1.4.6, and the continuity of Lg (Assumption 1.4.4) ensure that

$$\lim_{\delta\to0}\limsup_{\varepsilon\to0}\Gamma'_{\varepsilon,\delta}(g) = 0,$$

and the proof of tightness is concluded.

Step 2: identification of the limit. The same discussion with the simple example, except that when we get:

$$|I_\varepsilon| \leq \mathbb{E}\left[R^\varepsilon_{1,t_{p+1}}(f) + R^\varepsilon_{1,t_p}(f) + \int_{t_p}^{t_{p+1}}R^\varepsilon_{2,s}(f)\right]\|\varphi_1\|_\infty\ldots\|\varphi_p\|_\infty$$

by the same localization argument as in Step 1, we use the control on the rest term in Assumption 1.4.6, then imply that $I_\varepsilon\to0$, which proves the result of theorem.

1.4.2 From the abstract convergence result to the Langevin case

As the same, we start by constructing a perturbed test function: $f_\varepsilon\in C^\infty(\mathbb{T}^d\times\mathbb{R}^d)$. Let us look for f_ε in the following form (see [PSV77]):

$$f_\varepsilon(q,p) = f(q) + \varepsilon g_1(q,p) + \varepsilon^2 g_2(q,p). \tag{1.4.3}$$

Applying the generator L_ε, using the fact that f does not depend on p, and grouping terms with respect to powers of ε, we get

$$\begin{aligned} L_\varepsilon f_\varepsilon(q,p) = \frac{1}{\varepsilon}\left(p\cdot\nabla_q f - p\cdot\nabla_p g_1 + \frac{1}{\beta}\Delta_p g_1\right) \\ + \left(p\cdot\nabla_q g_1 - \nabla_q V_\varepsilon\cdot\nabla_p g_1 - p\cdot\nabla_p g_2 + \frac{1}{\beta}\Delta_p g_2\right) \\ + \varepsilon\left(p\cdot\nabla_q g_2 - \nabla_p g_2\cdot\nabla_q V_\varepsilon\right). \end{aligned} \tag{1.4.4}$$

As a consequence g_1 and g_2 should solve the following equations:

$$0 = p\cdot\nabla_q f - p\cdot\nabla_p g_1 + \frac{1}{\beta}\Delta_p g_1, \tag{1.4.5}$$

$$Lf(q) = p\cdot\nabla_q g_1 - \nabla_q V\cdot\nabla_p g_1 - p\cdot\nabla_p g_2 + \frac{1}{\beta}\Delta_p g_2.. \tag{1.4.6}$$

Therefore, we can define the perturbed test function by

$$f_\varepsilon(q,p) = f(q) + \varepsilon p\cdot\nabla_q f + \frac{1}{2}\varepsilon^2\nabla_q^2 f(p,p). \tag{1.4.7}$$

And then

$$\begin{aligned} &L_\varepsilon f_\varepsilon(q,p) - Lf(q) \\ &\quad = (\nabla_q V - \nabla_q V_\varepsilon)\cdot\nabla_q f + \frac{1}{2}\varepsilon\left(\nabla_q^3 f(q)(p,p,p) - \nabla_q^2 f\left(p,\nabla_q V_\varepsilon\right)\right). \end{aligned} \tag{1.4.8}$$

In the introduction, we are only giving ideas for the proof of the first rest term equation in Assumption 1.4.6 of Theorem 1.4.7 in the specific case of Langevin processes.

Since $f\in C^\infty(\mathbb{T}^d)$, there exists a $C_f = \max\left(\|\nabla f\|_\infty, \|\nabla^2 f\|_\infty\right)$ such that for all (q,p) and all $\delta\in(0,1/2)$

$$\begin{aligned} |f_\varepsilon(q,p) - f(q)| &= \varepsilon\,|p\cdot\nabla_q f(q)| + \frac{1}{2}\varepsilon^2\left|\nabla_q^2 f(q)\cdot(p,p)\right| \\ &\le C_f(\varepsilon\,|p| + \varepsilon^2\,|p|^2) \\ &\le \delta C_f + \frac{1}{\delta}C_f\varepsilon^2\,|p|^2, \end{aligned}$$

where we have used that for any $\delta>0$, $\varepsilon\,|p| \le \frac{1}{2}\delta + \frac{1}{2}\varepsilon^2\,|p|^2/\delta$. Therefore

$$\mathbb{E}\left[\sup_{t\in[0,T]} R^\varepsilon_{1,t}(f)\right] \le \delta C_f + \frac{1}{\delta}C_f\varepsilon^2\mathbb{E}\left[\sup_{t\in[0,T]}|P^\varepsilon_t|^2\right].$$

The proof is therefore reduced to establishing moment bounds. We will prove the following results:

Lemma 1.4.8 (Propagation of moments)**.** *Let $k \in \mathbb{R}$ with $k \geq 1$. There is a numerical constant $C_{\alpha,\beta,\|V\|_\infty}$ depending only on $k \geq 1$, $\beta > 0$ and $\|V\|_\infty$ such that for any $\varepsilon > 0$*

$$\sup_{t\geq 0} \mathbb{E}\left[|P_t^\varepsilon|^{2k}\right] \leq C_{k,\beta,\|V\|_\infty} \left(\mathbb{E}\left[|P_0^\varepsilon|^{2k}\right] + 1\right). \tag{1.4.9}$$

Lemma 1.4.9 (Moment of supremum)**.** *There is a numerical constant $C_{\beta,\|V\|_\infty,T}$ depending only on $\beta > 0, \|V\|_\infty$, and $T > 0$ such that for any $\varepsilon \in (0,1)$,*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |P_t^\varepsilon|^2\right] \leq \mathbb{E}\left[|P_0^\varepsilon|^2\right] + \frac{1}{\varepsilon} C_{\beta,\|V\|_\infty,T} \left(\mathbb{E}\left[|P_0^\varepsilon|^2\right] + 1\right)^{1/2}. \tag{1.4.10}$$

In particular, if $\lim_{\varepsilon\to 0} \varepsilon^2 \mathbb{E}\left[|P_0^\varepsilon|^2\right] = 0$, then

$$\lim_{\varepsilon\to 0} \varepsilon^2 \mathbb{E}\left[\sup_{0\leq t\leq T} |P_t^\varepsilon|^2\right] = 0.$$

Admitting these results for a moment, it is easy to conclude :
by assumption, $\lim_{\varepsilon\to 0} \varepsilon \mathbb{E}\left[|P_0^\varepsilon|^3\right] = 0$, which implies by Jensen's inequality that

$$\lim_{\varepsilon\to 0} \varepsilon \mathbb{E}\left[|P_0^\varepsilon|^2\right]^{3/2} = 0,$$

hence obviously $\lim_{\varepsilon\to 0} \varepsilon^2 \mathbb{E}\left[|P_0^\varepsilon|^2\right] = 0$.
From the key Lemma 1.4.9 we get that $\lim_{\varepsilon\to 0} \varepsilon^2 \mathbb{E}\left[\sup_{t\in[0,T]} |P_t^\varepsilon|^2\right] = 0$, hence

$$\limsup_{\varepsilon\to 0} \mathbb{E}\left[\sup_{t\in[0,T]} R_{1,t}^\varepsilon(f)\right] \leq \delta C_f,$$

which proves first limit in Assumption 1.4.6 since $\delta \in (0, 1/2)$ is arbitrary.

1.4.3 Proving the moment bounds

We now come back to the proofs of the moment bounds(Lemmas 1.4.8 and 1.4.9). It will prove useful to work with the Hamiltonian of the system rather than directly with P_t^ε. For any continuous $V : \mathbb{T}^d \to \mathbb{R}$ we denote by $\mathrm{osc}(V)$ the oscillation defined by

$$\mathrm{osc}(V) = \max V - \min V.$$

For convenience's sake we assume without loss of generality that $0 \leq V_\varepsilon(q) \leq \mathrm{osc}(V_\varepsilon)$. We will also write the Hamiltonian by $H_t^\varepsilon := H^\varepsilon\left(Q_t^\varepsilon, P_t^\varepsilon\right)$ for simplicity.

Proof of Lemma 1.4.8. Let us first discuss how to prove Lemma 1.4.8; for simplicity we

only consider here the case $k = 1$. By Itô's formula, for any smooth function $(t, h) \to \varphi(t, h)$, we thus get:

$$d\left(e^{\alpha t} H_t^\varepsilon\right) = \left(H_t^\varepsilon\right)\left(\alpha H_t^\varepsilon - \frac{2}{\varepsilon^2} H_t^\varepsilon + \frac{2}{\varepsilon^2} V_\varepsilon(Q_t^\varepsilon) + \frac{1}{\varepsilon^2 \beta}\right) e^{\alpha t} dt + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} H_t^\varepsilon P_t^\varepsilon e^{\alpha t} dW_t.$$

The choice $\alpha = 2/\varepsilon^2$ cancels the higher order term in the first bracket. We integrate in time, multiply by $e^{-\alpha t}$ and regroup the finite variation terms to get:

$$H_t^\varepsilon = H_0^\varepsilon + \int_0^t \left(\frac{2}{\varepsilon^2} V(Q_t^\varepsilon) + \frac{1}{\varepsilon^2 \beta}\right) e^{-\alpha(t-s)} ds + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t H_s^\varepsilon P_s^\varepsilon e^{-\alpha(t-s)} dW_s.$$

Since $(1/2)\left|P_s^\varepsilon\right|^2 \leq H_s^\varepsilon \leq (1/2)\left|P_s^\varepsilon\right|^2 + \operatorname{osc}(V_\varepsilon)$,

$$H_t^\varepsilon \leq H_0^\varepsilon + \left(1 - e^{-\alpha t}\right)\left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t P_s^\varepsilon e^{-\alpha(t-s)} dW_s. \tag{1.4.11}$$

To deal with the unboundedness of the momentum P, we define the following stopping times:

$$\tau_n := \inf\{t : |P_t| = n\}.$$

So when $s \leq \tau_n$, we have $|P_s^\varepsilon| \leq n$ and $H_s^\varepsilon \leq \left(\operatorname{osc}(V_\varepsilon) + \frac{n^2}{2}\right)$. This entails that $t \mapsto \int_0^{t \wedge \tau_n} P_s^\varepsilon dW_s$ is martingale. Then we get

$$\mathbb{E}\left[H_{t \wedge \tau_n}^\varepsilon\right] \leq \mathbb{E}\left[H_0^\varepsilon\right] + \left(1 - e^{-\alpha t}\right)\left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right).$$

Sending n to infinity, we apply Fatou's lemma to get

$$\mathbb{E}\left[H_t^\varepsilon\right] \leq \mathbb{E}\left[H_0^\varepsilon\right] + \left(1 - e^{-\alpha t}\right)\left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right).$$

Thus there exists a $C(M, \beta)$ such that for all ε,

$$\sup_t \mathbb{E}\left[H_t^\varepsilon\right] \leq C(M, \beta)\left(1 + \mathbb{E}\left[H_0^\varepsilon\right]\right), \tag{1.4.12}$$

concluding the proof. □

Proof of Lemma 1.4.9. Finally, the control on suprema of Lemma 1.4.9 may be proved along the following lines.

Recalling (1.4.11) for $\alpha = 2/\varepsilon^2$:

$$H_t^\varepsilon \leq H_0^\varepsilon + \left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s. \tag{1.4.13}$$

Define $M_t = \int_0^t P_s^\varepsilon dW_s$ and integrate by parts:

$$\begin{aligned}\left|\int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s\right| &= \left|\int_0^t e^{-\alpha(t-s)} dM_s\right| = \left|M_t - \alpha\int_0^t e^{-\alpha(t-s)} M_s ds\right| \\ &\leq |M_t| + \sup_{s\in[0,t]} |M_s| \\ &\leq 2\sup_{s\in[0,T]} |M_s|\,.\end{aligned}$$

Plugging this in (1.4.13) yields

$$\sup_{t\in[0,T]} H_t \leq H_0^\varepsilon + \left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} 2\sup_{t\in[0,T]} |M_t|\,. \tag{1.4.14}$$

By Doob's martingale maximal inequality, Itô's isometry and the bound (1.4.12) we get

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq t\leq T} |M_t^\varepsilon|^2\right] \leq 4\mathbb{E}\left[|M_T^\varepsilon|^2\right] = 4\mathbb{E}\left[\left|\int_0^T P_s^\varepsilon dW_s\right|^2\right] &= 4\mathbb{E}\left[\int_0^T (P_s^\varepsilon)^2 ds\right] \\ &\leq 8T\left(\operatorname{osc}(V_\varepsilon) + \sup_{t\in[0,T]} \mathbb{E}\left[H_t^\varepsilon\right]\right) \\ &\leq 8T\left(2\operatorname{osc}(V_\varepsilon) + \mathbb{E}\left[H_0^\varepsilon\right] + \frac{1}{\beta}\right).\end{aligned}$$

Injecting this in (1.4.14) and applying Cauchy–Schwarz inequality yields

$$\mathbb{E}\left[\sup_{t\in[0,T]} H_t\right] \leq \mathbb{E}\left[H_0^\varepsilon\right] + \left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + 8\frac{\sqrt{T\beta^{-1}}}{\varepsilon}\left(2\operatorname{osc}(V_\varepsilon) + \mathbb{E}\left[H_0^\varepsilon\right] + \frac{1}{\beta}\right)^{1/2}, \tag{1.4.15}$$

concluding the proof of (1.4.10). □

Chapter 2

Variance reduction by optimal reweighting of samples

2.1 Introduction on variance reduction

Let (X, Y) be a couple of random variables, and suppose that we are interested in computing the expected value $\mathbb{E}[\phi(Y)]$ - or more generally $\mathbb{E}[\phi(X, Y)]$ - for each ϕ in some class of test functions.

Since the distribution of $\phi(X, Y)$ is most often impossible to obtain in closed analytic form, a classical approach is to resort to Monte-Carlo integration: given an iid sample $(\mathbf{X}; \mathbf{Y}) = (X_1, ..., X_N; Y_1, ..., Y_N)$, the usual "naïve" Monte Carlo estimator is

$$\mathbf{\Phi}_{MC}(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \sum_{n=1}^{N} \phi(X_n, Y_n). \tag{2.1.1}$$

This estimator is unbiased, and its mean square error is given by its variance:

$$\mathrm{MSE}(\mathbf{\Phi}_{MC}) := \mathbb{E}\left[(\mathbf{\Phi}_{MC}(\mathbf{X}, \mathbf{Y}) - \mathbb{E}[\phi(X, Y)])^2\right] = \frac{1}{N} \mathrm{Var}(\phi(X, Y)).$$

The behaviour in N is inescapable and given by the CLT; however, over the years, many *variance reduction* techniques have been devised to reduce the constant multiplicative factor, using various kinds of additional hypotheses on the couple (X, Y). For a general overview of these techniques, see for example the survey paper of Glynn [Gly94], or the book [Ros13]. We introduce in this thesis a new technique for reducing variance, which can be seen as a variation on the classical post-stratification method, except that we do not have to fix strata. The method is based on two assumptions on the distribution of the couple (X, Y). The first assumption (see in Assumption 4.1.1): the distribution of the

first marginal X is exactly known:

$$X \sim \gamma := \mathcal{N}(0,1), \tag{2.1.2}$$

the standard Gaussian distribution, is essential to our method.

We note that we could easily accommodate other distributions than the standard Gaussian, which we consider for simplicity; the main point is that we know the distribution of X. Before introducing the second assumption, let us first recall a classical decomposition of variance. If we denote by

$$M_\phi(X) := \mathbb{E}\left[\phi(X,Y)|X\right], \qquad V_\phi(X) := \mathbb{E}\left[\left(\phi(X,Y) - M_\phi(X)\right)^2 \middle| X\right],$$

the mean and variance of $\phi(X,Y)$ conditionally on X, then the variance of $\phi(X,Y)$ may be rewritten as the sum of the expected conditional variance and the variance of the conditional expectation:

$$\mathrm{Var}\left(\phi(X,Y)\right) = \mathrm{Var}\left(M_\phi(X)\right) + \mathbb{E}\left[V_\phi(X)\right]. \tag{2.1.3}$$

The mean square error then reads:

$$\mathrm{MSE}(\mathbf{\Phi}_{MC}) = \frac{1}{N}\mathbb{E}\left[V_\Phi(X)\right] + \frac{1}{N}\,\mathrm{Var}(M_\Phi(X)). \tag{2.1.4}$$

We now state informally the second assumption (see Assumption 4.1.2): for the considered test function ϕ, the (random) conditional variance $V_\phi(X)$ is sufficiently 'small' as compared to the variance of the conditional expectation $\mathrm{Var}(M_\phi(X))$. Typically, our theoretical results will hold under the almost sure condition

$$V_\phi(X) \leq \mathrm{Var}(M_\phi(X))/c, \quad c \quad \text{large enough.} \tag{2.1.5}$$

Intuitively, (2.1.5) ensures that the majority of the variance of $\phi(X,Y)$ is due to X; since the distribution of X is simple and explicit, substantial variance reduction can be expected.

For the sake of comparison, let us first discuss how classical variance reduction methods can be applied in this setting.

2.1.1 Control variates

A control variate is a computable function ψ of X such that the mean square error of $\mathbb{E}\left[\left(\phi(X,Y) - \psi(X)\right)^2\right]$ is as small as possible (see [Owe13, Gla13] for a general introduction). Recent works have studied various techniques to find an optimized ψ using basic functions and a Monte Carlo approach (see for instance [Jou09, PS18]). The associated

estimator is then

$$\mathbf{\Phi}_{CV}(\mathbf{X},\mathbf{Y}) = 1/N\left(\sum_{n=1}^{N}\phi(X_n,Y_n) - \Psi(X_n)\right) + \int \Psi d\gamma,$$

whose mean square error (identical to variance since it is unbiased), satisfies

$$\begin{aligned}\mathrm{MSE}(\mathbf{\Phi}_{CV}(\mathbf{X},\mathbf{Y})) &= \mathrm{Var}(\mathbf{\Phi}_{CV}(\mathbf{X},\mathbf{Y}))\\ &= \frac{1}{N}\mathbb{E}\left[\left(\phi(X,Y) - \Psi(X) + \int \Psi d\gamma - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right].\end{aligned}$$

By the characteristic property of the conditional expectation

$$\mathbb{E}\left[\phi(X,Y)|X\right] = \underset{\tilde{\Psi}}{\mathrm{argmin}}\,\mathbb{E}\left[\left(\phi(X,Y) - \tilde{\Psi}(X)\right)^2\right],$$

this variance is minimal for

$$\Psi(X) = M_\phi(X) = \mathbb{E}\left[\phi(X,Y)|X\right],$$

in which case

$$\mathrm{Var}\left(\mathbf{\Phi}_{CV}^{\mathrm{opt}}(\mathbf{X},\mathbf{Y})\right) = \frac{1}{N}\mathbb{E}\left[V_\phi(X)\right] < \mathrm{MSE}(\mathbf{\Phi}_{MC}).$$

Let us note that, if a good control variate, close to the conditional expectation $\psi(X) \sim \mathbb{E}\left[\phi(X,Y)|X\right]$ is available, and if we try to apply our method to $\tilde{\phi}(X,Y) = \phi(X,Y) - \psi(X)$, $M_{\tilde{\phi}}$ is identically 0, so our second assumption, for example in its form (2.1.5), can not hold for $\tilde{\phi}$. In the numerical experiment done in Section 4.6 — see also Figure 4.2 — we compare the reweighting method with a natural affine control variate $\psi(X) = aX + b$ which is not sufficient to approximate correctly $\mathbb{E}\left[\phi(X,Y)|X\right]$; interestingly the reweighting method is able to overcome this issue in a generic way, without specific analytic approximation of $\mathbb{E}\left[\phi(X,Y)|X\right]$ contrary to what is required to improve the control variate.

2.1.2 Conditioning

This method uses the well known fact that conditioning reduces the variance.

We can estimate $\mathbb{E}\left[\phi(X,Y)\right]$ by

$$\mathbf{\Phi}_{\mathrm{cond}}(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[\phi(X,Y)|X = X_n\right],$$

where $(X_i)_{i\geq 1}$ are independently sampled from the distribution of X.

We find

$$\mathrm{MSE}(\mathbf{\Phi}_{\mathrm{cond}}) = \mathrm{Var}(\mathbf{\Phi}_{\mathrm{cond}}) = \mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N} M_\phi(X_i)\right) = \frac{1}{N}\mathrm{Var}\left(M_\phi(X)\right) < \mathrm{MSE}(\mathbf{\Phi}_{MC}).$$

It is immediately clear that *conditional Monte-Carlo* can not have higher variance than ordinary Monte-Carlo. The justification for the method is that $\mathbb{E}\left[\phi(X,Y)\right] = \mathbb{E}\left[M_\phi(X)\right]$, the function $M_\phi(\cdot)$ gives the conditional mean of Y and then we complete the job by Monte-Carlo. The method is called *conditioning* or *conditional Monte-Carlo*. The main requirement for conditioning is that we must be able to compute analytically $M_\phi(X)$.

Note that in the present context, we simply have to compute a one dimensional average with respect to γ, which can be done more efficiently with deterministic integration methods.

2.1.3 Stratification

The idea in stratified sampling is to split up the domain $\mathcal{S}$ of X into separate regions, on which we wish to calculate an expectation or integral into strata, take a sample of points from each such region, and combine the results to estimate $\mathbb{E}\left[\phi(X,Y)\right]$. Intuitively, if each region gets its "fair share" of points then we should get a better result than when they are all sampled randomly.

Let $(\mathcal{S}_i, 1 \leq i \leq K)$ be a partition of X. Then $\mathbb{E}\left[\phi(X,Y)\right]$ can be expressed as

$$\begin{aligned}\mathbb{E}\left[\phi(X,Y)\right] &= \sum_{i=1}^{K} \mathbb{E}\left[\mathbf{1}_{X\in\mathcal{S}_i}\phi(X,Y)\right] \\ &= \sum_{i=1}^{K} \mathbb{E}\left[\phi(X,Y)|X\in\mathcal{S}_i\right]\mathbb{P}\left[X\in\mathcal{S}_i\right].\end{aligned}$$

Note that $\mathbb{E}\left[\phi(X,Y)|X\in\mathcal{S}_i\right]$ can be interpreted as $\mathbb{E}\left[\phi(X^i,Y^i)\right]$ where (X^i,Y^i) is a random variable whose law is the law of (X,Y) conditioned by X belonging to $\mathcal{S}_i$.

When the numbers $p_i := \mathbb{P}\left[X\in\mathcal{S}_i\right]$ can be explicitly computed, the problem reduces to an estimation of $M_\phi^i(X) := \mathbb{E}\left[\phi(X,Y)|X\in\mathcal{S}_i\right]$. This can be done by "naïve" Monte Carlo as follows

$$\hat{M}_\phi^i(X) = \frac{1}{n_i}\left(\phi(X_1^i,Y_1^i) + ... + \phi(X_{n_i}^i,Y_{n_i}^i)\right),$$

where $\left(X_1^i,...,X_{n_i}^i\right)$ are independent copies of X^i, and n_i is the number of sample points $X\in\mathcal{S}_i$. Then an estimator $\mathbf{\Phi}_{\mathrm{stra}}(\mathbf{X},\mathbf{Y})$ of $\mathbb{E}\left[\phi(X,Y)\right]$ is

$$\mathbf{\Phi}_{\mathrm{stra}}(\mathbf{X},\mathbf{Y}) = \sum_{i=1}^{K} p_i \hat{M}_\phi^i(X).$$

The variance of this stratified estimator is given by:

$$\mathrm{Var}\left(\mathbf{\Phi}_{\mathrm{stra}}(\mathbf{X},\mathbf{Y})\right) = \sum_{i=1}^{K} p_i^2 \frac{\mathrm{Var}\left(M_\phi^i(X)\right)}{n_i}. \tag{2.1.6}$$

Fix the total number of simulations $\sum_{i=1}^{K} n_i = N$. The optimal choice of n_i minimising the variance above is explicit and given by:

$$n_{\mathrm{opt}}(i) = N \frac{p_i \left(\mathrm{Var}\left(M_\phi^i(X)\right)\right)^{\frac{1}{2}}}{\sum_{i=1}^{K} p_i \left(\mathrm{Var}\left(M_\phi^i(X)\right)\right)^{\frac{1}{2}}},$$

where $n_{\mathrm{opt}}(i)$ has no reason to be an integer and must be rounded in practice.

For this optimal choice of allocations, the variance of $\mathbf{\Phi}_{\mathrm{stra}}^{\mathrm{opt}}(\mathbf{X},\mathbf{Y})$ is given by

$$\begin{aligned}\mathrm{Var}\left(\mathbf{\Phi}_{\mathrm{stra}}^{\mathrm{opt}}(\mathbf{X},\mathbf{Y})\right) &= \frac{1}{N}\left(\sum_{i=1}^{K} p_i \left(\mathrm{Var}\left(M_\phi^i(X)\right)\right)^{\frac{1}{2}}\right)^2 \\ &\leq \frac{1}{N}\sum_{i=1}^{K} p_i \,\mathrm{Var}\left(M_\phi^i(X)\right),\end{aligned}$$

where in the last line we used Jensen inequality.

Note that this variance is smaller than the one obtained without stratification (with a "naïve Monte Carlo" approach). Indeed,

$$\begin{aligned}\mathrm{Var}\left(\phi(X,Y)\right) &= \mathbb{E}\left[(\phi(X,Y))^2\right] - (M_\phi(X))^2 \\ &= \sum_{i=1}^{K} p_i \mathbb{E}\left[(\phi(X,Y))^2 \,|X \in \mathcal{S}_i\right] - \left(\sum_{i=1}^{K} p_i M_\phi^i(X)\right)^2 \\ &= \sum_{i=1}^{K} p_i \,\mathrm{Var}\left(M_\phi^i(X)\right) + \sum_{i=1}^{K} p_i \left(M_\phi^i(X)\right)^2 - \left(\sum_{i=1}^{K} p_i M_\phi^i(X)\right)^2.\end{aligned}$$

By the convexity inequality for x^2, we have $\sum_{i=1}^{K} p_i \left(M_\phi^i(X)\right)^2 \geq \left(\sum_{i=1}^{K} p_i M_\phi^i(X)\right)^2$, then the inequality

$$\mathrm{Var}\left(\phi(X,Y)\right) \geq \sum_{i=1}^{K} p_i \,\mathrm{Var}\left(M_\phi^i(X)\right) \geq N \,\mathrm{Var}\left(\mathbf{\Phi}_{\mathrm{stra}}^{\mathrm{opt}}(\mathbf{X},\mathbf{Y})\right)$$

follows.

Remark 2.1.1. • *Formally, for a very large set of strata, the stratification method amounts to compute observables with respect to X on a deterministic grid (the strata) and then to simulate averages conditional on X using Monte Carlo. Hence*

the variance of the method is approximately the average of the conditional variance – with respect to X – of $\phi(X,Y)$.

- *The optimal n_i are not explicit (they depend on $\mathrm{Var}\left(M_\phi^i(X)\right)$);*
- *A bad choice of n_i could increase the variance;*
- *A safe choice is proportional allocation $n_i = Np_i$ (i.e. the n_i are only proportional to the known p_i, and not reweighted by the standard deviation in the strata as in the optimal choice). In this case,*

$$\mathrm{Var}\left(\mathbf{\Phi}_{strat}(\mathbf{X},\mathbf{Y})\right) = \frac{1}{N}\sum_{i=1}^{K} p_i \,\mathrm{Var}\left(M_\phi^i(X)\right) \leq \frac{1}{N}\,\mathrm{Var}\left(\phi(X,Y)\right).$$

Note that in practice, $n_i = Np_i$ may not be a integer and must be rounded.

Remark 2.1.2. *The method requires to be able to simulate (X,Y), conditional on X being in a strata, and to know the p_i. The method is also clearly useless when $M_\phi^i(X) = M_\phi^j(X)$ (for any $i,j = 1,...,K$) with fixed p_i, since no variance reduction is achieved.*

2.1.4 Post-stratification

As we have seen, the stratification strategy is guaranteed to reduce variance, but may only be applied if we know how to simulate (X,Y) conditional on $X \in \mathcal{S}_i$.

In post-stratification we sample X_i and assign X_i to their strata after the fact. If N is large, by the LLN (Law of Large Numbers), we have $N_i/N \to p_i$, thus $Nw_i \to 1$. If a stratum is "undersampled" in the sense that $N_i/N < p_i$, the idea of post-stratification is to compensate by assigning a greater weight to the X_i falling in this stratum, and vice versa for "oversampled" strata.

In that context, the post-stratification weights are defined by

$$w_i(\mathbf{X}) = \frac{p_i}{N_i}, \tag{2.1.7}$$

where $p_i = \mathbb{P}\left[X \in \mathcal{S}_i\right]$ and $N_i = \sum_{n=1}^{N} \mathbf{1}_{X_n \in \mathcal{S}_i}$.

Then an estimator $\mathbf{\Phi}_{\text{post-stra}}(\mathbf{X},\mathbf{Y})$ of $M_\phi(X)$ is

$$\mathbf{\Phi}_{\text{post-stra}}(\mathbf{X},\mathbf{Y}) = \sum_{n=1}^{N} w_{I_n}(\mathbf{X})\phi(X_n,Y_n), \tag{2.1.8}$$

where $I_n \in \mathbb{N}$ is such that $X_n \in \mathcal{S}_{I_n}$.

We will see how the post-stratification can be seen as a special case of our method in Section 2.2.2.

Remark 2.1.3. *The main difference with respect to stratification is that* $(N_i)_{i=1,...,K}$ *are now random. As compared to stratification, we don't assume we are able to simulate conditionally on* X*, nor do we assume we are able to simulate* X *conditional on being in a strata. Another difference is that the variance of* $\mathbf{\Phi}_{\text{post-stra}}$ *is not smaller than* $\mathbf{\Phi}_{MC}$ *in general.*

2.2 The reweighting idea

When we look at the weighted empirical estimator (2.1.8), we would like to be more general and find a good way to re-weight the samples. Since we do not have the liberty of choosing the X_i, but we know exactly their distribution γ, our main idea is to use the samples $\mathbf{X} = (X_1, ..., X_N)$ to devise random weights $(w_n(\mathbf{X}))_{1 \leq n \leq N}$ such that the empirical measure $\sum_n w_n(\mathbf{X})\delta_{X_i}$ is "as close as possible" to the true distribution γ. More precisely, for some distance dist () between distributions — the choice of which will be discussed below — we look for solutions of the minimization problem:

$$\text{minimize: } \text{dist}\left(\gamma, \sum_{n=1}^{N} w_n \delta_{X_n}\right) \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases} \tag{2.2.1}$$

This minimization problem typically admits a unique solution $(w_1(\mathbf{X}), ..., w_N(\mathbf{X}))$, which can be used instead of the naïve uniform weights $(1/N)$ to estimate $\mathbb{E}[\phi(X, Y)]$ by:

$$\mathbf{\Phi}_W(\mathbf{X}, \mathbf{Y}) = \sum_{n=1}^{N} w_n(\mathbf{X})\phi(X_n, Y_n).$$

The goal of the article presented in Chapter 4 is to show, both theoretically and empirically, that it indeed succeeds in reducing the variance with respect to the naïve Monte Carlo method.

2.2.1 Decomposition of the mean square error

The mean square error of our estimator is given by:

$$\begin{aligned}
\text{MSE}(\mathbf{\Phi}_W) &:= \mathbb{E}\left[(\mathbf{\Phi}_W(\mathbf{X}, \mathbf{Y}) - \mathbb{E}[\phi(X, Y)])^2\right] \\
&= \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})\phi(X_n, Y_n) - \sum_n w_n(\mathbf{X})M_\phi(X_n) + \sum_n w_n(\mathbf{X})M_\phi(X_n) - \mathbb{E}[\phi(X, Y)]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})(\phi(X_n, Y_n) - M_\phi(X_n))\right)^2\right] + \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})M_\phi(X_n) - \mathbb{E}[\phi(X, Y)]\right)^2\right]
\end{aligned}$$

since the cross terms vanish because $\mathbb{E}\left[\phi(X_n, Y_n) - M_\phi(X_n)|\mathbf{X}\right] = 0$. In the first term, we expand the square, condition on $\mathbf{X}$ and use the conditional independence of the $(Y_n)_{n\geq 1}$; we rewrite the second term using the notation $\eta_N^\star = \sum_n w_n(\mathbf{X})\delta_{X_n}$ and get

$$\begin{aligned}\mathrm{MSE}(\mathbf{\Phi}_W) &= \mathbb{E}\left[\left(\mathbf{\Phi}_W(\mathbf{X}, \mathbf{Y}) - \mathbb{E}\left[\phi(X, Y)\right]\right)^2\right] \\ &= \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2 V_\phi(X_n)\right] + \mathbb{E}\left[\left(\int M_\phi(x) d\eta_N^\star(x) - \int M_\phi(x)\gamma(dx)\right)^2\right] \end{aligned} \tag{2.2.2}$$

Let us note here that for the naïve choice $w_n(\mathbf{X}) = 1/N$, Equation (2.2.2) reduces to the decomposition (2.1.4) in two terms of the same order $1/N$: the first term becoming $\frac{1}{N}\mathbb{E}\left[V_\phi(X)\right]$, the second being $\frac{1}{N}\mathrm{Var}\left(M_\phi(X)\right)$. By choosing weights that minimize the distance between the reweighted measure $\eta_N^\star$ and γ, our goal is to make the second term of (2.2.2) negligible; for this we pay a price by increasing the first term. This informal statement will be made precise below, see in particular Corollary 4.1.16 and Remark 2.4.2 below.

In order to give rigorous statements, we make two additional assumptions concerning the test function ϕ and the distance we will use:

- The distance $\mathrm{dist}\,()$ may be written in operator norm form

$$\mathrm{dist}\,(\eta, \gamma) = \sup_{f\in\mathcal{D}} |\eta(f) - \gamma(f)| \tag{2.2.3}$$

  where $\mathcal{D}$ is a set of functions (typically a unit ball of test functions).

- There exist two constants m_ϕ, v_ϕ such that:

  - The conditional mean $x \mapsto M_\phi(x)$ satisfies $\left(M_\phi(\cdot) - c\right)/m_\phi \in \mathcal{D}$ for some constant c, where $\mathcal{D}$ is the set of functions defined in the previous assumption. If $\mathcal{D}$ is the unit ball associated with a norm $\|\ \|$, the optimal m_ϕ is given by

$$m_\phi := \inf_{c\in\mathbb{R}} \left\| M_\phi - \int M_\phi d\gamma \right\|. \tag{2.2.4}$$

  - The conditional variance V_ϕ satisfies

$$V_\phi(X_n) \leq v_\phi \qquad a.s. \tag{2.2.5}$$

Assuming this, as an immediate consequence of (2.2.2), we get

$$\mathbb{E}\left[\left(\mathbf{\Phi}_W(\mathbf{X}, \mathbf{Y}) - \mathbb{E}\left[\phi(X, Y)\right]\right)^2\right] \leq v_\phi \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] + m_\phi^2 \mathbb{E}\left[\mathrm{dist}\,(\eta_N^\star, \gamma)^2\right]. \tag{2.2.6}$$

We consequently propose to define and compute the weights $(w_n(\mathbf{X}))_{1\leq n\leq N}$ according to

$$\text{minimize: } \operatorname{dist}\left(\gamma, \sum_{n=1}^{N} w_n \delta_{X_n}\right)^2 + \delta \sum_n w_n^2, \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1, \end{cases} \tag{2.2.7}$$

or more simply to (2.2.1) which is obtained from the former by taking $\delta = 0$.

Remark 2.2.1 (On the choice $\delta = 0$)**.** *Depending on the choice of the distance dist(), solving (2.2.7) for $\delta \neq 0$, instead of $\delta = 0$ can be almost free or quite costly numerically. Moreover it requires the tuning of another parameter δ. As a consequence, for simplicity and homogeneity, we will mainly focus on the choice $\delta = 0$. From the discussion above, this choice is formally appropriate for the limit case where observables ϕ satisfies*

$$\frac{v_\phi}{m_\phi^2} \ll 1.$$

This is a special limit case of (2.1.5).

2.2.2 Reinterpreting post-stratification as a special case

The present work may be interpreted as a generalization of the post-stratification method presented in Section 2.1.4 to continuous state spaces. In the framework of Section 2.1.4, post-stratification can be define by first choosing a finite partition of $\mathbb{R}$, given by K 'strata', for instance the K-quantiles $x_{1/2} < \ldots < x_{K-1/2}$ defined by $\int_{x_{k-1/2}}^{x_{k+1/2}} d\gamma = 1/K = p_i$.

In that context, then the post-stratification weights can be defined by

$$\begin{cases} w_n(\mathbf{X}) = \frac{1}{KB_n(\mathbf{X})}, \\ B_n(\mathbf{X}) = \operatorname{card}\left\{X_m \in \operatorname{strat}(X_n),\, 1 \leq m \leq N\right\}, \end{cases} \tag{2.2.8}$$

where in the above $\operatorname{strat}(X_n)$ is the interval $[x_{k-1/2}, x_{k+1/2}[$ containing X_n. The latter post-stratification weights are defined so that the sum of the weights of particles in a given stratum is constant an equal to $1/K$.

Let us denote by $\mathcal{D}_K$ the space of functions on $\mathbb{R}$ that are constant on the srata $[x_{k-1/2}, x_{k+1/2}[$, for $k = 1 \ldots K$; consider the semi-norm over finite measures

$$p_K(\mu) = \sup_{\psi \in \mathcal{D}_K,\, \|\psi\|_\infty \leq 1} \mu(\psi);$$

then the post-stratification weights (2.2.8) is the solution to the minimization problem

obtained by setting $\mathrm{dist}(\eta, \gamma) = p_K(\eta - \gamma)$ in (2.2.1) that is

$$\text{minimize: } p_K\left(\gamma - \sum_{n=1}^{N} w_n \delta_{X_n}\right), \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1, \end{cases} \tag{2.2.9}$$

that moreover minimize the variance of the weight $\sum_n w_n^2 - 1$.

The remainder of this chapter is organized as follows. In Section 2.3 , since we will consider two distances, we need to introduce a few tools to describe them properly. We will describe our main results in Section 2.4 and give a glimpse of the main proofs in Section 2.5. In Section 2.6 we will briefly describe the numerical experiments to compare our method with naïve Monte Carlo method. Finally, we make a conclusion in Section 2.7.

2.3 Two choices for the distance

To specify completely the algorithm, we need to choose an appropriate distance between probability measures to use in the minimization problem (2.2.1)-(2.2.7). Among the many possible choices, see e.g.[GS02] for a review, we focus on two choices.

2.3.1 The L^2 method

The first distance uses the Hilbert structure of the space $L^2(\gamma)$. On this space, a natural choice would be the χ^2 divergence. However, $\sum_n w_n \delta_{X_n}$ has no density with respect to γ, so we need to regularize it. In our setting, a natural tool for regularizing is the following Ornstein-Uhlenbeck semigroup.

Let P_t denotes the semigroup of probability transitions of the Ornstein–Uhlenbeck process solution to the SDE $dX_t = -X_t dt + \sqrt{2} B_t$ where B_t is a standard Brownian motion, that is $\mathbb{E}[f(X_t)|X_0 \sim \eta] = \int P_t f(x) d\eta(x)$ for all bounded continuous test function f and any probability measure η. Then it holds

$$P_t(x, dz) = k_t(x, z)\gamma(dz).$$

This kernel k_t has an explicit expression and a nice spectral decomposition, see Lemma 4.3.1 for details.

Let us denote by

$$h = 1 - e^{-2t} \quad (t > 0)$$

the variance of X_t so that h can be seen as the square of a bandwidth parameter. Now we define precisely the h-norm on signed-measures.

Let $\mathcal{M}$ be the set of signed measures on $\mathbb{R}$ with a finite total mass, and let

$$\mathcal{S} = \left\{ \nu \in \mathcal{M}, \int e^{\frac{y^2}{4}} |\eta| (dy) < +\infty \right\}.$$

For all ν in $\mathcal{S}$, it can be checked that the regularization νP_t admits a square integrable density with respect to the Gaussian measure, so that

$$\|\nu\|_h := \left\| \frac{d(\nu P_t)}{d\gamma} \right\|_{L^2(\gamma)} < \infty.$$

This expression defines a norm on $\mathcal{S}$, that satisfies nice properties (see Theorem 4.3.4). For instance $\|\nu\|_h$ admits the dual representation

$$\|\nu\|_h = \sup_{\|\varphi\|_{L^2(\gamma)} \leq 1} \int \varphi \frac{d(\nu P_t)}{d\gamma} d\gamma = \sup_{\|\varphi\|_{L^2(\gamma)} \leq 1} \int P_t \varphi d\nu. \tag{2.3.1}$$

Moreover,

$$\|\cdot\|_{h'} \leq \|\cdot\|_h, \quad \text{when} \quad h \leq h'.$$

Finally, if $(\eta_k)_{k\in\mathbb{N}}$ and η are probability measures in $\mathcal{S}$, and if $\|\eta_k - \eta\|_h \to 0$, then η_k converges weakly to η.

As a consequence, the set of considered test functions $\mathcal{D}$ is the unit ball associated with $\|\|_h$, and we obtain easily that in (2.2.4):

$$m_\phi = \left\| P_t^{-1} \left(M_\phi - \int M_\phi d\gamma \right) \right\|_{L^2(\gamma)}. \tag{2.3.2}$$

The decomposition of the kernel function – or the fact that P_t is a regularizing kernel – then shows that in order for m_ϕ to be bounded, M_ϕ must be a very regular function, at least smooth on $\mathbb{R}$.

Then we discuss the minimization problem (2.2.1) when dist () is the h-norm, first from a deterministic point of view. Let $\mathbf{x} = (x_1, ..., x_N)$ be a vector in $\mathbb{R}^N$. We want to solve the following minimization problem :

$$\text{minimize:} \quad \left\| \sum_n w_n \delta_{x_n} - \gamma \right\|_h^2 + \delta \sum_n w_n^2, \qquad \text{subject to:} \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases} \tag{2.3.3}$$

Let us denote by Ω the simplex $\{w = (w_1, ..., w_N) | w_i \geq 0, \ \sum_{n=1}^N w_n = 1\}$, and let $F(w) =$

$\|\sum_n w_n \delta_{x_n} - \gamma\|_h$. It can be checked that, F may be rewritten as follows:

$$\begin{aligned} F(w) &= \sum_{n,m}^{N} w_n w_m k_{2t}(x_n, x_m) - 2\sum_{n=1}^{N} w_n + 1 \\ &= w^T Q w - 1, \end{aligned}$$

where Q is the $N \times N$ matrix whose components are given by

$$Q_{n,m} = k_{2t}(x_n, x_m), \quad \text{for any } 1 \leq n, m \leq N. \tag{2.3.4}$$

Thereby, the minimization problem is reduced to the following quadratic problem over a convex set:

$$\text{minimize: } w^\top (Q + \delta \mathrm{Id}) w \qquad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases} . \tag{2.3.5}$$

Remark 2.3.1. *Now the minimization problem* (2.3.3) *turns out to be a quadratic programming convex optimization problem* (2.3.5)*, which can be solved using standard methods, typically with a cubic polynomial complexity in terms of the sample size* N*. Solving the case* $\delta \neq 0$ *is in fact easier than the case* $\delta = 0$ *since larger* δ *simply improve the conditioning of the symmetric matrix underlying the quadratic programming problem.*

Moreover, uniqueness holds. If the x_n are pairwise distinct, in that case Q is positive definite, and the minimization problem (2.3.5) has a unique solution even if $\delta = 0$. This holds in particular with probability one if $(x_n)_{1 \leq n \leq N} = (X_n)_{1 \leq n \leq N}$ are iid samples of γ.

2.3.2 The Wasserstein method

Assume that you are in charge of the transport of goods between producers and consumers, whose respective spatial distributions are modelled by probability measures. The farther producers and consumers are from each other, the more difficult will be, and we would like to summarize the degree of difficulty with just one quantity. For this purpose, it is natural to consider the optimal transport cost between the two measures:

$$C(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int c(x, y) d\pi(x, y), \tag{2.3.6}$$

where $c(x, y)$ is the cost for transporting one unit of mass from x to y. For more details, see in [Vil08].

One can think of (2.3.6) as a kind of distance between μ and ν, strictly speaking, satisfy the

axioms of a distance function. Here the choice we investigate is the Wasserstein distance $\mathcal{W}_1$, defined classically as follows: Let (E, d) be a Polish metric space. For any two probability measures η_1, η_2 on E, the Wasserstein distance between η_1 and η_2 is defined by the formula

$$\begin{aligned}\mathcal{W}_1(\eta_1, \eta_2) &= \Big(\inf_{\pi \in \Pi} \int_E d(x_1, x_2) d\pi(x_1, x_2) \Big) \\ &= \inf \Big\{ \mathbb{E}\left[d(X_1, X_2) \right], \quad \text{Law}(X_1) = \eta_1, \text{Law}(X_2) = \eta_2 \Big\},\end{aligned} \tag{2.3.7}$$

where Π is the set of all couplings of η_1 and η_2.

Kantorovitch duality (see [Vil08]) implies that the latter distance is in fact an operator norm of the form

$$\mathcal{W}_1(\eta_1, \eta_2) = \|\eta_1 - \eta_2\|_{\text{Lip}} = \sup_{\|f\|_{\text{Lip}} \leq 1} \eta_1(f) - \eta_2(f),$$

where $\|f\|_{\text{Lip}} = \sup_{x,y} \frac{f(x)-f(y)}{d(x,y)}$ is the Lipschitz norm on the space of functions defined up to an additive constant. $\mathcal{D}$ in (2.2.3) is thus the the unit ball associated with Lipschitz norm. As a consequence, we obtain easily that in (2.2.4)

$$m_\phi = \|M_\phi\|_{\text{Lip}}$$

so that M_ϕ need only to be Lipschitz for m_ϕ to be bounded.

Then we discuss the minimization problem (2.2.1) when dist () is the $\mathcal{W}_1$-norm, first from a deterministic point of view. Let $\mathbf{x} = (x_1, ..., x_N)$ be a vector in $\mathbb{R}^N$. We want to solve the following minimization problem :

$$\text{minimize: } \left\| \sum_n w_n \delta_{x_n} - \gamma \right\|_{\mathcal{W}_1}^2 + \delta \sum_n w_n^2, \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases}$$

In Proposition 4.4.1, we will detail the following fact.

Let $\mathbf{x} = (x_1, ..., x_N)$ be a set of N distinct points in $\mathbb{R}$, let $(x_{(1)} < x_{(2)} \cdots < x_{(N)}$ be their ordered relabelling, and let

$$\begin{cases} y_n := \frac{1}{2}(x_{(n)} + x_{(n+1)}), 1 \leq n \leq N-1 \\ y_0 = -\infty, \\ y_N = \infty. \end{cases}$$

For $w = (w_1, ..., w_N)$ let $F(w)$ be the cost

$$F(w) = \mathcal{W}_1 \left(\sum_{n=1}^{N} w_n \delta_{x_n}, \gamma \right).$$

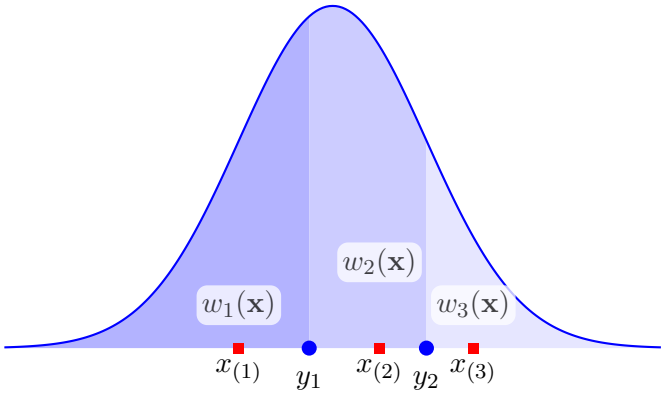
The minimization problem:

$$\text{minimize: } F(w) \quad \text{subject to: } w \in \Omega$$

has a unique solution $w(\mathbf{x}) = (w_1(\mathbf{x}), ..., w_N(\mathbf{x}))$, given by

$$w_n(\mathbf{x}) = \int_{y_{m-1}}^{y_m} \gamma(dz), \tag{2.3.8}$$

where m is the unique integer such that $x_n = x_{(m)}$.



Given the sample $(\mathbf{x})$, the optimal Wasserstein weights are obtained by computing the middle points $y_n = (x_{(n)} + x_{(n+1)})/2$, and letting $w_n = \gamma([y_{n-1}, y_n])$.

Figure 2.1: The optimal weights $w(\mathbf{x})$.

Thus, at least in dimension 1 for the choice $\text{dist}(x_1, x_2) = |x_1 - x_2|$, the Wasserstein distance leads to an explicit formula for the optimal weights $(w_n(X))_{1 \leq n \leq N}$, that can be computed with an complexity proportional to the sample size N. This leads to faster algorithms and more explicit bounds on the mean square error (see in Section 2.4) as compared to the L^2 case of the last section. However, for this optimal transport method, solving the case $\delta = 0$ is non explicit and thus harder (although still a convex optimization problem).

2.4 The main results

Recall that $\mathbf{X} = (X_1, ..., X_N)$ is an i.i.d. $\mathcal{N}(0,1)$ sequence in $\mathbb{R}$. We denote by

$$\overline{\eta}_N = \frac{1}{N}\sum_n \delta_{X_n}.$$

the empirical measure of the sample $\mathbf{X}$. The reweighted measure $\sum_n w_n(\mathbf{X})\delta_{X_n}$ will be denoted by:

$\eta^\star_{h,N}$, if the $w_n(\mathbf{X})$ solve (2.2.7) for the L^2 distance with parameter h and δ;
$\eta^\star_{\text{Wass},N}$, if the $w_n(\mathbf{X})$ solve (2.2.1) for the Wasserstein distance.

Let us recall (2.2.6):

$$\text{MSE}(\mathbf{\Phi}_W) \leq v_\phi \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] + m_\phi^2 \mathbb{E}\left[\text{dist}\left(\eta^\star_N, \gamma\right)^2\right].$$

We first focus on results on these optimally reweighted measures, shedding light on the behaviour of this bound and especially the second term in it.

We start by the L^2 minimization method. First remark that

Remark 2.4.1 (Potential improvement in the L^2 method)**.** *Assume that $V_\phi(x)$ is bounded above by a constant v_ϕ and set $\delta = v_\phi/m_\phi^2$. then the L^2 method is by construction not worse than the naïve Monte Carlo approach in terms of the upper bound in (2.2.6):*

$$\mathbb{E}\left[\left(\mathbf{\Phi}_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right] \leq v_\phi \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] + m_\phi^2 \mathbb{E}\left[dist\left(\eta^\star_N, \gamma\right)^2\right]. \tag{2.4.1}$$

But it is not necessarily better in terms of the MSE($\mathbf{\Phi}_W(X,Y)$) given in (2.2.2)

The following partial theoretical result, as well as numerical results below, suggests that it is however the case (i.e. variance reduction do occur) when $V_\phi(X)/\operatorname{Var}(M_\phi)$ is small enough (i.e. when Assumption 4.1.2 is satisfied). The behaviour of the method is given in the following.

Theorem (Theorem 4.1.9 and Corollary 4.1.10)**.** *For any fixed h, N and any $\delta \geq 0$, the optimization problem (2.2.7) with the distance $\|\cdot\|_h$ has almost surely a unique solution. The distance of the optimizer $\eta^\star_{h,N}$ to the target γ satisfies:*

- *If $1 > h_0 \geq h_N \gg \frac{1}{N}$, then*

$$\eta^*_{h_N,N} \xrightarrow{(d)} \gamma \quad \textit{in probability.}$$

- *If $\delta = 0$, there exists h_0 such that, for $h > h_0$,*

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] = o(1/N) \tag{2.4.2}$$

as N goes to infinity.

The second result, Equation (2.4.2), justifies our strategy in the sense that we managed to decrease significantly the second term in the decomposition (2.2.6) of the mean square error. Considering the first term in that decomposition leads naturally to the following (see in Conjecture 4.1.11), which is supported by numerical tests: for any h, there exists a constant C_h such that the optimal weights for the L^2 method with $\delta = 0$ satisfy

$$\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \leq \frac{C_h}{N}. \tag{2.4.3}$$

Also, we can have the following corollary.

Corollary (Corollary 4.1.13)**.** *Assume that (2.4.3) holds true (Conjecture 4.1.11). Let h be large enough, and assume that M_ϕ is regular enough so that m_ϕ (see (2.2.4)) which depends on h is finite – for instance M_ϕ is analytic. Assume also that $V_\phi(x)$ is bounded above by a constant v_ϕ. Then the L^2 method with $\delta \leq v_\phi/m_\phi^2$ satisfies*

$$MSE(\mathbf{\Phi}_W) = \mathbb{E}\left[\left(\Phi_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right] \leq \frac{C_h v_\phi}{N} + m_\phi^2 o(1/N)$$

for some numerical constant C_h and numerical $o(1/N)$.

Therefore, under the above assumptions, the L^2 method is asymptotically better than the naïve Monte Carlo approach in terms of MSE as soon as $v_\phi \leq \mathrm{Var}(M_\phi(X))/(C_h - 1)$.

We are able to prove similar results for the Wasserstein method.

Theorem (Theorem 4.1.15)**.** *For any N, the optimization problem with $\delta = 0$ has almost surely a unique solution. The distance $D = \mathcal{W}_1\left(\eta^\star_{Wass,N}, \gamma\right)$ of the optimizer $\eta^\star_{Wass,N}$ to the target γ satisfies for all integer $p \geq 1$ the moment bounds:*

$$\mathbb{E}\left[D^p\right] = \mathcal{O}^\star\left(\frac{1}{N^p}\right),$$

where $\mathcal{O}^\star$ means $\mathcal{O}$ up to logarithmic factors. In particular,

$$\eta^\star_{Wass,N} \xrightarrow[N\to+\infty]{Law} \gamma \quad in \quad in\ probability.$$

Moreover, the optimal weights satisfy:

$$\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \leq \frac{6}{N}.$$

Taking $p = 2$ in the previous theorem, we obtain the following corollary:

Corollary (Corollary 4.1.16)**.** *Assume M_ϕ is Lipschitz – so that $m_\phi < +\infty$, and that $V_\phi(x)$ is bounded above by a constant v_ϕ. Then the Wasserstein method with $\delta = 0$ satisfies*

$$MSE(\mathbf{\Phi}_W) = \mathbb{E}\left[\left(\Phi_W(\mathbf{X}, \mathbf{Y}) - \mathbb{E}\left[\phi(X, Y)\right]\right)^2\right] \leq \frac{c_0 v_\phi}{N} + m_\phi^2 o(1/N)$$

for some numerical constant $c_0 \leq 6$ and numerical $o(1/N)$.

Therefore, under the above assumptions, the Wasserstein method is asymptotically better than the naïve Monte Carlo approach in terms of MSE as soon as $v_\phi \leq \mathrm{Var}(M_\phi(X))/(c_0 - 1)$.

Remark 2.4.2. *A careful look at the proof shows that, if the X_i follow the uniform distribution on $[0, 1]$, the bound on $\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right]$ may be divided by 4, leading to $c_0 = 3/2$. Numerical tests suggest that even in the Gaussian case, this bound still holds true asymptotically in the sense that $N\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \to \frac{3}{2}$. Therefore, the Wasserstein method conjectured to be better than the naïve Monte Carlo approach as soon as $v_\phi \leq 2\,\mathrm{Var}(M_\phi(X))$.*

Remark 2.4.3 (Relative strength of the results)**.** *The results for both methods are quite similar. The main difference is that in the Wasserstein setting, we are actually able to control the variance of the weights with an explicit constant $c_0 = 6$; in the L^2 case we only conjecture that a similar result holds. This difference essentially comes from the fact that in our one dimensional setting, the optimal Wasserstein weights are explicit, and therefore much easier to study.*

2.5 Summary of main proofs

Let us recall that $\mathcal{O}^*$ means $\mathcal{O}$ up to logarithmic factors.

2.5.1 The idea of the proof of (2.4.2) in Theorem 4.1.9

Let us recall (2.4.2) : in the $\delta = 0$ case, for h sufficiently large,

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] = o(1/N).$$

and $\eta^\star_{h,N}$ is defined by minimizing $\|\sum w_n \delta_{X_n} - \gamma\|_h^2$ over all weight vectors. The main difficulty here is that this minimizer is not explicit. However, for any integer K, the kernel

k_t admits an explicit spectral decomposition in terms of Hermite polynomials (see e.g. [AS92, Chapter 22] for details), leading to the following formula for the distance:

$$\|\eta_N - \gamma\|_h^2 = \sum_{k=1}^{K} e^{-2kt} \left(\sum_n w_n h_k(X_n) \right)^2 + \sum_{k>K} e^{-2kt} \left(\sum_n w_n h_k(X_n) \right)^2 . \tag{2.5.1}$$

Let $w_n^K(\mathbf{X})$ be an optimizer of the *first*, finite dimensional term. If N is large enough with respect to K, then it is reasonable to expect that, with high probability, the value of this finite dimensional problem is zero.

We now let $M < N$ be an integer and decompose the real line $\mathbb{R}$ in M segments between the quantiles $(x_i)_{1 \leq i \leq M-1}$, where $F_\gamma(x_i) = \int_{-\infty}^{x_i} \gamma(dx) = i/M$, F_γ begin the repartition function. If there is at least one of the $(X_n)_{1 \leq n \leq N}$ in each of the M "bins" $(]x_{i-1}, x_i[)_{1 \leq i \leq M}$, we say that the sample is "M-spread" and we denote by G this "good event".

It will proven in Section 4.3.6, using elementary but tricky calculations and the fact that a polynomial of degree K has no more than K roots, that if $M > CK^{5/2}8^K$ for some universal numerical constant C, and if the "good event" is satisfied then almost surely

$$\sum_n w_n^K(\mathbf{X}) h_k(X_n) = 0, \quad \forall k = 1 \dots K.$$

Now in order to obtain an upper bound, we can replace the minimizer $w_n(\mathbf{X})$ by $w_n^K(\mathbf{X})$ on the 'good event' and by the naïve uniform weight on the 'bad event' so that:

$$\mathbb{E}\left[\left\| \eta_{h,N}^\star - \gamma \right\|_h^2 \right] \leq \underbrace{\mathbb{E}\left[\left\| \sum_n w_n^K(\mathbf{X}) \delta_{X_n} - \gamma \right\|_h^2 \mathbf{1}_G \right]}_{(a)} + \underbrace{\mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^2 \mathbf{1}_{G^c} \right]}_{(b)}$$

For the first term (a), on the good event G, we apply (2.5.1): by definition the first term of (2.5.1) vanishes and we bound the second term of (2.5.1) using Jensen's inequality.

For the term (b), on the bad event, we use Hölder's inequality and hypercontractivity of the Ornstein-Uhlenbeck semigroup. This yields an upper bound of the form

$$\mathbb{E}\left[\left\| \eta_{h,N}^\star - \gamma \right\|_h^2 \right] \leq N e^{-2tK} a(t) + \frac{1}{\sqrt{N}} b(t) \mathbb{P}\left[G^c\right]^{1/2} . \tag{2.5.2}$$

To conclude, we control the probability of the bad event $\mathbb{P}\left[G^c\right]$ by coupon collecting: suppose that $N > 4M \ln(M)$. For $\mathbf{X} = (X_1, ..., X_N)$, an iid gaussian sample, the probability

of the bad event G^c is small:

$$\mathbb{P}\left[G^c\right] = \mathbb{P}\left[\exists i, \forall n, X_n \notin (x_{i-1}, x_i)\right] \leq \frac{M}{M-1}\frac{1}{M^3}.$$

Finally, we fix $t > \ln 8$ and choose K and M as large as possible such that $M > CK^{\frac{5}{2}}8^K$ and $N > 4M \ln M$. It can then be checked that both (a) and (b) are $o(\frac{1}{N})$.

2.5.2 The proof of $\mathbb{E}\left[D^p\right] = \mathcal{O}^\star\left(\frac{1}{N^p}\right)$ in Theorem 4.1.15

Let us now consider the proof of the results for the Wasserstein method (see details in Theorem 4.1.15).

By Proposition 4.4.1, the optimal coupling between a Gaussian random variable X and the optimally reweighted empirical measure $\sum_n w_n(\mathbf{x})\delta_{x_n}$ is given by the piecewise constant transport map that sends each interval $]y_n, y_{n+1}[$ to x_n, so D has the explicit expression

$$D = \int \min_n |x - X_n| \, \gamma(dx).$$

We now let $M < N$ be an integer and decompose the real line $\mathbb{R}$ in M segments between the quantiles $(x_i)_{0\leq i\leq M}$, where $F_\gamma(x_i) = \int_{-\infty}^{x_i} \gamma(dx) = i/M$. As the same approach before (If there is at least one of the $(X_n)_{1\leq n\leq N}$ in each of the M "bins" $(]x_{i-1}, x_i[)_{1\leq i\leq M}$, we say that X is M-well spread and we denote by G this "good event"), then there exist $N(1), ..., N(M)$ such that $X_{N(i)} \in]x_{i-1}, x_i[$. Therefore

$$\begin{aligned}
D\mathbf{1}_G &= \int \min_n |x - X_n| \, \gamma(dx) \\
&\leq \sum_{i=1}^{M} \int_{x_{i-1}}^{x_i} \left|x - X_{N(i)}\right| \gamma(dx) \\
&\leq \frac{2}{M} x_{M-1} + 2\int_{x_{M-1}}^{\infty} |x - x_{M-1}| \, \gamma(dx) \\
&= \mathcal{O}^\star\left(\frac{1}{M}\right),
\end{aligned}$$

where the last line follows from known asymptotics as the quantile x_{n-1}. As stated in Lemma 4.3.12, a coupon collecting argument shows that the probability that the good event is not satisfied is small $\mathbb{P}\left[G^c\right] \leq \frac{M}{M-1}\frac{1}{M^{2p+1}}$, if $N > (2p+2)M \ln M$. Then we

combine these two cases, the result is as follows:

$$\begin{aligned}\mathbb{E}\left[D^p\right] &= \mathbb{E}\left[D^p \mathbf{1}_G\right] + \mathbb{E}\left[D^p \mathbf{1}_{G^c}\right] \\ &\leq \mathcal{O}^\star\left(\frac{1}{M^p}\right) + \sqrt{\mathbb{E}\left[D^{2p}\right]}\sqrt{\mathbb{P}\left[G^c\right]} \\ &\leq \mathcal{O}^\star\left(\frac{1}{N^p}\right),\end{aligned}$$

using a priori bounds on $\mathbb{E}\left[D^{2p}\right]$ (see Section 4.4.2 for details).

2.5.3 The proof of the control in l^2 of the optimal weights in Theorem 4.1.15

Let us now see how the explicit formula for the Wasserstein optimal weights can be used to control their l^2 norm. By Proposition 4.4.1, the weight of the optimal coupling is explicit:

$$w_n(\mathbf{X}) = F_\gamma(Y_{n+1}) - F_\gamma(Y_n),$$

where the Y_n are the middle points of the reordered sample $\left(X_{(1)}, ..., X_{(N)}\right)$ and F_γ is the cdf (*cumulative distribution function*) of the standard Gaussian distribution. By a rough upper bound, for $2 \leq n \leq N-1$,

$$w_n \leq F_\gamma^{-1}(X_{(n+1)}) - F_\gamma^{-1}(X_{(n-1)}).$$

The cdf F_γ maps the ordered sample $(X_{(1)}, ..., X_{(N)})$ to an ordered sample $(U_{(1)}, ..., U_{(N)})$ of the uniform distribution on $[0, 1]$, so

$$w_n \leq U_{(n+1)} - U_{(n-1)},$$

for $2 \leq n \leq N-1$, $w_1 \leq U_{(2)}$ and $w_n \leq 1 - U_{(N-1)}$. Then by simple calculation, It is easily to check the control in l^2 of the optimal weights.

2.6 Summary of numerical experiments

2.6.1 Comparison among naïve Monte Carlo, L^2 method and Wasserstein distance

We supplement our theoretical findings with numerical tests. In the first series of tests, we compare numerically the naïve Monte Carlo method, the L^2 method with various choices of the bandwidth, and the Wasserstein method, in the toy case where X itself is the variable of interest. For simplicity, and for homogeneity between the two methods, we

have chosen in this first series of numerical tests $\delta = 0$ (up to numerical precision) in the the L^2-method. This case is an idealized case which highlights concrete problems where $v_\phi \ll m_\phi^2$.

The full results may be found in Section 4.5. Figure 2.2 and Figure 2.3 show that both methods perform much better than the naïve Monte Carlo estimator. The L^2 method is often able to reduce significantly the statistical error, but the bandwidth parameter h must be chosen carefully, depending on N and the type of observable we are interested in. The parameter-free Wasserstein method is faster and more robust, but may be outperformed by a well-tuned L^2 method for very regular observables (here, cosinus).

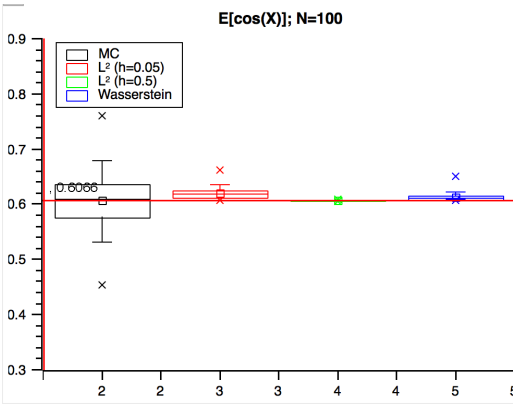


Figure 2.2: Comparsion among three methods in $\mathbb{E}\left[\cos(X)\right]$.

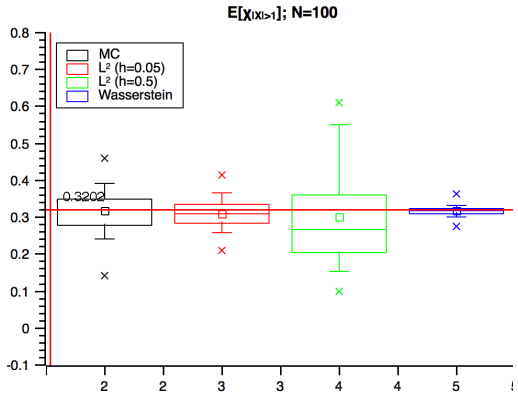


Figure 2.3: Comparsion among three methods in $\mathbb{E}\left[\chi_{|x|>1}\right]$.

Figure 2.4 and Figure 2.5 investigate the influence of the bandwidth parameter h on the L^2 method by test various values of the bandwidth. Figure 2.4 corresponds to the test function cos; a bias clearly appears in that case, and the estimator is better when h is quite large, with a trade-off at $h = .5$ ($h = .8$ is not as good). However, when we apply the method to estimate the expectation of a discontinuous function of X, here $x \mapsto \mathbf{1}_{|x|>1}$, which does not belong to the appropriate class of regularity, the picture is completely different and the best estimator is obtained for a much smaller $h \approx 0.05$, as can be seen

in Figure 2.5.

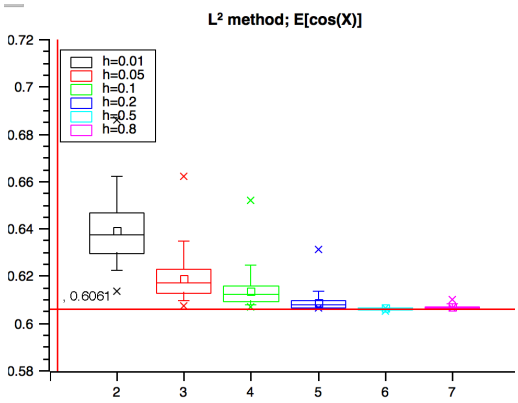


Figure 2.4: Test $x \mapsto \cos(x)$ with various bandwidth.

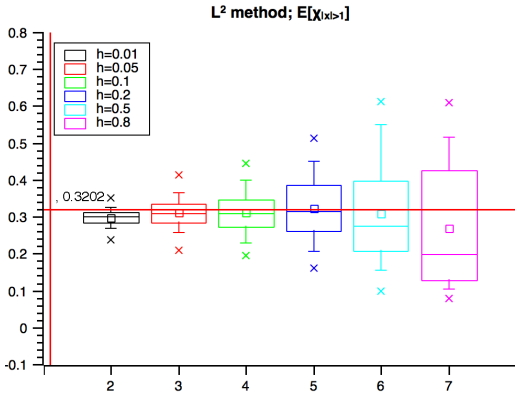


Figure 2.5: Test $x \mapsto \chi_{|x|>1}$ with various bandwidth.

2.6.2 Example of variance reduction

The second series of numerical tests in done in Section 4.6. We let G be a d-dimensional standard Gaussian vector, and assume we are interested in the distribution of a non-linear function

$$Y = F(G) = \frac{1}{\sqrt{d}} \sin\left(\frac{1}{\sqrt{d}} \sum_{i=1}^{d} G_i\right).$$

with parameters $d = 10$ and $N = 30$. Note that the linearisation of F gives

$$X = (DF)_0 G = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} G_i \sim \gamma,$$

so that the distribution of X is an explicit one dimensional unit Gaussian. We then use our method to estimate, for any fixed t, the cumulant generating function $\log \mathbb{E}\left[\exp(tF(G))\right]$, using X as our "control variable". In this more realistic setting, we focus on the more

robust Wasserstein method described in Section 2.3.2, and show how it can be compared to, and combined with, a more classical control variate approach. The additional control variate is obtained by adding

$$\log 1/N \sum_{n=1}^{N} w_n(\mathbf{X}) e^{tX_n} - t^2/2$$

to each of the considered estimator. Note that $t^2/2 = \log \int e^{tx} \gamma(dx)$ is the cumulant generating function of the gaussian distribution.

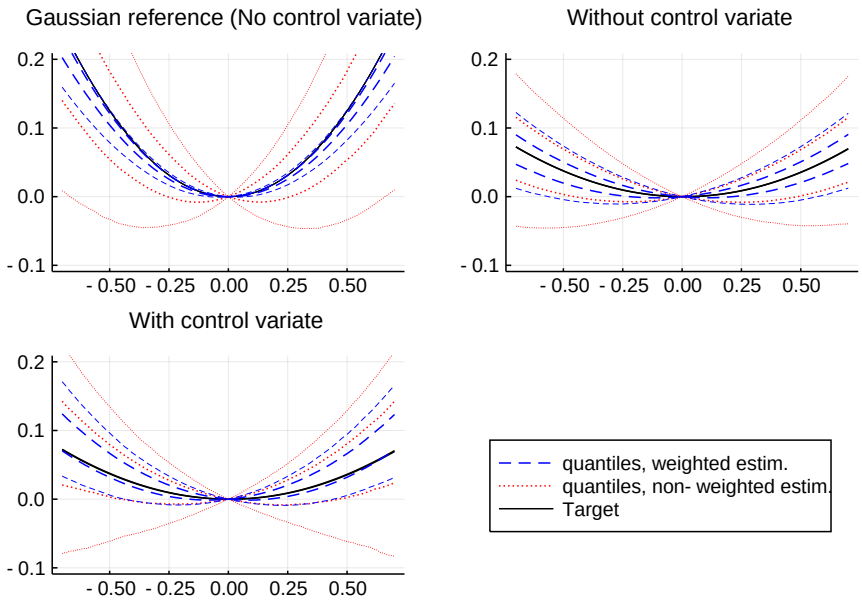


Figure 2.6: The figures above represents the $[.05, .25, .75, .95]$-quantile envelops of the different estimators of the cumulant generating functions for $\log \mathbb{E}\left[\exp(tF(G))\right]$.

In Figure 2.6, we clearly see that our -generic- reweighting method always reduces variance in a substantial way, without using any prior information of F. This non-linear example also shows that a standard variance reduction by a linear control variate maybe be useless. Note that the use of a control variate and the weighting method can be used simultaneously.

2.7 Conclusion

We have proposed *generic and robust* variance reduction techniques based on reweighting samples using a one dimensional Gaussian control variable. The latter can be seen as generalization of post-stratification methodologies and can outperform variance reduction by control variate even in simple situations.

Theoretically, the results, which prove effective variance reduction for both methods, are

quite similar. The main difference is that in the Wasserstein setting, we are actually able to control the variance of the weights with an explicit constant $c_0 \leq 6$; in the L^2 case we only conjecture that a similar result holds. This difference essentially comes from the fact that in our one dimensional setting, the optimal Wasserstein weights are explicit, and therefore much easier to study.

Numerically, we have observed (as theoretically suggested) that the L^2 approach, as compared to the Wasserstein approach, requires more regular observables and some tuning, and is more costly when the sample size become large. Note however, that the L^2 approach may be amenable to control variables X in higher dimension, where Wasserstein optimization – optimal transport – problems are known to be very cumbersome. This issue is left for future work.

Part II

A weak overdamped limit theorem for Langevin processes

Chapter 3

A weak overdamped limit theorem for Langevin processes

3.1 Introduction

This paper focuses on the overdamped asymptotics of Langevin dynamics. The Langevin Stochastic Differential Equation (SDE) describes the dynamics of a classical mechanical system perturbed by a stochastic thermostat. The system state at time $t \geq 0$ is encoded by its position Q_t and its momentum P_t. More formally, the equation reads:

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

where in the above, Q_t takes values in the d-dimensional torus $\mathbb{T}^d$, P_t takes values in $\times\mathbb{R}^d$, the function $V : \mathbb{T}^d \to \mathbb{R}$ is the particles' potential energy, $\beta > 0$ the inverse temperature, and $t \mapsto W_t \in \mathbb{R}^d$ is a standard d-dimensional Brownian motion. The term $\sqrt{2\beta^{-1}} dW_t$ is a fluctuation term bringing energy into the system, while this energy is dissipated through the friction term $-P_t dt$; the sum of these two terms forming the so-called thermostat part. The remaining terms are simply Newton's equation of motion. For more details on this equation, we refer to [LRS10, Section 2.2].

The case we consider here is the so-called overdamped asymptotics, where the time scale of the large damping due to friction is much smaller than the time scale of the Hamiltonian dynamics, so that the momentum becomes a fast variable compared to the slow position variable. We introduce a parameter ε for the ratio of the time scales, and consider

$$\begin{cases} dQ_t^\varepsilon = \frac{1}{\varepsilon} P_t^\varepsilon dt, \\ dP_t^\varepsilon = -\frac{1}{\varepsilon}\nabla V_\varepsilon(Q_t^\varepsilon)dt - \frac{1}{\varepsilon^2} P_t^\varepsilon dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}} dW_t. \end{cases} \tag{3.1.1}$$

Note that we allow the potential $V_\varepsilon \in C^1(\mathbb{T}^d)$ to depend on ε and will only suppose that it converges to a limit V; see below for a precise statement. The Markov generator L_ε associated with (3.1.1) is given by

$$L_\varepsilon f(q,p) := \frac{1}{\varepsilon^2}\left(\frac{1}{\beta}\Delta_p f - p\cdot\nabla_p f\right) + \frac{1}{\varepsilon}\left(p\cdot\nabla_q f - \nabla_q V_\varepsilon\cdot\nabla_p f\right), \tag{3.1.2}$$

where f denotes any smooth test function of the variables $(q,p)\in\mathbb{T}^d\times\mathbb{R}^d$.

Overdamped processes are stochastic dynamics on the system position $(Q_t)_{t\geq 0}$ only. The overdamped Langevin SDE is given by:

$$dQ_t = -\nabla V(Q_t)dt + \sqrt{2\beta^{-1}}dB_t, \tag{3.1.3}$$

where $V:\mathbb{T}^d\to\mathbb{R}^d$ is a potential energy, limit of V_ε when $\varepsilon\to 0$ in some appropriate sense, and $t\mapsto B_t\in\mathbb{R}^d$ is a standard d-dimensional Wiener process. The Markov generator L associated with (3.1.3) acts on smooth test functions f of the variable q as follows:

$$Lf(q) := -\nabla_q V\cdot\nabla_q f + \frac{1}{\beta}\Delta_q f.$$

Our main result is the proof of the convergence in distribution of the Langevin position process $(Q_t^\varepsilon)_{t\geq 0}$ towards its overdamped counterpart $(Q_t)_{t\geq 0}$, assuming the uniform convergence of the gradient potential as well as a control of moments of the initial kinetic energy.

Theorem 3.1.1 (Overdamped limit of the Langevin dynamics)**.** *For any $\varepsilon>0$, let $(Q_t^\varepsilon,P_t^\varepsilon)_{t\geq 0}\in\mathbb{T}^d\times\mathbb{R}^d$ be a weak solution to the SDE (3.1.1). Assume that the following conditions hold:*

1. *V_ε is $C^1(\mathbb{T}^d)$, and converges to V in the sense that $\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow[\varepsilon\to 0]{} 0$,*
2. *The following moment bound holds true:*

$$\lim_{\varepsilon\to 0}\varepsilon\mathbb{E}(|P_0^\varepsilon|^3) = 0$$

3. *The initial position distribution is converging to some limit: $Law(Q_0^\varepsilon)\xrightarrow[\varepsilon\to 0]{} Law(Q_0)$.*

Then, when $\varepsilon\to 0$, the process $(Q_t^\varepsilon)_{t\geq 0}\in C(\mathbb{R}_+\to\mathbb{T}^d)$ converges in distribution to the unique weak solution of the overdamped SDE (3.1.3).

Remark 3.1.2. *In Theorem 3.1.1, the space of trajectories $C(\mathbb{R}_+\mapsto\mathbb{T}^d)$ is endowed with uniform convergence on compact sets; making it Polish (metrizable for a separable and complete metric).*

The literature on diffusion approximations is very rich; we refer for instance to Stuart-Pavliotis in [PS08] for a recent pedagogical overview of related issues. Historically, a possible chain of seminal references is given by Stratonovich in [Str63], Khas'minskii in [Kha66], Papanicolaou-Varadhan in [PV73], as well as Papanicolaou-Kohler in [PK74]; complemented with the more modern viewpoint of Ethier-Kurtz in [EK86], Chapter 12 "Random evolutions".

In the present case, the momentum variable is averaged out with the diffusion approximation, so that the problem may be labeled as "diffusion approximation with averaging". Broadly speaking, the problem can be approached using strong or weak convergence techniques. For an example of the strong convergence approach, the results in [SSMD82] rely on estimating the dynamics of Q_t^ε and its limit using a Gronwall argument; this approach requires the Lipschitz continuity of ∇V_ε uniformly in ε. On the other hand, weak convergence results rely on the so-called "perturbed" test function or "corrector" approach, that have been developed since Panicolaou-Stroock-Varadhan in [PSV77]. The case of the overdamped limit (3.1.1) is not directly covered by these results. Indeed, the correctors are not bounded in the present case, due to the fact that the state space of the momentum variable is not compact.

In a series of papers [PV01, PV03, PV05], Pardoux-Veretennikov extend the classical diffusion approximation with averaging to the non-compact state space case. In the latter setting however, the slow variable has a dynamics independent of the fast one, which is not the case in the Langevin case (3.1.1).

We now give a physically motivated example that satisfies our assumptions but was not covered by previous works.

Example 3.1.3. *Let*

$$V_\varepsilon(q) = V(q) + \alpha_\varepsilon \chi(k_\varepsilon q),$$

where $\chi \in C^\infty(\mathbb{T}^d)$*, and the scaling coefficients* $k_\varepsilon \in \mathbb{N}$ *and* $\alpha_\varepsilon \in \mathbb{R}$ *satisfy*

$$k_\varepsilon \to \infty, \qquad \alpha_\varepsilon k_\varepsilon \to 0.$$

Physically, the potential $\alpha_\varepsilon \chi(k_\varepsilon q)$ *may model the interaction between a particle with unit energy and a periodic crystal of small period* k_ε^{-1}*, and small energy range of order* α_ε*. When* $k_\varepsilon \to +\infty$ *but* $\alpha_\varepsilon k_\varepsilon = 1$ *and* ε *is kept constant, the effective action of the periodic crystal on the particle can not be neglected, especially for grazing velocities co-linear to the principal directions of the crystal. Indeed, in the latter case, on times of order* 1*, the crystal exerts on the particle a total force also of order* 1*, making it deviating from its trajectory.*

Our result shows that the physically necessary condition $\alpha_\varepsilon k_\varepsilon \to 0$ *is in fact sufficient for neglecting the crystal effect in the overdamped regime. Note that if* $\alpha_\varepsilon k_\varepsilon^2 \to +\infty$*, when*

$\varepsilon \to 0$, then

$$\|\nabla V_\varepsilon - \nabla V\|_\infty \xrightarrow{\varepsilon \to 0} 0,$$

but still

$$\|\nabla^2 V_\varepsilon\|_\infty \sim \alpha_\varepsilon k_\varepsilon^2 \|\nabla^2 \chi\|_\infty \xrightarrow{\varepsilon \to 0} +\infty,$$

preventing ∇V_ε from being Lipschitz uniformly in ε; and hence forbidding results based on strong convergence.

In order to prove Theorem 3.1.1, we will establish a more general weak convergence result. We consider a sequence (indexed by a small parameter $\varepsilon > 0$) of Markov processes of the form $t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d$ taking value in the Skorokhod path space $\mathbb{D}_{\mathbb{T}^d \times \mathbb{R}^d}$. Our general convergence result, namely Theorem 3.3.5, gives general conditions under which $(Q_t^\varepsilon)_{t \geq 0}$ converges in distribution to the unique solution of a particular martingale problem. The proof follows the usual pattern: first we prove tightness for the family of distributions of (Q_t^ε), and then characterize the limit through martingale problems. For both steps, we use the perturbed test function method. The key sufficient criteria yielding the results of both steps is given in Assumption 3.3.4, which states that to any smooth $f : \mathbb{T}^d \to \mathbb{R}$, we can associate a perturbed test function $f_\varepsilon : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}$ such that for all $T > 0$,

$$\lim_{\varepsilon \to 0} \mathbb{E}\left(\sup_{t \leq T} |f(Q_t^\varepsilon) - f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| \right) = 0 \quad \text{and} \quad \lim_{\varepsilon \to 0} \mathbb{E}\left(\int_0^T |Lf(Q_t^\varepsilon) - L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)| \, dt \right) = 0.$$

Remark 3.1.4 (On the choice of the state space)**.** *Theorem 3.3.5 can be useful for càd-làg processes, which explains the fact that we work in Skorokhod space. We have chosen to work in $\mathbb{T}^d \times \mathbb{R}^d$ for notational simplicity, but Theorem 3.3.5 could be extended to more general product spaces of the type $E \times F$, where E and F are Polish spaces. If E is compact, the extension is straightforward. If E is locally compact, then one can work with $E \cup \{\infty\}$, the one point compactification of E at infinity (see [EK86, Chapter 4]). If E is not locally compact, then one needs to use Theorem 9.1 in [EK86, Chapter 3] instead of Theorem 3.2.12 below which is a corollary of the former. In the latter case: (i) the a priori compact containment condition (9.1) of Theorem 9.1 in [EK86, Chapter 3] has to be proven; and (ii) one has to show the tightness of $\left(Law\left(f(Q_t^\varepsilon)\right)_{t \geq 0}\right)_{\varepsilon \geq 0}$ for all f in a space of functions dense in $C_b(E)$ for the topology of uniform convergence on compacts. Such extensions to infinite dimensional spaces are left for future work.*

The paper is organized as follows. Section 3.2 starts with some notation and preliminaries. In Section 3.3, we state and prove the general convergence result Theorem 3.3.5. This general method is then applied in Section 3.4 to the overdamped Langevin limit, proving Theorem 3.1.1.

3.2 Notation and Preliminaries

In what follows, we introduce notation and recall some known results.

3.2.1 General notation

Let (E, d) be a Polish space, that is, a topological space which is metric, complete and separable. Denote $C(E)$ the Banach space of all continuous functions and $C_b(E)$ the Banach space of all bounded continuous functions. We denote by $\mathcal{P}(E)$ the space of probability measures on the Borel σ-field $\mathcal{B}(E)$. The notation $\mathcal{F}_t^X$ means the natural filtration of càd-làg processes $(X_t)_{t\geq 0}$, that is $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$. For any $(s,t) \in \mathbb{R} \times \mathbb{R}$, we denote by $s \wedge t$ the minimum of s and t.

3.2.2 The Skorokhod space

A càd-làg function (from the French "continu à droite, limité à gauche", also called RCLL for "right continuous with left limits") is a function defined on $\mathbb{R}_+$ that is everywhere right-continuous and has left limits everywhere. The collection of càd-làg functions on a given domain is known as the Skorokhod space. We denote $\mathbb{D}_E$ the space of càd-làg functions with values in a Polish space E. We recall that this path space $\mathbb{D}_E$ may be equipped with the Skorokhod topology (see Section 5 of [EK86, Chapter 3]): a family of trajectories $(q_s^\varepsilon)_{s\geq 0}$ indexed by ε converges to a limit trajectory $(q_s^0)_{s\geq 0}$ if there exists a sequence $(\lambda_\varepsilon)_{\varepsilon\geq 0}$ in the space of strictly increasing continuous bijections of $[0, \infty[$, such that for each $T > 0$: $\lim_{\varepsilon\to 0} \sup_{t\leq T} |\lambda_\varepsilon(t) - t| = 0$ and $\lim_{\varepsilon\to 0} \sup_{t\leq T} d\left(q_t^\varepsilon, q^0_{\lambda_\varepsilon(t)}\right) = 0$. The following result will be useful in the proof of Theorem 3.3.5.

Lemma 3.2.1. *Integration with respect to time is continuous with respect to the Skorokhod topology: if $(q_t^\varepsilon)_{t\geq 0}$ converges to $(q_t^0)_{t\geq 0}$ in $\mathbb{D}_E$, and $\psi : E \to \mathbb{R}$ is bounded and continuous, then for each $T > 0$,*

$$\int_0^T \psi(q_t^\varepsilon)dt \xrightarrow[\varepsilon\to 0]{} \int_0^T \psi(q_t^0)dt.$$

Proof. Let us denote by $J_T := \{t \in [0,T], q^0_{t^-} \neq q_t^0\}$ the countable set of jump times in $[0,T]$ of q^0. By definition of convergence in the Skorokhod space,

$$\lim_{\varepsilon\to 0} q_s^\varepsilon = q_s^0 \qquad \forall s \in [0,T] \setminus J_T.$$

Since J_T has Lebesgue measure 0 and ψ is continuous and bounded, dominated convergence yields the result. □

3.2.3 Martingale problems

Let us first recall some basics on martingales and stochastic calculus. Let $(\Omega, \mathcal{F}, \mathbf{P}, (\mathcal{F}_t)_{t\geq 0})$ a filtered probability space. A càd-làg real-valued process $(X_t)_{t\geq 0}$ is said to be adapted if X_t is $\mathcal{F}_t$-measurable for all $t \geq 0$, and is called a $(\mathcal{F}_t)_{t\geq 0}$-martingale if $\mathbb{E}(|X_t|\,|\mathcal{F}_s) < +\infty$ and $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for any $0 \leq s \leq t$.

We will often need the technical tool of localization by stopping times, to deal with the unboundedness of the momentum variable. We follow here the presentation of [EK86, Chapter 4].

Definition 3.2.2 (Local martingale)**.** *A càd-làg real-valued process $(X_t)_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, \boldsymbol{P}, (\mathcal{F}_t)_{t\geq 0})$ is called a* local martingale *with respect to $(\mathcal{F}_t)_{t\geq 0}$ if there exists a non-decreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of $(\mathcal{F}_t)_{t\geq 0}$-stopping times such that $\tau_n \to \infty$ $\boldsymbol{P}$-almost surely, and for every $n \in \mathbb{N}$, $\left(X_{t\wedge\tau_n}\right)_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$-martingale.*

Let us now state precisely what it means for a process to solve a martingale problem.

Definition 3.2.3 (Martingale problem)**.** *Let E be a Polish space. Let L be a linear operator mapping a given space $\mathcal{D} \subset C_b(E)$ into bounded measurable functions. Let μ be a probability distribution on E. A càd-làg process $(X_t)_{t\geq 0}$ with values in E solves the* martingale problem *for the generator L on the space $\mathcal{D}$ with initial measure μ — in short, X solves $\boldsymbol{MP}(L, D(L), \mu)$ — if $Law(X_0) = \mu$ and if, for any $\varphi \in \mathcal{D}$,*

$$t \mapsto M_t(\varphi) := \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s)ds \tag{3.2.1}$$

is a martingale with respect to the natural filtration $\left(\mathcal{F}_t^X = \sigma\left(X_s,\, 0 \leq s \leq t\right)\right)_{t\geq 0}$.
Moreover, the martingale problem $\boldsymbol{MP}(L, \mathcal{D}, \mu)$ is said to be well-posed *if:*

- *There exists a probability space and a càd-làg process defined on it that solves the martingale problem (existence);*
- *whenever two processes solve $\boldsymbol{MP}(L, \mathcal{D}, \mu)$, then they have the same distribution on $\mathbb{D}_E$ (uniqueness).*

3.2.4 Weak solutions of SDEs

Let $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \mapsto \mathbb{R}^{d\times n}$ be locally bounded. Consider a stochastic differential equation in $\mathbb{R}^d$ of the form:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \tag{3.2.2}$$

with an initial condition $\text{Law}(X_0) = \mu_0$. Let L be the formal generator

$$L := \sum_{i=1}^{d} b_i \partial_i + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij} \partial_i \partial_j, \tag{3.2.3}$$

where $a = \sigma\sigma^T$.

Definition 3.2.4 (Weak solution of the SDE)**.** *A continuous process $(X_t)_{t\geq 0}$ is a* weak solution *of (3.2.2) if there exists a filtered probability space $(\Omega, \mathcal{F}, \boldsymbol{P}, (\mathcal{F}_t)_{t\geq 0})$ such that:*

- *$t \mapsto W_t$ is a $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion, that is, an $(\mathcal{F})_{t\geq 0}$-adapted process such that $Law(W_{t+h} - W_t | \mathcal{F}_t) = \mathcal{N}(0, h)$.*
- *X is a continuous, $(\mathcal{F}_t)_{t\geq 0}$-adapted process and satisfies the stochastic integral equation*

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dW_s \quad a.s.$$

We now quote two results from [EK86] concerning existence and uniqueness of solutions to SDEs and martingale problems. The first is an existence result, and can be found in [EK86, Section 5.3] (Corollary 3.4 and Theorem 3.10).

Theorem 3.2.5. *Assume that b, σ are continuous. If there exists a constant K such that for any $t \geq 0$, $x \in \mathbb{R}^d$:*

$$|\sigma|^2 \leq K(1 + |x|^2); \tag{3.2.4}$$

$$x \cdot b(x) \leq K(1 + |x|^2), \tag{3.2.5}$$

then there exists a weak solution of the stochastic differential equation (3.2.2) corresponding to (σ, b, μ), which is also solution of the martingale problem $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$, $C_c^\infty(\mathbb{R}^d)$ being the set of smooth functions with compact support.

Remark 3.2.6. *For the Langevin equation (3.1.1)) we first remark that the latter can be set in $\mathbb{R}^d \times \mathbb{R}^d$ using the $\mathbb{Z}^d$-periodic extension of V_ε. Then $b(q,p) = \left(\frac{1}{\varepsilon}p, -\frac{1}{\varepsilon}\nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p\right)$ and $\sigma = (0, \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\,\mathrm{Id}_{\mathbb{R}^d})$ are continuous since $V_\varepsilon \in C^1(\mathbb{R}^d)$. Moreover, $|\sigma|^2 = \sigma\sigma^\top = (0, \frac{2}{\beta\varepsilon^2}\,\mathrm{Id}_{\mathbb{R}^d})$, and on the other hand*

$$(q,p) \cdot b(q,p) = \frac{1}{\varepsilon}pq - \frac{1}{\varepsilon}p\nabla V_\varepsilon(q) - \frac{1}{\varepsilon^2}p^2 \leq \frac{1}{2\varepsilon}(1 + \|\nabla V_\varepsilon\|_\infty)(1 + |p|^2 + |q|^2),$$

which implies the existence of weak solution of (3.1.1) in $\mathbb{R}^d$. One then obtains existence of a weak solution in $\mathbb{T}^d$ of the original (3.1.1) using the canonical continuous mapping $\mathbb{R}^d \to \mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$.

The next result follows from [EK86] (Theorem 1.7 in Section 8.1) and [SV07] (Theorem 10.2.2 and the discussion following their Corollary 10.1.2) .

Theorem 3.2.7. *Assume that the bounds (3.2.4) and (3.2.5) hold. Suppose that $a := \sigma\sigma^\top$ is continuous and uniformly elliptic:*

$$\exists C_a > 0, \forall \xi \in \mathbb{R}^d, \forall x \in \mathbb{R}^d, \quad \xi^\top a(x)\xi \geq C_a|\xi|^2.$$

Then for any initial condition μ, there is a unique weak solution of the stochastic differential equation (3.2.2); which is also the unique solution of the martingale problem $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$.

Remark 3.2.8. *For the overdamped Langevin equation (3.1.3), we remark again that the latter can be set in $\mathbb{R}^d$ using the $\mathbb{Z}^d$-periodic extension of V_ε. One then obtains well-posedness of the martingale problem $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ in $\mathbb{R}^d$ since ∇V is bounded and continuous by assumption. This solution obviously solves $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ in $\mathbb{T}^d$. The fact that uniqueness of $\boldsymbol{MP}(L, C_c^\infty(\mathbb{R}^d), \mu)$ implies uniqueness of $\boldsymbol{MP}(L, C_c^\infty(\mathbb{T}^d), \mu)$ is technically less obvious. It can be treated using the localization technique of Theorem A.1.1 stated in appendix. More precisely, using the notation of Theorem A.1.1, one can defines the covering of $\mathbb{R}^d$ by the open sets*

$$U_k := \left\{(x_1, \dots, x_d) \in \mathbb{R}^d | \, |x_i - k_i/8| \leq 1/4 \ \forall i = 1 \dots d\right\}$$

where $k \in \mathbb{Z}^d$ and then remark that by partition of unity for smooth functions, any $\varphi \in C_c^\infty(\mathbb{R}^d)$ can be written as a finite sum of smooth functions with compact support in each given U_k, $k \in \mathbb{Z}^d$.

3.2.5 Convergence in distribution

As we said before, we are interested here in proving convergence in distribution for processes. Let us briefly recall several key results that will be used later.

For completeness, we start by recalling the very classical Prohorov theorem, characterizing relative compactness by tightness (see for example Section 2 in[EK86, Chapter 3]).

Theorem 3.2.9 (Prohorov theorem)**.** *Let $(\mu_\varepsilon)_\varepsilon$ be a family of probability measures on a Polish space E. Then the following are equivalent:*

1. *$(\mu_\varepsilon)_\varepsilon$ is relatively compact for the topology of convergence in distribution.*
2. *$(\mu_\varepsilon)_\varepsilon$ is tight, that is to say, for any $\delta > 0$, there is a compact set K_δ such that*

$$\inf_\varepsilon \mu_\varepsilon(K_\delta) \geq 1 - \delta.$$

Over the years several relative compactness criteria in Skorokhod space have been developed. We will use the following one [EK86, Theorem 8.6, Chapter 4].

Theorem 3.2.10 (Kurtz-Aldous tightness criterion)**.** *Consider a family of stochastic processes $((X_t^\varepsilon)_{t\geq 0})_\varepsilon$ in $\mathbb{D}_{\mathbb{R}}$. Assume that $\left(Law(X_0^\varepsilon)\right)_\varepsilon$ is tight. $\forall \delta \in (0,1)$ and $T>0$, there exists a family of nonnegative random variable $\Gamma_{\varepsilon,\delta}$, such that: $\forall\ 0\leq t\leq t+h\leq t+\delta\leq T$*

$$\mathbb{E}\Big(|X_{t+h}^\varepsilon - X_t^\varepsilon|^2 | \mathcal{F}_t^{X^\varepsilon}\Big) \leq \mathbb{E}(\Gamma_{\varepsilon,\delta}|\mathcal{F}_t^{X^\varepsilon}); \tag{3.2.6}$$

with

$$\lim_{\delta\to 0}\sup_{\varepsilon}\mathbb{E}(\Gamma_{\varepsilon,\delta}) = 0. \tag{3.2.7}$$

Then the family of distributions $(Law((X_t^\varepsilon)_{t\geq 0}))_\varepsilon$ is tight.

Remark 3.2.11 (On using sequences)**.** *If $\varepsilon>0$ is a real number and that instead of (3.2.7), one considers the condition $\lim_{\delta\to 0}\limsup_{\varepsilon\to 0^+}\mathbb{E}(\Gamma_{\varepsilon,\delta})=0$, then the conclusion becomes the following: $(Law((X_t^{\varepsilon_n})_{t\geq 0}))_{\varepsilon_n}$ is tight for any $(\varepsilon_n)_{n\geq 1}$-sequence such that $\varepsilon_n>0$ and $\lim_{n\to+\infty}\varepsilon_n = 0$. This version will be the one used in the present paper.*

If the processes, say $(Q_t^\varepsilon)_{t\geq 0}$, is defined in a general state space E, it is natural to consider the image processes $(f(Q_t^\varepsilon))_{t\geq 0}$ for various observables, or test functions, f. The following result enables us to recover the tightness for the original process from the tightness of the observed processes (Corollary 9.3 Chapter 3 in [EK86]).

Theorem 3.2.12 (Tightness from observables)**.** *Let E be a compact Polish space and $((Q_t^\varepsilon)_{t\geq 0})_{\varepsilon>0}$ be a family of stochastic processes in $\mathbb{D}_E$. Assume that there is an algebra of test functions $\mathcal{D}\subset C_b(E)$, dense for the uniform convergence, such that for any $f\in\mathcal{D}$, $((f(Q_t^\varepsilon))_{t\geq 0})_{\varepsilon>0}$ is tight in $\mathbb{D}_{\mathbb{R}}$. Then $(Law(Q_t^\varepsilon)_{t\geq 0})_{\varepsilon>0}$ is tight in $\mathbb{D}_E$.*

Remark 3.2.13. *Again, the above theorem will be used for families indexed by sequences $(\varepsilon_n)_{n\geq 1}$ such that $\varepsilon_n>0$ and $\lim_{n\to+\infty}\varepsilon_n=0$.*

Finally, the following two lemmas will be useful when we considering martingale problems. The first one states that the distribution of jumps of càd-làg processes have atoms in a countable set (see Lemma 7.7 Chapter 3 in [EK86]).

Lemma 3.2.14. *Let $(X_t)_{t\geq 0}$ be a random process in the Skorokhod path space $\mathbb{D}_E$. The set of instants where no jump occurs almost surely:*

$$\mathcal{C}_{Law(X)} := \{t\in\mathbb{R}^+ | \mathbb{P}(X_{t^-}=X_t)=1\},$$

has countable complement in $\mathbb{R}^+$. In particular, it is a dense set.

The second one is a very useful way to check whether a process is a martingale or not (see page 174 in Ethier-Kurtz[EK86]).

Lemma 3.2.15 (Martingale equivalent condition)**.** *Let $(M_t)_{t\geq 0}$ and $(X_t)_{t\geq 0}$ be two càdlàg proceses and let $\mathcal{C}$ be an arbitrary dense subset of $\mathbb{R}_+$. Then $(M_t)_{t\geq 0}$ is $\mathcal{F}_t^X$-martingale if and only if*

$$\mathbb{E}\big[(M_{t_{k+1}} - M_{t_k})\varphi_k(X_{t_k})...\varphi_1(X_{t_1})\big] = 0,$$

for any time ladder $t_1 \leq ... \leq t_{k+1} \in \mathcal{C} \subset \mathbb{R}_+$, $k \geq 1$, and $\varphi_1, ..., \varphi_k \in C_b(E)$.

3.3 A general perturbed test function method

In this section, we consider a sequence of stochastic processes, indexed by a small parameter $\varepsilon > 0$, of the form

$$t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d,$$

taking value in the Skorokhod path space $\mathbb{D}_{\mathbb{T}^d\times\mathbb{R}^d}$ associated with the (Polish) product state space $\mathbb{T}^d \times \mathbb{R}^d$. Our goal is to describe a general framework to prove the convergence of the (slow) variables Q towards a well-identified dynamics. We use standard tightness arguments and characterization through martingale problems, emphasizing the technical role of perturbed test functions.

3.3.1 Notation and Assumptions

For each ε, we consider a càd-lag process $t \mapsto (Q_t^\varepsilon, P_t^\varepsilon) \in \mathbb{T}^d \times \mathbb{R}^d$. The natural filtration of the full process and the process $(Q_t^\varepsilon)_{t\geq 0}$ are denoted respectively by $\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon} := \sigma\left((Q_s^\varepsilon, P_s^\varepsilon), 0 \leq s \leq t\right)$, and $\mathcal{F}_t^{Q^\varepsilon} := \sigma\left(Q_s^\varepsilon, 0 \leq s \leq t\right)$. We now state the key assumptions that will imply convergence in distribution of the process $(Q_t^\varepsilon)_{t\geq 0}$ towards the solution of a martingale problem.

Assumption 3.3.1 (Generator of the process $(Q_t^\varepsilon, P_t^\varepsilon)$)**.** *There exists a linear operator L_ε acting on $C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$ which is the extended Markov generator of $(Q_t^\varepsilon, P_t^\varepsilon)_{t\geq 0}$ in the sense that, for all $f \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$, $L_\varepsilon f$ is locally bounded and*

$$t \mapsto M_t^\varepsilon(f) := f(Q_t^\varepsilon, P_t^\varepsilon) - f(Q_0^\varepsilon, P_0^\varepsilon) - \int_0^t L_\varepsilon f(Q_s^\varepsilon, P_s^\varepsilon)ds$$

is a $(\mathcal{F}_t^{Q^\varepsilon, P^\varepsilon})_{t\geq 0}$-local martingale.

Assumption 3.3.2 (The limit process)**.** *There exists a linear operator L mapping $C^\infty(\mathbb{T}^d)$ to $C(\mathbb{T}^d)$ such that the martingale problem $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$ is well-posed for any initial condition μ.*

Assumption 3.3.3 (Initial condition)**.** *The initial condition $(Law(Q_0^\varepsilon))_{\varepsilon>0}$ converge to a limit μ_0, when $\varepsilon \to 0$.*

Assumption 3.3.4 (Existence of perturbed test functions). *For all $f \in C^\infty(\mathbb{T}^d)$, there exists a perturbed test function $f_\varepsilon \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$, such that for all T, the rest terms*

$$R^\varepsilon_{1,t}(f) := |f(Q^\varepsilon_t) - f_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)| \quad \text{and} \quad R^\varepsilon_{2,t}(f) := |Lf(Q^\varepsilon_t) - L_\varepsilon f_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)|$$

satisfy the following bounds:

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\sup_{0\le t\le T} R^\varepsilon_{1,t}(f)\right) = 0, \tag{3.3.1}$$

$$\lim_{\varepsilon\to 0} \mathbb{E}\left(\int_0^T R^\varepsilon_{2,t}(f)dt\right) = 0. \tag{3.3.2}$$

3.3.2 The general convergence theorem

We are now in position to state our main abstract result.

Theorem 3.3.5. *Under the Assumptions 3.3.1, 3.3.2, 3.3.3, and 3.3.4, the family $\left(Law(Q^\varepsilon_t)_{t\ge 0}\right)_{\varepsilon>0}$ converges when $\varepsilon \to 0$ to the unique solution of martingale problem $\boldsymbol{MP}(L, C^\infty(\mathbb{T}^d), \mu)$.*

The proof follows the classical pattern, in two steps: we first prove that the processes Q^ε_t are relatively compact in $\mathbb{D}_{\mathbb{T}^d}$; then we show that any possible limit must solve the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$.

3.3.2.1 Step one: The proof of tightness.

We want to prove that for each sequence $(\varepsilon_n)_{n\ge 1}$ satisfying $\lim_n \varepsilon_n = 0$, $(Law(Q^{\varepsilon_n}_t))_{n\ge 1}$ is tight. By Theorem 3.2.12, it is enough to prove the tightness of $(\mathrm{Law}\,(f(Q^{\varepsilon_n}_t)))_{n\ge 1}$ for all $f \in C^\infty(\mathbb{T}^d)$. The latter fact will follow from Theorem 3.2.10, if we are able to construct, for any function $f \in C^\infty(\mathbb{T}^d)$ and any $\varepsilon, \delta > 0$ and any $T > 0$, a random variable $\Gamma_{\varepsilon,\delta}(f)$ such that for all $0 \le t \le t+h \le t+\delta \le T$, one has

$$\mathbb{E}\left[\left(f(Q^\varepsilon_{t+h}) - f(Q^\varepsilon_t)\right)^2 \middle| \mathcal{F}^{Q^\varepsilon}_t\right] \le \mathbb{E}\left[\Gamma_{\varepsilon,\delta}(f) \middle| \mathcal{F}^{Q^\varepsilon}_t\right], \tag{3.3.3}$$

$$\text{where} \quad \lim_{\delta\to 0} \limsup_{\varepsilon\ge 0} \mathbb{E}\left[\Gamma_{\varepsilon,\delta}(f)\right] = 0. \tag{3.3.4}$$

We claim that the following variant:

Lemma 3.3.6. *For any $g \in \mathcal{C}^\infty(\mathbb{T}^d)$, and any $\delta, \varepsilon, T > 0$, there exists a random variable*

$\Gamma'_{\varepsilon,\delta}(g)$ *such that for all* $0 \leq t \leq t+h \leq t+\delta \leq T$,

$$\left|\mathbb{E}\left[g(Q^\varepsilon_{t+h}) - g(Q^\varepsilon_t)\middle|\mathcal{F}^{Q^\varepsilon}_t\right]\right| \leq \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g)\middle|\mathcal{F}^{Q^\varepsilon}_t\right], \tag{3.3.5}$$

$$\textit{where} \qquad \lim_{\delta\to 0}\limsup_{\varepsilon\geq 0}\mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g)\right] = 0. \tag{3.3.6}$$

is a sufficient condition. Indeed, the required estimates (3.3.3), (3.3.4) will follow easily from the basic decomposition

$$\left(f(Q^\varepsilon_t) - f(Q^\varepsilon_{t+h})\right)^2 = \left(f(Q^\varepsilon_{t+h})\right)^2 - \left(f(Q^\varepsilon_t)\right)^2 - 2f(Q^\varepsilon_t)\left(f(Q^\varepsilon_{t+h}) - f(Q^\varepsilon_t)\right).$$

since we get

$$\mathbb{E}\left[\left(f(Q^\varepsilon_{t+h}) - f(Q^\varepsilon_t)\right)^2\middle|\mathcal{F}^{Q^\varepsilon}_t\right] \leq \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(f^2)\middle|\mathcal{F}^{Q^\varepsilon}_t\right] + 2\|f\|_\infty\mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(f)\middle|\mathcal{F}^{Q^\varepsilon}_t\right], \tag{3.3.7}$$

and it is enough to let $\Gamma_{\varepsilon,\delta}(f) = \Gamma'_{\varepsilon,\delta}(f^2) + 2\|f\|_\infty\Gamma'_{\varepsilon,\delta}(f)$ to conclude.

Let us now prove the Lemma 3.3.6. Let g be an arbitrary smooth function, and let g_ε be the perturbed test function given by Assumption 3.3.4. An elementary rewriting leads to

$$\begin{aligned} g(Q^\varepsilon_{t+h}) - g(Q^\varepsilon_t) = {} & \left(g(Q^\varepsilon_{t+h}) - g_\varepsilon(Q^\varepsilon_{t+h}, P^\varepsilon_{t+h})\right) - \left(g(Q^\varepsilon_t) - g_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)\right) \\ & - \int_t^{t+h}\left(Lg(Q^\varepsilon_s) - L_\varepsilon g_\varepsilon(Q^\varepsilon_s, P^\varepsilon_s)\right)ds + \int_t^{t+h} Lg(Q^\varepsilon_s)ds \\ & - M^\varepsilon_t(g_\varepsilon) + M^\varepsilon_{t+h}(g_\varepsilon), \end{aligned} \tag{3.3.8}$$

where $(M^\varepsilon_t(g_\varepsilon))_{t\geq 0}$ is a local $\mathcal{F}^{Q^\varepsilon,P^\varepsilon}$-martingale by Assumption 3.3.1. Let τ_n be an associated localizing sequence of stopping times. Applying (3.3.8) at times $t\wedge\tau_n$ and $(t+h)\wedge\tau_n$, we get

$$\begin{aligned} g(Q^\varepsilon_{(t+h)\wedge\tau_n}) - g(Q^\varepsilon_{t\wedge\tau_n}) = {} & g(Q^\varepsilon_{(t+h)\wedge\tau_n}) - g_\varepsilon\left(Q^\varepsilon_{(t+h)\wedge\tau_n}, P^\varepsilon_{(t+h)\wedge\tau_n}\right) - \left(g(Q^\varepsilon_{t\wedge\tau_n}) - g_\varepsilon(Q^\varepsilon_{t\wedge\tau_n}, P^\varepsilon_{t\wedge\tau_n})\right) \\ & - \int_t^{t+h}\left(Lg(Q^\varepsilon_s) - L_\varepsilon g_\varepsilon(Q^\varepsilon_s, P^\varepsilon_s)\right)\mathbf{1}_{s\leq\tau_n}ds + \int_t^{t+h} Lg(Q^\varepsilon_s)\mathbf{1}_{s\leq\tau_n}ds \\ & - M^\varepsilon_{t\wedge\tau_n}(g_\varepsilon) + M^\varepsilon_{(t+h)\wedge\tau_n}(g_\varepsilon). \end{aligned}$$

Taking the conditional expectation with respect to $\mathcal{F}^{Q^\varepsilon}_t$, the martingale terms cancel out, and we get:

$$
\begin{aligned}
&\left|\mathbb{E}\left[g(Q^\varepsilon_{(t+h)\wedge\tau_n}) - g(Q^\varepsilon_{t\wedge\tau_n})\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| \\
&\quad\le \left|\mathbb{E}\left[g(Q^\varepsilon_{(t+h)\wedge\tau_n}) - g_\varepsilon(Q^\varepsilon_{(t+h)\wedge\tau_n}, P^\varepsilon_{(t+h)\wedge\tau_n})\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| + \left|\mathbb{E}\left[g(Q^\varepsilon_{t\wedge\tau_n}) - g_\varepsilon(Q^\varepsilon_{t\wedge\tau_n}, P^\varepsilon_{t\wedge\tau_n})\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| \\
&\qquad + \int_t^{t+h}\left|\mathbb{E}\left[Lg(Q^\varepsilon_s) - L_\varepsilon g_\varepsilon(Q^\varepsilon_s, P^\varepsilon_s)|\mathcal{F}_t^{Q^\varepsilon}\right]\right| ds + h\sup_{q\in\mathbb{T}^d}|Lg(q)| \\
&\quad\le \mathbb{E}\left[R^\varepsilon_{1,(t+h)\wedge\tau_n} + R^\varepsilon_{1,t\wedge\tau_n}\middle|\mathcal{F}_t^{Q^\varepsilon}\right] + \int_t^{t+h}\mathbb{E}\left[R^\varepsilon_{2,s}|\mathcal{F}_t^{Q^\varepsilon}\right] ds + \delta\sup_{q\in\mathbb{T}^d}|Lg(q)| \\
&\quad\le 2\mathbb{E}\left[\sup_{s\in[0,T]} R^\varepsilon_{1,s}\middle|\mathcal{F}_t^{Q^\varepsilon}\right] + \int_0^T\mathbb{E}\left[R^\varepsilon_{2,s}|\mathcal{F}_t^{Q^\varepsilon}\right] ds + \delta\sup_{q\in\mathbb{T}^d}|Lg(q)|\,.
\end{aligned}
$$

The right hand side does not depend on n any longer. On the left hand side, we apply dominated convergence for $n\to\infty$ to get

$$
\left|\mathbb{E}\left[g(Q^\varepsilon_{(t+h)}) - g(Q^\varepsilon_t)\middle|\mathcal{F}_t^{Q^\varepsilon}\right]\right| \le \mathbb{E}\left[\Gamma'_{\varepsilon,\delta}(g)\middle|\mathcal{F}_t^{Q^\varepsilon}\right]
$$

for $\Gamma'_{\varepsilon,\delta}(g) = 2\sup_{[0,T]} R^\varepsilon_{1,t} + \int_0^T R^\varepsilon_{2,s}ds + \delta\|Lg\|_\infty$. The controls on the rest terms given by Assumption 3.3.4, and the continuity of Lg (Assumption 3.3.2) ensure that $\lim_{\delta\to 0}\limsup_{\varepsilon\to 0}\Gamma'_{\varepsilon,\delta}(g) = 0$, and the proof of tightness is concluded.

3.3.2.2 Step two: identification of the limit

In this step, we suppose that a sequence $Q^n_t = Q^{\varepsilon_n}_t$ converges in distribution to a limit Q^0_t, and we prove that necessarily, Q^0 solves the martingale problem for the generator L. Let $f\in C^\infty(\mathbb{T}^d)$, we have to check that

$$
M_t(Q^0_t) := f(Q^0_t) - f(Q^0_0) - \int_0^t Lf(Q^0_s)ds \tag{3.3.9}
$$

is a martingale with respect to $\mathcal{F}_t^{Q^0} = \sigma(Q^0_s, 0\le s\le t)$. Consider a time sequence $0\le t_1\le\cdots\le t_p\le t_{p+1}$ for $p\ge 1$, taken in the continuity set $\mathcal{C}_{\mathrm{Law}(Q)}$ given by Lemma 3.2.14. Recall that $\mathcal{C}_{\mathrm{Law}(Q)}$ is dense in $\mathbb{R}$. Let $\varphi_1,...,\varphi_p\in C_b(\mathbb{T}^d)$ be p test functions. By Lemma 3.2.15, it is enough to prove that

$$
I_0 := \mathbb{E}\left[\left(f(Q^0_{t_{p+1}}) - f(Q^0_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^0_s)ds\right)\varphi_1(Q^0_{t_1})\cdots\varphi_p(Q^0_{t_p})\right] = 0.
$$

Let I_ε be the corresponding quantity for $\varepsilon>0$, that is,

$$
I_\varepsilon := \mathbb{E}\left[\left(f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s)ds\right)\varphi_1(Q^\varepsilon_{t_1})\cdots\varphi_p(Q^\varepsilon_{t_p})\right].
$$

Let us first show that I_ε converges to 0. We first condition on $\mathcal{F}_{t_p}^{Q^\varepsilon}$ to get:

$$|I_\varepsilon| \leq \mathbb{E}\left[\mathbb{E}\left[\left|f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s)ds\right| \left|\mathcal{F}_{t_p}^{Q^\varepsilon}\right.\right] \left|\varphi_1(Q^\varepsilon_{t_1})\right| \cdots \left|\varphi_p(Q^\varepsilon_{t_p})\right|\right]$$
$$\leq \mathbb{E}\left[\mathbb{E}\left[\left|f(Q^\varepsilon_{t_{p+1}}) - f(Q^\varepsilon_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(Q^\varepsilon_s)ds\right| \left|\mathcal{F}_{t_p}^{Q^\varepsilon}\right.\right]\right] \|\varphi_1\|_\infty \cdots \|\varphi_p\|_\infty.$$

Using again the perturbed test function f_ε and the decomposition (3.3.8), we get by the same localization argument as in Step 1 that

$$|I_\varepsilon| \leq \mathbb{E}\left[R^\varepsilon_{1,t_{p+1}}(f) + R^\varepsilon_{1,t_p}(f) + \int_{t_p}^{t_{p+1}} R^\varepsilon_{2,s}(f)\right] \|\varphi_1\|_\infty ... \|\varphi_p\|_\infty.$$

The estimates on the rest term from Assumption 3.3.4 then imply that $I_\varepsilon \to 0$.

Let us now prove that I_ε converges to I_0. Let $\Phi : \mathbb{D}_{\mathbb{T}^d} \to \mathbb{R}$ be the functional

$$\Phi : (q_t)_{t\geq 0} \mapsto \left(f(q_{t_{p+1}}) - f(q_{t_p}) - \int_{t_p}^{t_{p+1}} Lf(q_s)ds\right) \varphi_1(q_{t_1}) \cdots \varphi_p(q_{t_p})$$

so that $I_\varepsilon = \mathbb{E}\left[\Phi((Q^\varepsilon_t)_{t\geq 0})\right]$ and $I_0 = \mathbb{E}\left[\Phi((Q^0_t)_{t\geq 0})\right]$. Let us first check that, if $q^0 \in \mathbb{D}_{\mathbb{T}^d}$ satisfies $q^0_{t_k^-} = q^0_{t_k}$ for each $1 \leq k \leq p+1$, then the functional Φ is continuous at the trajectory q^0. Indeed, since Lf is continuous and bounded by Assumption 3.3.2, Lemma 3.2.1 shows that the map $(q_t)_{t\geq 0} \mapsto \int_{t_p}^{t_{p+1}} Lf(q_s)ds$ is continuous with respect to Skorokhod topology; moreover, by assumption, q^0 is continuous at the time t_k for each $1 \leq k \leq p+1$, so the map $(q_t)_{t\geq 0} \mapsto \varphi_k(q_{t_k})$ is continuous at $q^0 \in \mathbb{D}_{\mathbb{T}^d}$.

Let now $(\varepsilon_n)_{n\geq 1}$ be any sequence such that $\varepsilon_n \to 0$ and $(Q^{\varepsilon_n}_t)_{t\geq 0}$ converges in distribution to $(Q^0_t)_{t\geq 0}$. The Skorokhod representation theorem (Theorem 1.8 in [EK86, Chapter 3]) ensures that one can construct a probability space where the distribution of $(Q^{\varepsilon_n}_t)_{t\geq 0}$ for each n is unchanged but for which $\lim_{n\to+\infty} Q^{\varepsilon_n} = Q^0$ almost surely in $\mathbb{D}_{\mathbb{T}^d}$. Since $t_k \in \mathcal{C}_{\mathrm{Law}(Q^0)}$ for each $k = 1 \dots p+1$, Ψ is almost surely continuous at Q^0 and we can apply the dominated convergence theorem to obtain $\lim_{n\to+\infty} I_{\varepsilon_n} = I_0$. Since the choice of the vanishing sequence $(\varepsilon_n)_{n\geq 1}$ is arbitrary, we conclude that $\lim_{\varepsilon\to 0} I_\varepsilon = I_0$. The limit process thus solves the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$.

3.3.2.3 Conclusion.

For each sequence $(\varepsilon_n)_{n\geq 1}$ satisfying $\lim_n \varepsilon_n = 0$, we have proven that $(Law(Q^{\varepsilon_n}_t))_{n\geq 1}$ is tight and that any converging subsequence is solution to the martingale problem $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$. By uniqueness of the latter according to Assumption 3.3.2, this identifies the limit, showing that $(Law(Q^{\varepsilon_n}_t))_{n\geq 1}$ converges to the solution of $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$. Since the sequence $(\varepsilon_n)_{n\geq 1}$ is arbitrary and convergence in distribution is metrizable, $(Law(Q^\varepsilon_t))_{\varepsilon>0}$ also con-

verges to the solution of $\mathbf{MP}(L, C^\infty(\mathbb{T}^d), \mu)$, proving Theorem 3.3.5.

3.4 Overdamped limit of the Langevin dynamics

In the section, we will use the perturbed test function method presented in last section to prove Theorem 3.1.1. We will first state the key estimates on $(|P_t^\varepsilon|)_{t\geq 0}$. These estimates are then used to check the assumptions of our general Theorem 3.3.5 in the specific case of Langevin processes. In a last section we will detail the proof of the key estimates.

3.4.1 Some moments estimates for Langevin processes

We start by giving a few facts about the solution to the Langevin SDE (3.1.1). We first check that the operator L_ε acting on $\mathcal{C}^\infty(\mathbb{T}^d, \mathbb{R}^d)$ by

$$L_\varepsilon f(q,p) := \frac{1}{\varepsilon^2}\left(\frac{1}{\beta}\Delta_p f - p\cdot\nabla_p f\right) + \frac{1}{\varepsilon}\left(p\cdot\nabla_q f - \nabla_q V_\varepsilon\cdot\nabla_p f\right)$$

is the generator the process, in the sense that Assumption 3.3.1 holds.

Proposition 3.4.1. *If $(Q_t^\varepsilon, P_t^\varepsilon)_{t\geq 0}$ is a weak solution of the Langevin SDE (3.1.1), then for any smooth function $f : \mathbb{T}^d \times \mathbb{R}^d \to \mathbb{R}^d$, the process*

$$t \mapsto M_t^\varepsilon(f) = f(Q_t^\varepsilon, P_t^\varepsilon) - f(Q_0^\varepsilon, P_0^\varepsilon) - \int_0^t L_\varepsilon f(Q_s^\varepsilon, P_s^\varepsilon)ds,$$

is a $(\mathcal{F}_t^{Q^\varepsilon,P^\varepsilon})_{t\geq 0}$-local martingale.

Proof. This is a very classical result. By Itô calculus we write

$$df_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon) = L_\varepsilon f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)dt + \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\nabla_p f_\varepsilon(Q_t^\varepsilon, P_t^\varepsilon)dW_t.$$

Defining the sequence of $\left(\mathcal{F}_t^{Q^\varepsilon,P^\varepsilon}\right)_{t\geq 0}$-stopping time

$$\tau_n = \inf\{t\geq 0, |P_t^\varepsilon| \geq n\}, \tag{3.4.1}$$

which converge almost surely to infinity, we obtain that

$$M_t^{\varepsilon,n}(f_\varepsilon) := \frac{1}{\varepsilon}\sqrt{2\beta^{-1}}\int_0^t \nabla_p f_\varepsilon(Q_s^\varepsilon, P_s^\varepsilon)1_{s\leq\tau_n}dW_s$$

is a $\left(\mathcal{F}_t^{Q^\varepsilon,P^\varepsilon}\right)_{t\geq 0}$-martingale for any $n\geq 0$, which is the definition of a local martingale. □

We now state several bounds on the momentum variable P_t^ε, which are the key technical estimates needed later to control the rest terms appearing in the perturbed test function

method. For any continuous $V : \mathbb{T}^d \to \mathbb{R}$ we denote by $\mathrm{osc}(V)$ the oscillation defined by

$$\mathrm{osc}(V) = \max V - \min V.$$

Lemma 3.4.2 (Propagation of moments)**.** *For any $\gamma \geq 1$, any $M > 0$ and any $\beta > 0$, there is a numerical constant $C(\gamma, M, \beta)$ such that for any $\varepsilon > 0$, if $\mathrm{osc}(V_\varepsilon) \leq M$, then*

$$\sup_{t \geq 0} \mathbb{E}\left[|P_t^\varepsilon|^{2\gamma}\right] \leq C(\gamma, M, \beta)\left(\mathbb{E}\left[|P_0^\varepsilon|^{2\gamma}\right] + 1\right). \tag{3.4.2}$$

Lemma 3.4.3 (Moment of suprema)**.** *For any $M > 0$, any $\beta > 0$ and any $T > 0$, there is a numerical constant $C(M, \beta, T)$ such that for any $\varepsilon \in (0, 1)$, if $\mathrm{osc}(V_\varepsilon) \leq M$, then*

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} |P_t^\varepsilon|^2\right] \leq \mathbb{E}\left[|P_0^\varepsilon|^2\right] + \frac{1}{\varepsilon} C(M, \beta, T)\left(\mathbb{E}\left[|P_0^\varepsilon|^2\right] + 1\right)^{1/2}. \tag{3.4.3}$$

In particular, if $\lim_{\varepsilon \to 0} \varepsilon^2 \mathbb{E}\left[|P_0^\varepsilon|^2\right] = 0$, then

$$\lim_{\varepsilon \to 0} \varepsilon^2 \mathbb{E}\left[\sup_{0 \leq t \leq T} |P_t^\varepsilon|^2\right] = 0.$$

The proofs of these estimates use classical techniques of stochastic calculus and are postponed to Section 3.4.3.

3.4.2 The perturbed test functions in the Langevin case

In this section we apply the general method described in Section 3.3 to the specific Langevin case, in order to prove Theorem 3.1.1.

We will use the following standard notation for multidimensional derivatives:

$$\nabla^k f(p_1, \ldots, p_k) := \sum_{i_1, \ldots, i_k = 1}^{d} \partial_{i_1} \ldots \partial_{i_k} f \times p_1^{i_1} \times \ldots \times p_k^{i_k}$$

where in the above $p_1, \ldots, p_k \in \mathbb{R}^d$. Note that as usual $\Delta f = \mathrm{Tr}\left(\nabla^2 f\right)$.

We first construct explicitly, for any $f \in C^\infty(\mathbb{T}^d)$, a perturbed test function $f_\varepsilon \in C^\infty(\mathbb{T}^d \times \mathbb{R}^d)$. Let us look for f_ε in the following form (see [PSV77])

$$f_\varepsilon(q, p) = f(q) + \varepsilon g_1(q, p) + \varepsilon^2 g_2(q, p). \tag{3.4.4}$$

Applying the generator L_ε, using the fact that f does not depend on p, and grouping

terms with respect to powers of ε, we get

$$\begin{aligned} L_\varepsilon f_\varepsilon(q,p) &= \frac{1}{\varepsilon} p\cdot\nabla_q[f(q)+\varepsilon g_1(q,p)+\varepsilon^2 g_2(q,p)] - \frac{1}{\varepsilon}\nabla_q V(q)\cdot\nabla_p[\varepsilon g_1(q,p)+\varepsilon^2 g_2(q,p)] \\ &\quad - \frac{1}{\varepsilon^2} p\cdot\nabla_p[\varepsilon g_1(q,p)+\varepsilon^2 g_2(q,p)] + \frac{1}{\varepsilon^2\beta}\Delta_p[\varepsilon g_1(q,p)+\varepsilon^2 g_2(q,p)] \\ &= \frac{1}{\varepsilon}\left(p\cdot\nabla_q f - p\cdot\nabla_p g_1 + \frac{1}{\beta}\Delta_p g_1\right) \\ &\quad + \left(p\cdot\nabla_q g_1 - \nabla_q V_\varepsilon\cdot\nabla_p g_1 - p\cdot\nabla_p g_2 + \frac{1}{\beta}\Delta_p g_2\right) \\ &\quad + \varepsilon\left(p\cdot\nabla_q g_2 - \nabla_p g_2\cdot\nabla_q V_\varepsilon\right). \end{aligned} \tag{3.4.5}$$

In order for $L_\varepsilon f_\varepsilon$ to converge to Lf, the ε^{-1}-order terms should vanish, and the ε^0-order terms should converge at least formally to $L(f)$. As a consequence g_1 and g_2 should solve the following equations:

$$0 = p\cdot\nabla_q f - p\cdot\nabla_p g_1 + \frac{1}{\beta}\Delta_p g_1, \tag{3.4.6}$$

$$Lf(q) = p\cdot\nabla_q g_1 - \nabla_q V\cdot\nabla_p g_1 - p\cdot\nabla_p g_2 + \frac{1}{\beta}\Delta_p g_2.. \tag{3.4.7}$$

The function $g_1(q,p) = p\cdot\nabla_q f(q)$ clearly solves (3.4.6). With this choice, (3.4.7) becomes

$$Lf(q) = \nabla_q^2 f(p,p) - \nabla_q V\cdot\nabla_q f - p\cdot\nabla_p g_2 + \frac{1}{\beta}\Delta_p g_2.$$

Since $Lf(q) = \frac{1}{\beta}\Delta_q f - \nabla_q V\cdot\nabla_q f$, it is easy to check that $g_2(q,p) = \frac{1}{2}\nabla_q^2 f(p,p)$ solves the equation.

Therefore, in view of Eq. (3.4.4), we defined the perturbed test function by :

$$f_\varepsilon(q,p) = f(q) + \varepsilon p\cdot\nabla_q f + \frac{1}{2}\varepsilon^2\nabla_q^2 f(p,p). \tag{3.4.8}$$

With this choice, we get using previous calculations and the last line of (3.4.5)

$$\begin{aligned} &L_\varepsilon f_\varepsilon(q,p) - Lf(q) \\ &\quad = (\nabla_q V - \nabla_q V_\varepsilon)\cdot\nabla_q f + \varepsilon\left(p\cdot\nabla_q g_2 - \nabla_p g_2\cdot\nabla_q V_\varepsilon\right) \\ &\quad = (\nabla_q V - \nabla_q V_\varepsilon)\cdot\nabla_q f + \frac{1}{2}\varepsilon\left(\nabla_q^3 f(q)(p,p,p) - \nabla_q^2 f\left(p,\nabla_q V_\varepsilon\right)\right). \end{aligned} \tag{3.4.9}$$

We now need to show that Assumption 3.3.4 holds for this choice of a perturbed test function, that is, we want to show that the differences $f_\varepsilon - f$ and $L_\varepsilon f_\varepsilon - Lf$ are small in the following appropriate sense. Recalling the notation

$$R^\varepsilon_{1,t}(f) = \left|f(Q^\varepsilon_t) - f_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)\right|, \qquad R^\varepsilon_{2,t}(f) = \left|Lf(Q^\varepsilon_t) - L_\varepsilon f_\varepsilon(Q^\varepsilon_t, P^\varepsilon_t)\right|,$$

we need to prove that

$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\sup_{0\le t\le T} R^{\varepsilon}_{1,t}(f)\Big)=0, \tag{3.4.10}$$

$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\int_0^T R^{\varepsilon}_{2,t}(f)dt\Big)=0. \tag{3.4.11}$$

Since $f\in C^\infty(\mathbb{T}^d)$, there exists a $C_f=\max\left(\|\nabla f\|_\infty,\|\nabla^2 f\|_\infty\right)$ such that for all (q,p) and all $\delta\in(0,1/2)$

$$\begin{aligned}|f_\varepsilon(q,p)-f(q)| &= \varepsilon\,|p\cdot\nabla_q f(q)|+\frac{1}{2}\varepsilon^2\left|\nabla_q^2 f(q)\cdot(p,p)\right|\\ &\le C_f(\varepsilon\,|p|+\varepsilon^2\,|p|^2)\\ &\le \delta C_f+\frac{1}{\delta}C_f\varepsilon^2\,|p|^2\,,\end{aligned}$$

where we have used that for any $\delta>0$, $\varepsilon\,|p|\le\frac{1}{2}\delta+\frac{1}{2}\varepsilon^2\,|p|^2/\delta$. Therefore

$$\mathbb{E}\left[\sup_{t\in[0,T]}R^{\varepsilon}_{1,t}(f)\right]\le\delta C_f+\frac{1}{\delta}C_f\varepsilon^2\mathbb{E}\left[\sup_{t\in[0,T]}|P^{\varepsilon}_t|^2\right].$$

By assumption, $\lim_{\varepsilon\to0}\varepsilon\mathbb{E}\left[|P^{\varepsilon}_0|^3\right]=0$, so $\varepsilon^2\mathbb{E}\left[|P^{\varepsilon}_0|^2\right]\le\varepsilon^{4/3}(\varepsilon\mathbb{E}\left[|P^{\varepsilon}_0|^3\right])^{2/3}$ also goes to zero by Jensen's inequality. By the key Lemma 3.4.3 this entails that the last term in the previous display disappears in the limit and we get

$$\limsup_{\varepsilon\to0}\mathbb{E}\left[\sup_{t\in[0,T]}R^{\varepsilon}_{1,t}(f)\right]\le\delta C_f,$$

which proves (3.4.10) since δ is arbitrary.

We now turn to the proof of (3.4.11), that is, we want to compare $L_\varepsilon f_\varepsilon$ and Lf. By the expression (3.4.9), we have for some constant $C_f=\max\left(\|\nabla f\|_\infty,\|\nabla^2 f\|_\infty,\|\nabla^3 f\|_\infty\right)$

$$|L_\varepsilon f_\varepsilon(q,p)-Lf(q)|\le C_f\|\nabla_q V-\nabla_q V_\varepsilon\|_\infty+C_f\varepsilon\left(|p|^3+\|\nabla_q V_\varepsilon\|_\infty|p|\right).$$

We get rid of the product term with Young's inequality $ab\le a^3/3+\frac{2}{3}b^{3/2}\le a^3+b^{3/2}$ and get

$$\mathbb{E}\left[R^{\varepsilon}_{2,t}\right]\le C_f\|\nabla_q V-\nabla_q V_\varepsilon\|_\infty+\varepsilon C_f\mathbb{E}\left[2\,|P^{\varepsilon}_t|^3+\|\nabla V_\varepsilon\|_\infty^{3/2}\right].$$

We integrate in t to obtain

$$\int_0^T\mathbb{E}\left[R^{\varepsilon}_{2,t}\right]dt\le C_f\|\nabla_q V-\nabla_q V_\varepsilon\|_\infty T+\varepsilon C_f T\left(\sup_{t\in[0,T]}\mathbb{E}\left[2\,|P^{\varepsilon}_t|^3\right]+\|\nabla V_\varepsilon\|_\infty^{3/2}\right).$$

By assumption, $\lim_{\varepsilon\to 0}\varepsilon\mathbb{E}\left[|P_0^\varepsilon|^3\right]=0$, and by the uniform convergence of ∇V_ε to ∇V we can find a uniform bound M such that $\mathrm{osc}(V_\varepsilon)\leq M$ for all ε, so we may apply Lemma 3.4.2 with $\gamma=3/2$ and get

$$\lim_{\varepsilon\to 0}\varepsilon\sup_{t\in[0,T]}\mathbb{E}\left[|P_t^\varepsilon|^3\right]=0,$$

for any $T\geq 0$. Together with the convergence of ∇V_ε to ∇V this yields

$$\lim_{\varepsilon\to 0}\int_0^T\mathbb{E}\left[R_{2,t}^\varepsilon\right]dt=0.$$

from which (3.4.11) follows.

3.4.3 Proofs of the moment bounds

We now come back to the proofs of the moment bounds (Lemmas 3.4.2 and 3.4.3). It will prove useful to work with the Hamiltonian of the system rather than directly with P_t^ε. For convenience's sake we assume without loss of generality that $0\leq V_\varepsilon(q)\leq\mathrm{osc}(V_\varepsilon)$.

Definition 3.4.4 (Hamiltonian)**.** *We denote by H^ε the Hamiltonian of the system:*

$$H^\varepsilon(q,p)=\frac{1}{2}|p|^2+V_\varepsilon(q).$$

We will also write $H_t^\varepsilon:=H^\varepsilon(Q_t^\varepsilon,P_t^\varepsilon)$.

By Itô's formula,

$$\begin{aligned}dH_t^\varepsilon&=P_t^\varepsilon dP_t^\varepsilon+\nabla_q V_\varepsilon(Q_t^\varepsilon)dQ_t^\varepsilon+\frac{1}{2}\sum_{i,j=1}^d d\langle(P^\varepsilon)^i,(P^\varepsilon)^j\rangle_t\\&=\left(-\frac{1}{\varepsilon^2}|P_t^\varepsilon|^2+\frac{1}{\varepsilon^2\beta}\right)dt+\frac{1}{\varepsilon}\sqrt{2\beta^{-1}}P_t^\varepsilon dW_t\qquad(3.4.12)\\&=\left(-\frac{2}{\varepsilon^2}H_t^\varepsilon+\frac{2}{\varepsilon^2}V_\varepsilon(Q_t^\varepsilon)+\frac{1}{\varepsilon^2\beta}\right)dt+\frac{\sqrt{2\beta^{-1}}}{\varepsilon}P_t^\varepsilon dW_t.\qquad(3.4.13)\end{aligned}$$

Again, by Itô's formula, we thus get for any smooth function $(t,h)\mapsto\phi(t,h)$

$$d\phi(t,H_t^\varepsilon)=\partial_t\phi(t,H_t^\varepsilon)dt+\partial_h\phi(t,H_t^\varepsilon)dH_t^\varepsilon+\frac{1}{\varepsilon^2\beta}\partial_h^2\phi(t,H_t^\varepsilon)|P_t^\varepsilon|^2dt.\qquad(3.4.14)$$

Proof of Lemma 3.4.2. Let $\gamma\geq 1$. We apply (3.4.14) to $\phi(t,x)=e^{\alpha t}h^\gamma$ and plug in (3.4.13) to get:

$$\begin{aligned}d(e^{\alpha t}(H_t^\varepsilon)^\gamma)=\gamma(H_t^\varepsilon)^{\gamma-1}&\left(\frac{\alpha}{\gamma}H_t^\varepsilon-\frac{2}{\varepsilon^2}H_t^\varepsilon+\frac{2}{\varepsilon^2}V_\varepsilon(Q_t^\varepsilon)+\frac{1}{\varepsilon^2\beta}\right)e^{\alpha t}dt\\&+\frac{\sqrt{2\beta^{-1}}}{\varepsilon}\gamma(H_t^\varepsilon)^{\gamma-1}P_t^\varepsilon e^{\alpha t}dW_t+\frac{\gamma(\gamma-1)}{\varepsilon^2\beta}(H_t^\varepsilon)^{\gamma-2}|P_t^\varepsilon|^2e^{\alpha t}dt.\end{aligned}$$

The choice

$$\alpha = 2\gamma/\varepsilon^2$$

cancels the higher order term in the first bracket. We integrate in time, multiply by $e^{-\alpha t}$ and regroup the finite variation terms to get:

$$(H_t^\varepsilon)^\gamma = (H_0^\varepsilon)^\gamma + \int_0^t \left(\gamma(H_s^\varepsilon)^{\gamma-1}\left(\frac{2}{\varepsilon^2}V_\varepsilon(Q_s^\varepsilon) + \frac{1}{\varepsilon^2\beta}\right) + \frac{\gamma(\gamma-1)}{\varepsilon^2\beta}(H_s^\varepsilon)^{\gamma-2}\left|P_s^\varepsilon\right|^2\right)e^{-\alpha(t-s)}ds$$
$$+ \frac{\sqrt{2\beta^{-1}}}{\varepsilon}\int_0^t \gamma(H_s^\varepsilon)^{\gamma-1}P_s^\varepsilon e^{-\alpha(t-s)}dW_s.$$

Since $(1/2)\left|P_s^\varepsilon\right|^2 \le H_s^\varepsilon \le (1/2)\left|P_s^\varepsilon\right|^2 + \mathrm{osc}(V_\varepsilon)$,

$$\begin{aligned}(H_t^\varepsilon)^\gamma \le (H_0^\varepsilon)^\gamma + \frac{2\gamma}{\varepsilon^2}\left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma}{\beta}\right)\int_0^t (H_s^\varepsilon)^{\gamma-1}e^{-\alpha(t-s)}ds \\ + \frac{\sqrt{2\beta^{-1}}}{\varepsilon}\int_0^t (H_s^\varepsilon)^{\gamma-1}P_s^\varepsilon e^{-\alpha(t-s)}dW_s.\end{aligned} \tag{3.4.15}$$

To deal with the unboundedness of the momentum P, we define the following stopping times:

$$\tau_n := \inf\{t : |P_t| = n\}. \tag{3.4.16}$$

When $s \le \tau_n$, we have $|P_s^\varepsilon| \le n$ and $H_s^\varepsilon \le (\mathrm{osc}(V_\varepsilon) + \frac{n^2}{2})$. This entails that $t \mapsto \int_0^{t\wedge\tau_n}(H_s^\varepsilon)^{\gamma-1}P_s^\varepsilon dW_s$ is martingale. Writing (3.4.15) at time $t\wedge\tau_n$ and taking expectations, the martingale part disappears; recalling that $\alpha = 2\gamma/\varepsilon^2$ we get

$$\begin{aligned}\mathbb{E}\left[(H_{t\wedge\tau_n}^\varepsilon)^\gamma\right] &\le \mathbb{E}\left[(H_0^\varepsilon)^\gamma\right] + \left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma}{\beta}\right)\alpha\mathbb{E}\left[\int_0^{t\wedge\tau_n}(H_s^\varepsilon)^{\gamma-1}e^{-\alpha(t-s)}ds\right] \\ &\le \mathbb{E}\left[(H_0^\varepsilon)^\gamma\right] + \left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma}{\beta}\right)\sup_{s\le t}\mathbb{E}\left[(H_s^\varepsilon)^{\gamma-1}\right]ds.\end{aligned}$$

Sending n to infinity, we apply Fatou's lemma to get

$$\mathbb{E}\left[(H_t^\varepsilon)^\gamma\right] \le \mathbb{E}\left[(H_0^\varepsilon)^\gamma\right] + \left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma}{\beta}\right)\sup_{s\le t}\mathbb{E}\left[(H_s^\varepsilon)^{\gamma-1}\right],$$

and thus

$$\sup_{t\ge 0}\mathbb{E}\left[(H_t^\varepsilon)^\gamma\right] \le \mathbb{E}\left[(H_0^\varepsilon)^\gamma\right] + \left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma}{\beta}\right)\sup_{t\ge 0}\mathbb{E}\left[(H_s^\varepsilon)^{\gamma-1}\right]. \tag{3.4.17}$$

We are now ready to conclude. Say that γ is good if there exists a $C(\gamma, M, \beta)$ such that for all ε,

$$\sup_t \mathbb{E}\left[(H_t^\varepsilon)^\gamma\right] \le C(\gamma, M, \beta)(1 + \mathbb{E}\left[(H_0^\varepsilon)^\gamma\right],$$

whenever $\mathrm{osc}(V_\varepsilon) \le M$. The bound (3.4.17) immediately shows that $\gamma = 1$ is good. If γ is

good and $\gamma \leq \gamma' \leq \gamma + 1$, using the elementary inequality $x^a \leq 1 + x^b$ valid for any $x > 0$ and any $1 \leq a < b$, we get

$$\begin{aligned}
\sup_{t\geq 0} \mathbb{E}\left[(H_t^\varepsilon)^{\gamma'}\right] &\leq \mathbb{E}\left[(H_0^\varepsilon)^{\gamma'}\right] + \left(\mathrm{osc}(V_\varepsilon) + \frac{\gamma'}{\beta}\right) \sup_{t\geq 0} \mathbb{E}\left[(H_s^\varepsilon)^{\gamma'-1}\right] \\
&\leq \mathbb{E}\left[(H_0^\varepsilon)^{\gamma'}\right] + \left(M + \frac{\gamma'}{\beta}\right)\left(1 + \sup_{t\geq 0} \mathbb{E}\left[(H_s^\varepsilon)^{\gamma}\right]\right) \\
&\leq \mathbb{E}\left[(H_0^\varepsilon)^{\gamma'}\right] + \left(M + \frac{\gamma'}{\beta}\right)\left(1 + C(\gamma, M, \beta)\mathbb{E}\left[(H_0^\varepsilon)^{\gamma}\right]\right) \\
&\leq \mathbb{E}\left[(H_0^\varepsilon)^{\gamma'}\right] + \left(M + \frac{\gamma'}{\beta}\right)\left(1 + C(\gamma, M, \beta)\left(1 + \mathbb{E}\left[(H_0^\varepsilon)^{\gamma'}\right]\right)\right)
\end{aligned}$$

showing that γ' is itself good. Therefore all $\gamma \geq 1$ are good. Using the bounds $(1/2)p^2 \leq H^\varepsilon(q,p) \leq (1/2)p^2 + M$ it is easy to translate this into bounds on $\mathbb{E}\left[|P_t^\varepsilon|^{2\gamma}\right]$, concluding the proof of Lemma 3.4.2. □

Proof of Lemma 3.4.3. Let us fix an arbitrary $T > 0$, and prove (3.4.3), that is, prove the existence of a numerical constant $C(\beta, M, T)$ such for any $\varepsilon \in (0, 1)$,

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |P_t^\varepsilon|^2\right] \leq \mathbb{E}\left[|P_0^\varepsilon|^2\right] + \frac{1}{\varepsilon} C(\beta, M, T)\left(\mathbb{E}\left[|P_0^\varepsilon|^2\right] + 1\right)^{1/2} \tag{3.4.18}$$

whenever $\mathrm{osc}(V_\varepsilon) \leq M$. As before, since $2H_t^\varepsilon - 2M \leq (P_t^\varepsilon)^2 \leq 2H_t^\varepsilon$, it is enough to prove the statement with H_t^ε instead of $|P_t^\varepsilon|^2$.

We start by recalling (3.4.15) for $\gamma = 1$ and $\alpha = 2/\varepsilon^2$:

$$H_t^\varepsilon \leq H_0^\varepsilon + \left(\mathrm{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} \int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s. \tag{3.4.19}$$

Recall that this led by a localization argument to the following bound (3.4.17):

$$\sup_{t\geq 0} \mathbb{E}\left[H_t^\varepsilon\right] \leq \mathbb{E}\left[H_0^\varepsilon\right] + \left(M + \frac{1}{\beta}\right). \tag{3.4.20}$$

In order to control the expectation of the supremum, we must control the stochastic integral. Define $M_t = \int_0^t P_s^\varepsilon dW_s$ and integrate by parts:

$$\begin{aligned}
\left|\int_0^t e^{-\alpha(t-s)} P_s^\varepsilon dW_s\right| = \left|\int_0^t e^{-\alpha(t-s)} dM_s\right| &= \left|M_t - \alpha \int_0^t e^{-\alpha(t-s)} M_s ds\right| \\
&\leq |M_t| + \sup_{s\in[0,t]} |M_s| \\
&\leq 2 \sup_{s\in[0,T]} |M_s|.
\end{aligned}$$

Plugging this in (3.4.19) yields

$$\sup_{t\in[0,T]} H_t \leq H_0^\varepsilon + \left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + \frac{\sqrt{2\beta^{-1}}}{\varepsilon} 2 \sup_{t\in[0,T]} |M_t| \,. \tag{3.4.21}$$

By Doob's martingale maximal inequality, Itô's isometry and the bound (3.4.20) we get

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T} |M_t^\varepsilon|^2\right] &\leq 4\mathbb{E}\left[|M_T^\varepsilon|^2\right] = 4\mathbb{E}\left[\left|\int_0^T P_s^\varepsilon dW_s\right|^2\right] = 4\mathbb{E}\left[\int_0^T (P_s^\varepsilon)^2 ds\right] \\
&\leq 8T\left(\operatorname{osc}(V_\varepsilon) + \sup_{t\in[0,T]} \mathbb{E}\left[H_t^\varepsilon\right]\right) \\
&\leq 8T\left(2\operatorname{osc}(V_\varepsilon) + \mathbb{E}\left[H_0^\varepsilon\right] + \frac{1}{\beta}\right).
\end{aligned}$$

Injecting this in (3.4.21) and applying Cauchy–Schwarz inequality yields

$$\mathbb{E}\left[\sup_{t\in[0,T]} H_t\right] \leq \mathbb{E}\left[H_0^\varepsilon\right] + \left(\operatorname{osc}(V_\varepsilon) + \frac{1}{\beta}\right) + 8\frac{\sqrt{T\beta^{-1}}}{\varepsilon}\left(2\operatorname{osc}(V_\varepsilon) + \mathbb{E}\left[H_0^\varepsilon\right] + \frac{1}{\beta}\right)^{1/2}, \tag{3.4.22}$$

concluding the proof of (3.4.18). □

Part III

Reducing variance by reweighting samples

Chapter 4

Reducing variance by reweighting samples

4.1 Introduction

4.1.1 The framework

Let (X, Y) be a couple of random variables, and say that we are interested in computing the expected value $\mathbb{E}[\phi(Y)]$ — or more generally $\mathbb{E}[\phi(X, Y)]$ — for each ϕ in some class of test functions. Since the distribution of $\phi(X, Y)$ is most often impossible to obtain in closed analytic form, a classical approach is to resort to Monte-Carlo integration: given an iid sample $(\mathbf{X}; \mathbf{Y}) = (X_1, ..., X_N; Y_1, ..., Y_N)$, the usual "naïve" Monte Carlo estimator is

$$\mathbf{\Phi}_{MC}(\mathbf{X}, \mathbf{Y}) = \frac{1}{N} \sum_{n=1}^{N} \phi(X_n, Y_n). \tag{4.1.1}$$

This estimator is unbiased, and its mean square error is given by its variance:

$$\mathrm{MSE}(\mathbf{\Phi}_{MC}) := \mathbb{E}\left[(\mathbf{\Phi}_{MC}(\mathbf{X}, \mathbf{Y}) - \mathbb{E}[\phi(X, Y)])^2\right] = \frac{1}{N} \mathrm{Var}(\phi(X, Y)).$$

The behaviour in N is inescapable and given by the CLT; however, over the years, many *variance reduction* techniques have been devised to reduce the constant multiplicative factor, using various kinds of additional hypotheses on the couple (X, Y). For a general overview of these techniques, see for example the survey paper of Glynn [Gly94], or the book [Ros13]. We introduce in this paper new techniques for reducing variance, which can be seen as a variation on the classical post-stratification method, except that we do not have to fix strata. The method is based on two assumptions on the distribution of the couple (X, Y).

Assumption 4.1.1. *The distribution of the first marginal X is exactly known: $X \sim \gamma :=$*

$\mathcal{N}(0,1)$, *the standard Gaussian distribution.*

We note that we could easily accomodate other distributions than the standard Gaussian, which we consider for simplicity; the main point is that we know the distribution of X.

Before introducing the second assumption, let us first recall a classical decomposition of the variance. If we denote by

$$M_\phi(X) := \mathbb{E}\left[\phi(X,Y)|X\right], \qquad V_\phi(X) := \mathbb{E}\left[(\phi(X,Y) - M_\phi(X))^2 \middle| X\right],$$

the mean and variance of $\phi(X,Y)$ conditionally on X, then the variance of $\phi(X,Y)$ may be rewritten as the sum of the expected conditional variance and the variance of the conditional expectation, so that the mean square error reads:

$$\mathrm{MSE}(\mathbf{\Phi}_{MC}) = \frac{1}{N}\mathbb{E}\left[V_\Phi(X)\right] + \frac{1}{N}\mathrm{Var}(M_\Phi(X)). \tag{4.1.2}$$

We now state informally and unprecisely the second assumption, see Corollary 4.1.16 and Remark 4.1.17 below for possible more precise statements.

Assumption 4.1.2. *For the considered test function ϕ, the (random) conditional variance $V_\phi(X)$ is sufficiently 'small' compared to the variance of the conditional expectation* $\mathrm{Var}(M_\phi(X))$.

Under our two assumptions, we devise a *generic method* that estimates $\mathbb{E}\left[\phi(X,Y)\right]$ with a smaller variance than the naïve method (4.1.1).

Since we do not have the liberty of choosing the values of $(X_n)_{1\leq n\leq N}$, but we know exactly their distribution γ, our main idea is to use the samples $\mathbf{X} = (X_1, ..., X_N)$ to devise *random weights* $(w_n(\mathbf{X}))_{1\leq n\leq N}$ such that the empirical measure $\sum_n w_n(\mathbf{X})\delta_{X_i}$ is "as close as possible" to the true distribution γ. For instance, for some distance dist () between distributions – the choice of which will be discussed below — we may look for solutions of the minimization problem:

$$\text{minimize: } \mathrm{dist}\left(\gamma, \sum_{n=1}^N w_n\delta_{X_n}\right) \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases} \tag{4.1.3}$$

This minimization problem typically admits a unique solution $(w_1(\mathbf{X}), ..., w_N(\mathbf{X}))$, which can be used instead of the naïve uniform weights $(1/N)$ to estimate $\mathbb{E}\left[\phi(X,Y)\right]$ by:

$$\mathbf{\Phi}_W(\mathbf{X},\mathbf{Y}) = \sum_{n=1}^N w_n(\mathbf{X})\phi(X_n,Y_n).$$

In the remainder of this paper, we study this estimator to show, both theoretically and empirically, that it can indeed succeed in reducing the variance with respect to the naïve

Monte Carlo method.

4.1.2 A decomposition of the mean square error

The mean square error of our estimator is given by:

$$\begin{aligned}
&\mathrm{MSE}(\mathbf{\Phi}_W) := \mathbb{E}\left[(\mathbf{\Phi}_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right])^2\right] \\
&= \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})\phi(X_n,Y_n) - \sum_n w_n(\mathbf{X})M_\phi(X_n) + \sum_n w_n(\mathbf{X})M_\phi(X_n) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})(\phi(X_n,Y_n) - M_\phi(X_n))\right)^2\right] + \mathbb{E}\left[\left(\sum_n w_n(\mathbf{X})M_\phi(X_n) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right]
\end{aligned}$$

since the cross terms vanish because $\mathbb{E}\left[\phi(X_n,Y_n) - M_\phi(X_n)|\mathbf{X}\right] = 0$. In the first term, we expand the square, condition on $\mathbf{X}$ and use the conditional independence of the (Y_n); we rewrite the second term using the notation $\eta_N^\star = \sum_n w_n(\mathbf{X})\delta_{X_n}$ and get

$$\begin{aligned}
&\mathbb{E}\left[(\mathbf{\Phi}_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right])^2\right] \\
&\qquad = \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2 V_\phi(X_n)\right] + \mathbb{E}\left[\left(\int M_\phi(x)d\eta_N^\star(x) - \int M_\phi(x)\gamma(dx)\right)^2\right] \qquad (4.1.4)
\end{aligned}$$

Let us note here that for the naïve choice $w_n(\mathbf{X}) = 1/N$, Equation (4.1.4) reduces to the decomposition (4.1.2) in two terms of the same order $1/N$. By choosing weights that minimize the distance between the reweighted measure $\eta_N^\star$ and γ, our goal is to make the second term of the right hand side of (4.1.4) negligible; for this we pay a price by increasing the first term. If $V_\phi(X)$ is small enough in a suitable sense, this price is expected to be small enough to still be able to decrease the global mean square error. It is the purpose of the main theoretical results of this paper to make this informal statement precise; see Section 4.1.5, in particular Corollary 4.1.13 (conditional on the validity of Conjecture 4.1.11) as well as Corollary 4.1.16 and Remark 4.1.17.

In order to give rigorous statements, we make two additional assumptions concerning the test function ϕ and the distance we will use.

Assumption 4.1.3 (The distance)**.** *The distance dist*$(\,)$ *may be written in operator norm form*

$$dist(\eta,\gamma) = \sup_{f\in\mathcal{D}} |\eta(f) - \gamma(f)| \qquad (4.1.5)$$

where $\mathcal{D}$ *is a set of functions (typically a unit ball of test functions).*

Assumption 4.1.4 (The test function)**.** *There exist two constants* m_ϕ, v_ϕ *such that:*

- *The conditional mean* $x \mapsto M_\phi(x)$ *is in the set* $\mathcal{D}$ *defined in the previous as-*

sumption, up to an affine transformation; more precisely, there exists c such that $(M_\phi(\cdot) - c)/m_\phi \in \mathcal{D}$. If $\mathcal{D}$ is the unit ball associated with a norm, the optimal constant m_φ is the associated distance between the line $\{M_\phi - c,\, c \in \mathbb{R}\}$ and 0.

- *The conditional variance V_ϕ satisfies*

$$V_\phi(X_n) \leq v_\phi \qquad a.s. \tag{4.1.6}$$

Assuming this, we get, as an immediate consequence of (4.1.4):

$$\mathbb{E}\left[(\mathbf{\Phi}_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right])^2\right] \leq v_\phi \mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] + m_\phi^2 \mathbb{E}\left[\mathrm{dist}\left(\eta_N^\star, \gamma\right)^2\right]. \tag{4.1.7}$$

We consequently propose to define and compute the weights $(w_n(\mathbf{X}))_{1\leq n\leq N}$ according to

$$\text{minimize: } \mathrm{dist}\left(\gamma, \sum_{n=1}^N w_n \delta_{X_n}\right) + \delta \sum_n w_n^2, \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1, \end{cases} \tag{4.1.8}$$

or more simply to (4.1.3) which is obtained from the former by taking $\delta = 0$.

Remark 4.1.5 (On Assumption 4.1.4)**.** *One could weaken the almost sure bound (4.1.6) to a moment condition at the price of stronger constraints on the weights, using for instance Hölder's inequality. Such cases won't be treated here and are left for future work.*

Remark 4.1.6 (On the choice $\delta = 0$)**.** *Depending on the choice of the distance dist( ), solving (4.1.8) for $\delta \neq 0$, instead of $\delta = 0$ — which is exactly (4.1.3) — can be almost free or quite costly numerically, as will be detailed in the next section. Moreover it requires the tuning of another parameter δ. As a consequence, for simplicity and homogoneity, we will mainly focus on the choice $\delta = 0$. From the discussion above, this choice is formally appropriate in the limiting case of observables ϕ for which $v_\phi \ll m_\phi^2$.*

To specify completely the algorithm, we need to choose an appropriate distance between probability measures to use in the minimization problem (4.1.3)-(4.1.8). Among the many possible choices, see e.g. [**?**] for a review, we focus on two choices.

4.1.3 A regularized L^2 distance

The first distance uses the Hilbert structure of the space $L^2(\gamma)$. In this setting, a natural choice for comparing a probability measure η with the target Gaussian distribution γ would be the χ^2 divergence $\int (\frac{d\nu}{d\gamma} - 1)^2 d\gamma$. Since this is degenerate if η is discrete, we need to mollify η in some way before taking this divergence. A natural way of doing this in

$L^2(\gamma)$ is to use the Mehler kernel

$$K_h(x,y) := h^{-1/2}\exp(-(x-\sqrt{1-h}y)^2/(2h))\exp(x^2/2) \tag{4.1.9}$$
$$= h^{-1/2}\exp\left(-\frac{\sqrt{1-h}}{2h}(x-y)^2 + \frac{\sqrt{1-h}}{2+2\sqrt{1-h}}\left(x^2+y^2\right)\right).$$

and map η to $\int K_h(x,y)d\eta(y)$: in probabilistic terms, we replace η by ηP_t where P_t is the Ornstein–Uhlenbeck semigroup, before taking the χ^2 divergence. More formally, for any

$$h = 1 - e^{-2t} \in (0,1),$$

we will see below in Section 4.3.2 that the formula

$$\|\nu\|_h = \left\|\frac{d(\nu P_t)}{d\gamma}\right\|_{L^2(\gamma)} = \left\|\int K_h(x,y)\,\nu(dy)\right\|_{L^2(\gamma)}, \tag{4.1.10}$$

defines a norm on signed measures ν satisfying some moment conditions, and we can set

$$\mathrm{dist}(\eta_1,\eta_2) = \|\eta_1 - \eta_2\|_h$$

in (4.1.3). The latter distance has a variational representation as follows

$$\|\nu\|_h = \sup_{\|f\|_{L^2(\gamma)}\leq 1} \nu P_t f,$$

so that the set $\mathcal{D}$ in (4.1.5) is the 'regularized' image by the Ornstein–Uhlenbeck semigroup of the unit ball of the Hilbert space $L^2(\gamma)$, and the optimal constant m_ϕ in Assumption 4.1.4 is defined by (see also Remark 4.3.5):

$$m_\phi := \sup_{\|\nu\|_h\leq 1\,\nu(1)=0} \nu M_\phi = \left\|P_t^{-1}\left(M_\phi - \int M_\phi \gamma\right)\right\|_{L^2(\gamma)}. \tag{4.1.11}$$

Finally, we will detail in Section 4.3.4 how the minimization problem (4.1.8) turns out to be a quadratic programming convex optimization problem, which can be solved using standard methods, typically with a cubic polynomial complexity in terms of the sample size N. Solving the case $\delta \neq 0$ is in fact easier than the case $\delta = 0$ since larger δ simply improve the conditioning of the symmetric matrix underlying the quadratic porgramming problem.

4.1.4 An optimal transport distance

The second choice we investigate is the Wasserstein distance $\mathcal{W}_1$, defined classically as follows:

Definition 4.1.7 (Wasserstein distance)**.** *Let (E, d) be a Polish metric space. For any two probability measures η_1, η_2 on E, the Wasserstein distance between η_1 and η_2 is defined by the formula*

$$\begin{aligned}\mathcal{W}_1(\eta_1, \eta_2) &= \Big(\inf_{\pi \in \Pi} \int_E d(x_1, x_2) d\pi(x_1, x_2) \Big) \\ &= \inf \Big\{ \mathbb{E}\left[d(X_1, X_2) \right], \quad Law(X_1) = \eta_1, Law(X_2) = \eta_2 \Big\},\end{aligned} \tag{4.1.12}$$

where Π is the set of all couplings of η_1 and η_2.

Kantorovitch duality (see [Vil08]) implies that the latter distance is in fact an operator norm of the form

$$\mathcal{W}_1(\eta_1, \eta_2) = \|\eta_1 - \eta_2\|_{\mathrm{Lip}} = \sup_{\|f\|_{\mathrm{Lip}} \leq 1} \eta_1(f) - \eta_2(f),$$

where $\|f\|_{\mathrm{Lip}} = \sup_{x,y} \frac{f(x)-f(y)}{d(x,y)}$ is the Lipschitz seminorm. This is consistent with choosing for $\mathcal{D}$ in (4.1.5) the set of 1-Lipschitz functions.

Finally, we will see in Proposition 4.4.1 that, at least in dimension 1 for the choice $d(x_1, x_2) = |x_1 - x_2|$, the latter distance leads to an *explicit formula* for the optimal weights $(w_n(\mathbf{X}))_{1 \leq n \leq N}$, that can be computed with a complexity proportional to the sample size N. This leads to faster algorithms and more explicit bounds on the mean square error as compared to the L^2 case of the last section. However, for this optimal transport method, solving the case $\delta \neq 0$ is non explicit and thus harder (although still a convex optimization problem).

Remark 4.1.8 (On other Wasserstein distances)**.** *We do not really lose generality here by only considering the $\mathcal{W}_1$ distance. Indeed, as can be seen from the proof of Proposition 4.4.1 below, the optimal weights would be the same for any distance $\mathcal{W}_p$, $p \geq 1$.*

4.1.5 Theoretical results

Recall that $\mathbf{X} = (X_1, ..., X_N)$ is an i.i.d. $\mathcal{N}(0, 1)$ sequence in $\mathbb{R}$. We denote by

$$\overline{\eta}_N = \frac{1}{N} \sum_n \delta_{X_n}.$$

the empirical measure of the sample $\mathbf{X}$. The reweighted measure $\sum_n w_n(\mathbf{X}) \delta_{X_n}$ will be denoted by:

- $\eta^\star_{h,N}$, if the $w_n(\mathbf{X})$ solve (4.1.8) for the L^2 distance with parameter h and δ;
- $\eta^\star_{\mathrm{Wass},N}$, if the $w_n(\mathbf{X})$ solve (4.1.3) for the Wasserstein distance (we will only consider the case $\delta = 0$).

We first focus on results on these optimally reweighted measures, shedding light on the behaviour of the bound (4.1.7) and especially the second term in it. We start by the L^2 minimization method.

Theorem 4.1.9 (The L^2 method)**.** *For any fixed h, N and any $\delta \geq 0$, the optimization problem (4.1.8) with the distance $\|\cdot\|_h$ has almost surely a unique solution. The distance of the optimizer $\eta^\star_{h,N}$ to the target γ satisfies:*

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] \leq \mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^2\right] = \frac{1}{N}\left(\frac{1}{h} - 1\right). \tag{4.1.13}$$

Moreover, in the case $\delta = 0$, there exists a numerical h_0 such that, for all $h > h_0$,

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] = o(1/N) \tag{4.1.14}$$

as N goes to infinity.

The following result, where the window size h is allowed to depend on h, is an easy consequence of (4.1.13).

Corollary 4.1.10. *If $(h_N)_{N\geq 1}$ is bounded away from 1 and satisfies $Nh_N \to \infty$, then*

$$\eta^\star_{h_N,N} \xrightarrow{(d)} \gamma \quad \textit{in probability.}$$

The bound given by Equation (4.1.14) justifies our strategy in the sense that we managed to decrease significantly the second term in the decomposition (4.1.7) of the mean square error. Considering the first term in that decomposition leads naturally to the following conjecture, which is supported by numerical tests.

Conjecture 4.1.11. *For any h, there exists a constant C_h such that the optimal weights for the L^2 method with $\delta = 0$ satisfy*

$$\limsup_N N\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \leq C_h.$$

Remark 4.1.12 (The conjecture holds true for $h = 0$)**.** *A quick computation based on (4.1.9) and Section 4.3.4 shows that the quadratic minimisation problem (4.1.8) in the case $\delta = 0$ and $h = 0$ is equivalent to the minimization over the simplex of the diagonal quadratic form*

$$(w_1, \ldots, w_N) \mapsto \sum_{n=1}^N w_n^2 \exp\left(\frac{1}{2}X_n^2\right),$$

which is solved by $w_n(\mathbf{X}) = Y_n/(\sum_m Y_m)$, *where* $Y_m = \exp\left(-\frac{1}{2}X_m^2\right)$. *Therefore*

$$\mathbb{E}\left[N\sum_n w_n^2(\mathbf{X})\right] = \mathbb{E}\left[\frac{N\sum_n Y_n^2}{(\sum_m Y_m)^2}\right] = \mathbb{E}\left[\frac{N^2 Y_1^2}{(\sum_m Y_m)^2}\right].$$

The random variable $S_N = N^2Y_1^2/(\sum_n Y_n)^2 \leq N^2$ *converges almost surely to* $Y_1^2/\mathbb{E}[Y_1]^2$ *by the law of large numbers. Moreover, the* S_N *are uniformly integrable. Indeed,*

$$\mathbb{E}[S_N \mathbf{1}_{S_N>K}] \leq N^2\mathbb{P}[S_N > K] \leq N^2\mathbb{P}\left[\frac{1}{N}\sum_n Y_n < K^{-1/2}\right],$$

where in the last inequality we have used $Y_1 \leq 1$. *If* $K^{-1/2} < \mathbb{E}[Y_1]$, *the last probability is exponentially small in* N *by Hoeffding's inequality so that the uniform integrability follows. Consequently*

$$\lim_{N\to+\infty} \mathbb{E}\left[N\sum_n w_n(\mathbf{X})^2\right] = \frac{\mathbb{E}[Y_1^2]}{\mathbb{E}[Y_1]^2} = \frac{2}{\sqrt{3}},$$

and the conjecture holds with $C_0 = 2/\sqrt{3}$.

Corollary 4.1.13. *Assume that Conjecture 4.1.11 holds true. Let* h *be larger than the numerical constant* h_0 *of Theorem 4.1.9, and assume that* M_ϕ *is regular enough so that (4.1.11) is finite, i.e.* $m_\phi < +\infty$ *– for instance* M_ϕ *is analytic. Assume also that* $V_\phi(x)$ *is bounded above by a constant* v_ϕ. *Then the* L^2 *method with* $\delta \leq v_\phi/m_\phi^2$ *satisfies*

$$MSE(\mathbf{\Phi}_W) = \mathbb{E}\left[(\Phi_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}[\phi(X,Y)])^2\right] \leq \frac{C_h v_\phi}{N} + m_\phi^2 o(1/N)$$

for some numerical constant C_h *and numerical* $o(1/N)$.

Therefore, under the above assumptions, the L^2 *method is asymptotically better than the naïve Monte Carlo approach in terms of MSE as soon as* $v_\phi \leq \mathrm{Var}(M_\phi(X))/(C_h - 1)$.

Remark 4.1.14 (On the optimal choice of h)**.** *The question of the best choice for the smoothing parameter* h *is not easy to tackle: in the upper bound (4.1.7),* h *appears in the weights* via C_h, *in the distance and in* m_ϕ. *We will give below theoretical and empirical evidence that the best choice is related to the regularity of the test function* ϕ, *and that smaller* h *are needed if* ϕ *is very irregular.*

We are able to prove similar but more complete results for the Wasserstein method.

Theorem 4.1.15 (The Wasserstein method)**.** *For any* N, *the optimization problem with* $\delta = 0$ *has almost surely a unique solution. The distance* $D = \mathcal{W}_1\left(\eta^\star_{Wass,N}, \gamma\right)$ *of the optimizer* $\eta^\star_{Wass,N}$ *to the target* γ *satisfies for all integer* $p \geq 1$ *the moment bounds:*

$$\mathbb{E}[D^p] = \mathcal{O}^\star\left(\frac{1}{N^p}\right),$$

where $\mathcal{O}^\star$ means $\mathcal{O}$ up to logarithmic factors. In particular,

$$\eta^\star_{Wass,N} \xrightarrow[N\to+\infty]{Law} \gamma \quad in \quad in\ probability.$$

Moreover, the optimal weights satisfy:

$$\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \leq \frac{6}{N}.$$

Corollary 4.1.16. *Assume M_ϕ is Lipschitz – so that $m_\phi < +\infty$, and that $V_\phi(x)$ is bounded above by a constant v_ϕ. Then the Wasserstein method with $\delta = 0$*

$$MSE(\mathbf{\Phi}_W) = \mathbb{E}\left[\left(\Phi_W(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right] \leq \frac{c_0 v_\phi}{N} + m_\phi^2 o(1/N)$$

for some numerical constant $c_0 \leq 6$ and numerical $o(1/N)$.

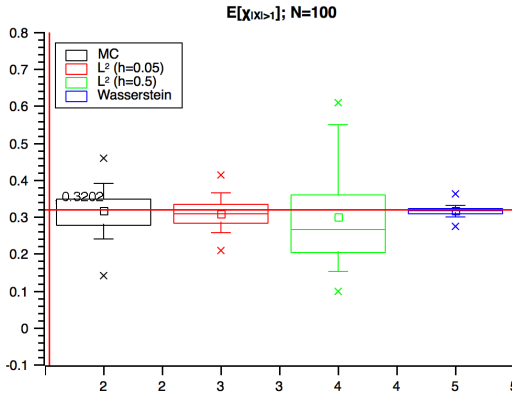
Therefore, under the above assumptions, the Wasserstein method is asymptotically better than the naïve Monte Carlo approach in terms of MSE as soon as $v_\phi \leq \mathrm{Var}(M_\phi(X))/(c_0 - 1)$.

Remark 4.1.17 (The optimal c_0)**.** *A careful look at the proof shows that, if the X_i follow the uniform distribution on $[0,1]$, the bound on $\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right]$ may be divided by 4, leading to $c_0 = 3/2$. Numerical tests suggest that even in the Gaussian case, this bound still holds true asymptotically in the sense that $N\mathbb{E}\left[\sum_n w_n(\mathbf{X})^2\right] \to \frac{3}{2}$. Therefore, we conjecture that the Wasserstein method is better than the naïve Monte Carlo approach as soon as $v_\phi \leq 2\,\mathrm{Var}(M_\phi(X))$.*

4.1.6 Numerical experiments

We supplement our theoretical findings with numerical tests. In the first series of tests, we compare numerically the naïve Monte Carlo method, the L^2 method with various choices of the bandwidth, and the Wasserstein method, in the toy case where X itself is the variable of interest. For simplicity, and for homogeneity between the two methods, we have chosen in this first serie of numerical tests $\delta = 0$ (up to numerical precision) in the the L^2-method. This case is an idealized case of concrete problems where $v_\phi \ll m_\phi^2$.

The full results may be found in Section 4.5. Figure 4.1 shows that both methods perform much better than the naïve Monte Carlo estimator. The L^2 method is often able to reduce significantly the statistical error, but the bandwidth parameter h must be chosen carefully, depending on N and the type of observable we are interested in. The parameter-free Wasserstein method is faster and more robust, but may be outperformed by a well-tuned L^2 method for very regular observables (for example the cosine function).



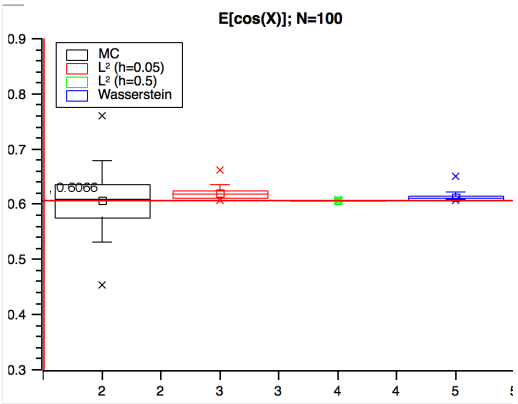


Figure 4.1: Comparison of methods for $\phi(X)$ and $\delta = 0$

The second series of numerical tests in done in Section 4.6. We let G be a d-dimensional standard Gaussian vector, and assume we are interested in the distribution of a non-linear function $Y = F(G)$. We linearize F near 0 and let $X = (DF)_0 G$, so that the distribution of X is an explicit one dimensional Gaussian. We then use our method to estimate, for any fixed t, the cumulant generating function $\log \mathbb{E}\left[\exp(tF(G))\right]$, using X as our "control variable". In this more realistic setting, we focus on the more robust Wasserstein method, and show how it can be compared to, and combined with, a more classical control variate approach.

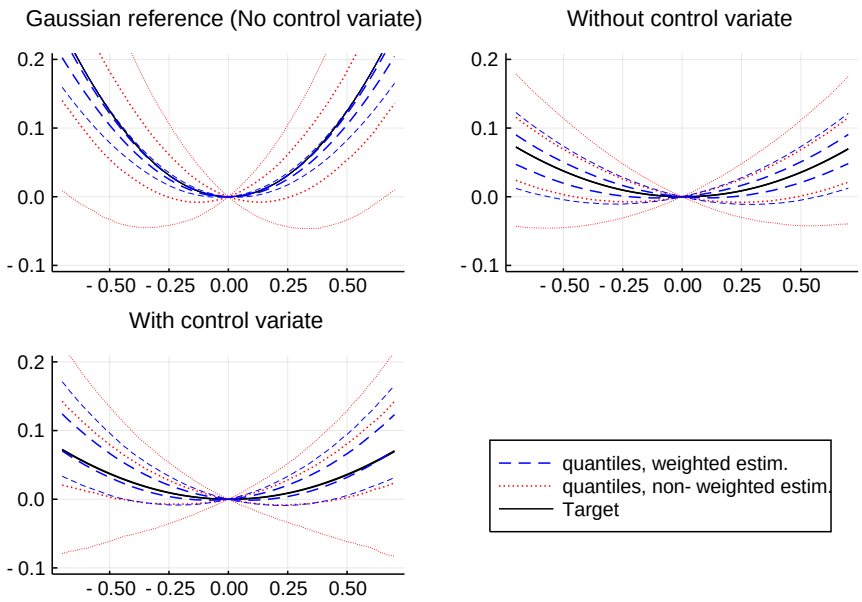


Figure 4.2: The figures above represents the $[.05, .25, .75, .95]$-quantile enveloppes of the different estimators of the cumulant generating functions for $F(G)$.

In Figure 4.2, we clearly see that our reweighting method always reduces variance in a substantial way, without using any prior information of F. This non-linear example also shows that a standard variance reduction by a linear control variate may be useless.

4.1.7 Conclusion

We have proposed *generic and robust* variance reduction techniques based on reweighting samples using a one dimensional Gaussian control variable. The latter can be seen as generalization of post-stratification methodologies and can outperform variance reduction by control variate even in simple situations.

Theoretically, the results, which prove effective variance reduction for both methods, are quite similar. The main difference is that in the Wasserstein setting, we are actually able to control the variance of the weights with an explicit constant $c_0 \leq 6$; in the L^2 case we only conjecture that a similar result holds. This difference essentially comes from the fact that in our one dimensional setting, the optimal Wasserstein weights are explicit, and therefore much easier to study.

Numerically, we have observed (as theoretically suggested) that the L^2 approach, as compared to the Wasserstein approach, requires more regular obervables and some tuning, and is more costly when the sample size become large. Note however, that the L^2 approach may be amenable to control variables X in higher dimension, where Wasserstein optimization — optimal transport — problems are known to be very cumbersome. This

issue is left for future work.

4.1.8 Outline of the paper

In Section 4.2, we briefly discuss how our method fits in the landscape of variance reduction techniques, and how it can be seen as complementing control variates and generalizing post-stratification. In Section 4.3 we discuss the L^2 method; the results on the Wasserstein approach are established in Section 4.4. The first numerical tests, considering only the gaussian variable X, are presented in Section 4.5. Finally, the tests on more realistic models are presented in Section 4.6.

4.2 Comparison with classical methods

4.2.1 Comparison to variance reduction with control variates

The method presented in this work can be seen as an alternative to control variates. More precisely, as we are about to explain now, they can be interesting as a complement to control variates when the latter is either not efficient or too expensive.

Within the framework of Section 4.1.1, a control variate is a computable function ψ of X such that the mean square error of $\mathbb{E}\left[(\phi(X,Y)-\psi(X))^2\right]$ is as small as possible (see [Owe13, Gla13] for a general introduction). Recent works have studied various techniques to find an optimized ψ using basic functions and a Monte Carlo approach (see for instance [Jou09, PS18]). The associated estimator is then

$$\mathbf{\Phi}_{CV}(\mathbf{X},\mathbf{Y}) = \frac{1}{N}\sum_{n=1}^{N}\phi(X_n,Y_n)-\psi(X_n)+\int\psi d\gamma,$$

whose mean square error (identical to variance since it is unbiased), satisfies

$$\mathrm{MSE}(\mathbf{\Phi}_{CV}(\mathbf{X},\mathbf{Y})) = \frac{1}{N}\mathbb{E}\left[\left(\phi(X,Y)-\psi(X)+\int\psi d\gamma-\mathbb{E}\left[\phi(X,Y)\right]\right)^2\right].$$

This quantity is classically minimized by choosing $\psi(X)=\mathbb{E}\left[\phi(X,Y)|X\right]$, up to an irrelevant additive constant.

As a consequence, if a good control variate, close to the conditional expectation $\psi(X)\sim\mathbb{E}\left[\phi(X,Y)|X\right]$ is available, and if we try to apply our method to $\tilde{\phi}(X,Y)=\phi(X,Y)-\psi(X)$, then Assumption 4.1.2 *will not hold* for $\tilde{\phi}$. In the numerical experiment done in Section 4.6 — see also Figure 4.2 — we compare the reweighting method with a natural affine control variate $\Psi(X)=aX+b$ which is not sufficient to approximate correctly $\mathbb{E}\left[\phi(X,Y)|X\right]$; interestingly the reweighting method is able to overcome this issue in a generic way, without

specific analytic approximation of $\mathbb{E}[\phi(X,Y)|X]$ contrary to what is required to improve the control variate.

The latter discussion suggests that the present variance reduction method based on re-weighting samples will be useful in one of the following two situations:

- The available control variates behaves very poorly.
- One is interested in estimating $\mathbb{E}[\phi(X,Y)]$ for a large class of test functions ϕ, making the calculation of control variates very costful.

4.2.2 Comparison to post-stratification variance reduction

The present work may be interpreted as a generalization of post-stratification methods to continuous state spaces. In the framework of Section 4.1.1, post-stratification can be defined by first choosing a finite partition of $\mathbb{R}$, given by K 'strata', for instance the K-quantiles $x_{1/2} < \ldots < x_{K-1/2}$ defined by $\int_{x_{k-1/2}}^{x_{k+1/2}} d\gamma = 1/K$.

In that context, the post-stratification weights are defined by

$$\begin{cases} w_n(\mathbf{X}) = \frac{1}{KB_n(\mathbf{X})}, \\ B_n(\mathbf{X}) = \text{card}\left\{X_m \in \text{strat}(X_n),\, 1 \leq m \leq N\right\}, \end{cases} \tag{4.2.1}$$

where in the above $\text{strat}(X_n)$ is the interval $[x_{k-1/2}, x_{k+1/2}[$ containing X_n. The latter post-stratification weights are defined so that the sum of the weights of particles in a given stratum is constant an equal to $1/K$.

We can then check the following:

Lemma 4.2.1. *Let us denote by $\mathcal{D}_K$ the space of functions on $\mathbb{R}$ that are constant on the srata $[x_{k-1/2}, x_{k+1/2}[$, for $k = 1 \ldots K$. Consider the semi-norm over finite measures*

$$p_K(\mu) = \sup_{\psi \in \mathcal{D}_K,\, \|\psi\|_\infty \leq 1} \mu(\psi).$$

Then the post-stratification weights (4.2.1) is the solution to the minimization problem obtained by setting $dist(\eta,\gamma) = p_K(\eta - \gamma)$ in (4.1.3) that is

$$\textit{minimize: } p_K\left(\gamma - \sum_{n=1}^{N} w_n \delta_{X_n}\right), \quad \textit{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1, \end{cases} \tag{4.2.2}$$

that moreover minimize the variance of the weight $\sum_n w_n^2 - 1$.

Proof. It's easy to check that, by definition,

$$p_K\left(\gamma - \sum_{n=1}^{N} w_n \delta_{X_n}\right) = \sum_{k=1}^{K} \left| \frac{1}{K} - \sum_n w_n \mathbf{1}_{X_n \in [x_{k-1/2}, x_{k+1/2}[} \right|$$

As a consequence, the weights that are solution to the minimization (4.2.2) are exactly those such that for all $1 \leq m \leq n$

$$\sum_n w_n \mathbf{1}_{X_n \in \mathrm{strat}(X_m)} = 1/K$$

which means that the sum of the weights of particles in the same stratum are equal to $1/K$. Now the unique minimum of $\sum_{n\in I} w_n^2$ under the constraint that $\sum_{n\in I)} w_n = c_0$ is constant is given by uniform weights $w_n = c_0/\mathrm{card}(I)$ since by Jensen

$$\sum_{n\in I} (c_0/\mathrm{card}(I))^2 = \left(\sum_{n\in I} w_n\right)^2 /\mathrm{card}(I) \leq \sum_{n\in I} w_n^2. \qquad \square$$

As a consequence, the methods presented in this work can be interpreted as extensions of the post-stratification methods from the semi-norm p_K to the norms $\|\,.\,\|_h$ or to the Wasserstein distance (which is in fact a norm) $\mathcal{W}_1$.

4.3 The L^2 method

In this section, after recalling a few classical facts and formulae on the Hilbert space $L^2(\gamma)$, we define precisely the h-norm on signed-measures, and study in Section 4.3.4 the corresponding optimization problem. Finally, we prove in Section 4.3.5 and 4.3.6 the results announced in Theorem 4.1.9.

4.3.1 Useful tools in L^2

We start by recalling a few useful definitions and results concerning the standard Gaussian Hilbert space $L^2(\gamma)$.

Orthogonalizing the standard polynomial basis with respect to the scalar product $\langle f, g\rangle_\gamma = \int f(x)g(x)d\gamma(x)$ gives rise to the classical family of *Hermite polynomials* (H_n), see e.g. [AS92, Chapter 22] for details: a Hilbert basis of $L^2(ga)$, where H_n is a polynomial of degree n, with the normalization $\langle H_m, H_n\rangle = m!\mathbf{1}_{m=n}$. We write h_n the corresponding orthonormal basis $h_n = (n!)^{-1/2}H_n$.

Recall the definition of the Mehler kernel (4.1.9):

$$\begin{aligned} K_h(x,y) &= h^{-1/2}\exp(-(x-\sqrt{1-h}y)^2/(2h))\exp(x^2/2) \\ &= h^{-1/2}\exp\left(-\frac{\sqrt{1-h}}{2h}(x-y)^2+\frac{\sqrt{1-h}}{2+2\sqrt{1-h}}\left(x^2+y^2\right)\right). \end{aligned}$$

It will be useful to introduce another parameter t such that $h = 1-e^{-2t}$; we let

$$k_t(x,y) = K_h(x,y) = K_{1-e^{-2t}}(x,y).$$

The classical formula of Mehler gives the spectral decomposition of this kernel.

Lemma 4.3.1 (Mehler's formula)**.** *For all $(x,y)\in\mathbb{R}^2$ and $t>0$,*

$$k_t(x,y) = \sum_{n=0}^{\infty} e^{-nt}h_n(x)h_n(y) = k_t(y,x). \tag{4.3.1}$$

It is also classical to interpret this kernel as the probability density kernel of the Ornstein-Uhlenbeck semigroup with respect to the standard Gaussian.

Lemma 4.3.2. *Let P_t denotes the semigroup of probability transitions of the Ornstein–Uhlenbeck process solution to the SDE $dX_t = -X_t dt + \sqrt{2}B_t$ where B_t is a standard Brownian motion, that is $\mathbb{E}[f(X_t)|X_0\sim\eta] = \int P_t f(x)d\eta(x)$ for all bounded continuous test function f and any probability measure η. Then it holds*

$$P_t(x,dz) = k_t(x,z)\gamma(dz).$$

Proof. X_t has the same distribution as $e^{-t}X_0 + \sqrt{1-e^{-2t}}G$ for a standard Gaussian random variable G. Hence recalling (4.1.9) and $h = 1-e^{-2t}$ it yields $\mathbb{E}[f(X_t)|X_0=x] = \int f(z)k_t(x,z)\gamma(dz)$. □

Let us collect a few consequences of this representation.

Lemma 4.3.3. *Let $\gamma = \mathcal{N}(0,1)$, then*

$$\int K_h(x,y)\,\gamma(dy) = \int K_h(x,y)\gamma(dx) = 1; \tag{4.3.2}$$

$$\int k_s(x,y)k_t(y,z)\gamma(dy) = k_{s+t}(x,z); \tag{4.3.3}$$

$$\int\int K_h(x,y)^2\gamma(dx)\gamma(dy) = \frac{1}{h}. \tag{4.3.4}$$

Proof. For the first equality, we integrate (4.3.1) with respect to one of the variables:

$$\begin{aligned}\int k_t(x,y)\gamma(dy) &= \int \sum_{n=0}^{\infty} e^{-nt} h_n(x)h_n(y)\gamma(dy) \\ &= \sum_{n=0}^{\infty} e^{-nt} h_n(x) \int h_n(y)\gamma(dy) = h_0(x) = 1.\end{aligned}$$

The second equality is another way of expressing the semigroup property for the Ornstein Uhlenbeck process: For all $x,\ z \in \mathbb{R}$

$$\begin{aligned}\int k_t(x,y)k_s(y,z)\gamma(dy) &= \int \sum_{m,n} e^{-mt-ns} h_m(x)h_m(y)h_n(y)h_n(z)\gamma(dy) \\ &= \sum_{n} e^{-n(t+s)} h_n(x)h_n(z) \\ &= k_{t+s}(x,z).\end{aligned}$$

Applying this to the special case $x = z$ and $t = s$, for $h = 1 - e^{-2t}$, and integrating with respect to $\gamma(dx)$ yields

$$\begin{aligned}\int\int K_h(x,y)^2\gamma(dx)\gamma(dy) &= \int k_{2t}(y,y)\gamma(dy) = \int \sum_{n=0}^{\infty} e^{-2nt} h_n(y)^2\gamma(dy) \\ &= \sum_{n=0}^{\infty} e^{-2nt} = \frac{1}{1-e^{-2t}} = \frac{1}{h}.\square\end{aligned}$$

4.3.2 The h-norm: theoretical properties

We gather the definition and main properties of the h-norm in the following result.

Theorem 4.3.4. *Let*

$$h = 1 - e^{-2t}, \qquad t > 0.$$

Let $\mathcal{M}$ be the set of signed measures on $\mathbb{R}$ with a finite total mass, and let

$$\mathcal{S} = \left\{ \nu \in \mathcal{M}, \int \exp\left(\frac{y^2}{4}\right) |\eta|\,(dy) < +\infty \right\}.$$

1. *For any $\nu \in \mathcal{S}$, the function $x \mapsto \int K_h(x,y)d\nu(y)$ is in $L^2(\gamma)$, so that*

$$\|\nu\|_h := \left\| \frac{d\,(\nu P_t)}{d\gamma} \right\|_{L^2(\gamma)} = \left\| \int K_h\,(x,y)\,\nu(dy) \right\|_{L^2(\gamma)} < \infty.$$

2. *Let us denote $\mathcal{D} := \{P_t\varphi; \varphi \in L^2(\gamma), \|\varphi\|_{L^2(\gamma)} \le 1\}$ the unit ball of the space $\{\psi = P_t\varphi; \varphi \in L^2(\gamma)\}$ endowed with the norm $\left\|P_t^{-1}\psi\right\|_{L^2(\gamma)}$. Then $\|\nu\|_h$ admits the dual*

representation

$$\|\nu\|_h = \sup_{\|\varphi\|_{L^2(\gamma)}\leq 1} \int \varphi \frac{d(\nu P_t)}{d\gamma} d\gamma = \sup_{\|P_t^{-1}\psi\|_{L^2(\gamma)}\leq 1} \int \psi d\nu = \sup_{\psi\in\mathcal{D}} \int \psi d\nu. \tag{4.3.5}$$

3. *The map $\nu \mapsto \|\nu\|_h$ is a norm on $\mathcal{S}$.*

4. *$\|\cdot\|_{h'} \leq \|\cdot\|_h$, when $h \leq h'$.*

5. *If $(\eta_k)_{k\in\mathbb{N}}$ and η are probability measures in $\mathcal{S}$, and if $\|\eta_k - \eta\|_h \to 0$, then η_k converges weakly to η.*

Remark 4.3.5 (On the set $\mathcal{D}$)**.** *The set $\mathcal{D}$ is the image of the unit ball of L^2 by the Ornstein-Uhlenbeck semigroup and consists of very regular functions. Indeed, its coefficents on the Hermite basis must decrease geometrically: $\psi \in \mathcal{D}$ if and only if*

$$\|\psi\|_{L^2(\gamma)} = \sum_{k\geq 0} \mathrm{e}^{kt} \left(\int_{\mathbb{R}} h_k \psi d\gamma \right)^2 \leq 1.$$

The set $\mathcal{D}$ contains of course all conveniently normalized polynomials, as well as many explicit non-polynomial functions. For instance, the cosine function is in $\mathcal{D}$, as can be checked thanks to the computations of $P_t e^{i\lambda\cdot}$ in the proof below.

Proof. 1. By Minkowski's integral inequality, we have

$$\begin{aligned}\|\nu\|_h &= \left(\int \left(\int K_h(x,y)\nu(dy) \right)^2 \gamma(dx) \right)^{\frac{1}{2}} \\ &\leq \int \left(\int (K_h(x,y))^2 \gamma(dx) \right)^{\frac{1}{2}} |\nu|(dy).\end{aligned}$$

By (4.3.3) from Lemma 4.3.3 applied with $x = z$ and $s = t$, and a quick computation using the explicit formula for k_{2t}, we can rewrite the innermost integral as follows:

$$\int K_h(x,y)^2 \gamma(dx) = k_{2t}(y,y) = (1-e^{-4t})^{-1/2} \exp\left(\left(\frac{e^{-2t}}{1+e^{-2t}} \right) y^2 \right).$$

Therefore, $\|\nu\|_h$ is finite whenever $\exp\left(\left(\frac{e^{-2t}}{2(1+e^{-2t})} \right) y^2 \right)$ is $|\nu|$ integrable. In particular it is finite for all h if $|\nu|$ integrates $\exp(x^2/4)$.

2. The first equation uses the Hilbert structure of $L^2(\gamma)$ and the second one follows from the fact that the Ornstein–Uhlenbeck semigroup P_t is self adjoint in $L^2(\gamma)$.

3. Homogeneity and sub-additivity follow easily from the dual expression (4.3.5). Since the positivity $\|\eta\|_h \geq 0$ is obvious, it is enough to prove that $\|\eta\|_h = 0$ implies

$\eta = 0$. We prove this fact using characteristic functions. Denote by $\mathcal{F}(f)$ the Fourier transform, for any function $f : \mathbb{R} \to \mathbb{C}$

$$\mathcal{F}(f)(\xi) := \int e^{-2\pi i x\cdot\xi} f(x)dx.$$

We recall that, for any $a > 0$,

$$\mathcal{F}(e^{-ax^2})(\xi) = \sqrt{\frac{\pi}{a}} \exp\left(-\frac{\xi^2}{\frac{a}{\pi^2}}\right). \tag{4.3.6}$$

Now, remark that the Ornstein Uhlenbeck semigroup P_t leaves the set of functions $\{x \mapsto ce^{i\lambda x}\}$ invariant: indeed, for any $\lambda \in \mathbb{R}$, let $\tilde{\lambda} = \frac{\lambda}{\sqrt{1-h}}$ and $c = e^{\frac{h\lambda^2}{2(1-h)}}$, we have

$$\begin{aligned} P_t\left(ce^{i\tilde{\lambda}\cdot}\right)(x) &= \mathbb{E}\left[ce^{i\tilde{\lambda}\left(\sqrt{1-h}x+\sqrt{h}G\right)}\right] = ce^{i\sqrt{1-h}\tilde{\lambda}x}\mathbb{E}\left[e^{i\tilde{\lambda}\sqrt{h}G}\right] \\ &= ce^{i\lambda x}\int e^{-i(-\sqrt{h}\tilde{\lambda})x}\gamma(dx), \end{aligned}$$

by Fourier transform (4.3.6), we get

$$\begin{aligned} P_t\left(ce^{i\tilde{\lambda}\cdot}\right)(x) &= ce^{i\lambda x}\int e^{-i(-\sqrt{h}\tilde{\lambda})x}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\ &= ce^{i\lambda x}e^{-\frac{h\tilde{\lambda}^2}{2}} \\ &= e^{i\lambda x}. \end{aligned}$$

Since $\left\|ce^{i\tilde{\lambda}\cdot}\right\|_{L^2(\gamma)} = ce^{-\frac{\tilde{\lambda}^2}{2}} \leq |c|$, then for any $\lambda \in \mathbb{R}$

$$\begin{aligned} \left|\nu(e^{i\lambda\cdot})\right| &= \left|\nu\left(P_t\left(ce^{i\tilde{\lambda}\cdot}\right)\right)\right| \leq |c| \sup_{\|\phi\|_{L^2}\leq 1} |(\nu)\left(P_t\left(\phi\right)\right)| \\ &= |c|\left\|\nu\right\|_h. \end{aligned} \tag{4.3.7}$$

Therefore, if $\|\nu\|_h = 0$, then $\left|\nu(e^{i\lambda\cdot})\right| = 0$ for all λ, which implies $\nu = 0$.

4. Let $h' = 1 - e^{-2t'}$, then $t \leq t'$. By definition of $\|\cdot\|_h$, we have

$$\begin{aligned} \|\nu\|_{h'} &= \sup_{\|\varphi\|_{L^2}\leq 1} \nu\left(P_{t'}\varphi\right) = \sup_{\|\varphi\|_{L^2}\leq 1} \nu\left(P_t P_{t'-t}\varphi\right) \\ &\leq \sup_{\varphi:\left\|P_{t'-t}\varphi\right\|\leq 1} \nu\left(P_t P_{t'-t}\varphi\right) \\ &= \sup_{\|\varphi\|_{L^2}\leq 1} \nu\left(P_t\varphi\right) \\ &= \|\nu\|_h, \end{aligned}$$

where we have use that $\|P_{t'-t}\varphi\|_{L^2} \leq \|\varphi\|_{L^2}$ by Jensen's inequality.

5. Let (η_k) and η be probability measures in $\mathcal{S}$ such that $\|\eta_k - \eta\| \to 0$. For any λ and any k, we apply (4.3.7) to $\nu = \eta_k - \eta$ and let k go to infinity. This implies that $\eta_k(e^{i\lambda\cdot})$ converges to $\eta(e^{i\lambda\cdot})$ for all λ, so η_k converges to η in distribution. □

4.3.3 Choice of the bandwidth h

We have seen in Theorem 4.3.4 that the mapping $h \mapsto \|\,\|_h$ is decreasing, whereas the mapping $h \mapsto m_\phi := \sup_{\|\nu\|_h \leq 1} \nu(M_\phi)$ is increasing. If one tries to minimize the second term $m_\phi^2 \mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right]$ in the upper bound (4.1.7), one can easily check that its derivative with respect to h has the same sign as

$$\frac{d}{dh} \ln m_\phi^2 + \frac{d}{dh} \ln \mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] \tag{4.3.8}$$

On the other hand, in the Hermite polynomials orthonormal basis, we have the simple formula (see Remark 4.3.5) $m_\phi^2 = \sum_{k\geq 1} \mathrm{e}^{kt} \left(\int h_k M_\phi d\gamma\right)^2$ with $h = 1 - \mathrm{e}^{-2t}$, so that

$$\frac{d}{dh} m_\phi^2 = \frac{dt}{dh} \sum_{k\geq 1} k\mathrm{e}^{kt} \left(\int h_k M_\phi d\gamma\right)^2,$$

and thus the ratio $\frac{d}{dh} \ln m_\phi^2 = m_\phi^{-2} \frac{dm_\phi^2}{dh}$ can be interpreted as a strong measure of the irregularity of $x \mapsto M_\phi(x)$ — the more the observable M_ϕ is 'irregular', the more the high frequency modes are relatively large and the larger $\frac{d}{dh} \ln m_\phi^2$ is. As a consequence, for any fixed h, a less regular observable M_ϕ renders the gradient (4.3.8) strictly positive, showing that the minimizer of $h \mapsto m_\phi^2 \mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right]$ is attained for *smaller* h — that is, as expected, for smaller kernel bandwidths. This monotony between the best choice of bandwidth h and the regularity of the observable will be observed numerically in Section 4.5.

4.3.4 The L^2 method as a quadratic programming problem

We now discuss the minimization problem (4.1.3) when dist () is the h-norm, first from a deterministic point of view. Let $\mathbf{x} = (x_1, ..., x_N)$ be a vector in $\mathbb{R}^N$. We want to solve the following minimization problem :

$$\text{minimize: } \left\|\sum_n w_n \delta_{x_n} - \gamma\right\|_h^2 + \delta \sum_n w_n^2, \quad \text{subject to: } \begin{cases} w_n \geq 0, \\ \sum_n w_n = 1. \end{cases}$$

Let us denote by Ω the simplex $\{w = (w_1, ..., w_N) | w_i \geq 0,\ \sum_{n=1}^N w_n = 1\}$, and let $F(w) = \|\sum_n w_n \delta_{x_n} - \gamma\|_h^2$. By definition, F may be rewritten as follows:

$$\begin{aligned} F(w) &= \left\| \sum_{n=1}^N w_n K_h(y, x_n) - 1 \right\|_{L^2(\gamma(dy))}^2 \\ &= \int \left(\sum_{n=1}^N w_n K_h(y, x_n) - 1 \right)^2 \gamma(dx) \\ &= \sum_{n,m}^N w_n w_m \int K_h(y, x_n) K_h(y, x_m) \gamma(dx) - 2 \sum_{n=1}^N w_n \int K_h(y, x_n) \gamma(dx) + 1 \end{aligned}$$

By (4.3.3) in Lemma 4.3.3, we have

$$\begin{aligned} F(w) &= \sum_{n,m}^N w_n w_m k_{2t}(x_n, x_m) - 2 \sum_{n=1}^N w_n + 1 \\ &= w^\top Q w - 1, \end{aligned}$$

where Q is the $N \times N$ matrix whose components are given by

$$Q_{n,m} = k_{2t}(x_n, x_m), \quad \text{for any } 1 \leq n, m \leq N. \tag{4.3.9}$$

For future reference, let us note that using Mehler's formula, $k_{2t}(x_m, x_n) = \sum_k e^{-2kt} h_k(x_m) h_k(x_n)$, so that we can also write, for any weight vector $(w_1, ..., w_N)$,

$$\begin{aligned} \left\| \sum_{n=1}^N w_n \delta_{x_n} - \gamma \right\|_h^2 = w^\top Q w - 1 &= \sum_{k \geq 0} e^{-2kt} \left(\sum_n w_n h_k(x_n) \right)^2 - 1 \\ &= \sum_{k \geq 1} e^{-2kt} \left(\sum_n w_n h_k(x_n) \right)^2. \end{aligned} \tag{4.3.10}$$

The minimization problem is therefore reduced to the following quadratic problem over a convex set:

$$\text{minimize: } w^\top (Q + \delta \mathrm{Id}) w \quad \text{subject to: } w \in \Omega. \tag{4.3.11}$$

Proposition 4.3.6 (The quadratic programming problem)**.** *If the x_n are pairwise distinct, then Q is positive definite, and the minimization problem (4.3.11) has a unique solution even if $\delta = 0$.*

This holds in particular with probability one if the $(x_n)_{1 \leq n \leq N} = (X_n)_{1 \leq n \leq N}$ are iid samples of γ.

Remark 4.3.7. *Here the solution may or may not be in the interior of the simplex: there are vectors $\mathbf{x} = (x_1, ..., x_N)$ for which some of the components of the optimal weight vector*

are zero.

Proof. For any column vector $a = (a_1, ..., a_N)$, we need to prove that $a^T Q a = 0$ if and only if $a = 0$. The same calculation leading to (4.3.10) yields

$$a^T Q a = \sum_k e^{-2kt} \left(\sum_m a_m h_k(x_m) \right)^2 = 0,$$

which implies that, for all integer k, $\sum_{m=1}^N a_m h_k(x_m) = 0$. Let P_n be the Lagrange cardinal polynomial that satisfies $P_n(x_m) = \delta_{nm}$. Since P_n may be decomposed on the basis of the h_k, it holds that $\sum_m a_m P_n(x_m) = 0$, so a_n must be zero. Since n is arbitrary, $a = 0$. This shows that Q is positive definite.

We are therefore optimizing a strictly convex function over a compact convex set: the minimizer exists and is unique. □

4.3.5 A first comparison with the naïve empirical measure.

Recall that $\overline{\eta}_N = \frac{1}{N} \sum_n \delta_{X_n}$ and $\eta^\star_{h,N} = \sum_n w_n(\mathbf{X}) \delta_{X_n}$ denote respectively the naïve and L^2-reweighted empirical measure.

Proof of Theorem 4.1.9. The existence and uniqueness of the minimizer follow from Proposition 4.3.6.

We now want to establish (4.1.13), that is,

$$\mathbb{E}\left[\left\| \eta^\star_{h,N} - \gamma \right\|_h^2 \right] \leq \mathbb{E}\left[\left\| \overline{\eta}_N - \gamma \right\|_h^2 \right] = \frac{1}{N} \left(\frac{1}{h} - 1 \right).$$

The first inequality follows from Jensen's inequality. Indeed, by definition, the (w_n) solve (4.3.11), so that almost surely,

$$\left\| \eta^\star_{h,N} - \gamma \right\|_h^2 + \delta \sum_n w_n(\mathbf{X})^2 \leq \left\| \overline{\eta}_N - \gamma \right\|_h^2 + \delta / N.$$

Jensen's inequality on the weights (w_n) then implies

$$1/N^2 = \left(\sum_n w_n(\mathbf{X})/N \right)^2 \leq \sum_n w_n^2(\mathbf{X})/N$$

so that

$$\left\| \eta^\star_{h,N} - \gamma \right\|_{h_N} \leq \left\| \overline{\eta}_N - \gamma \right\|_h.$$

To compute the expected value of $\left\| \overline{\eta}_N - \gamma \right\|_h^2$, we use the spectral decomposition (4.3.10) to write

$$\mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^2\right] = \frac{1}{N^2}\sum_{n,m}^{N}\mathbb{E}\left[\sum_{k=1}^{\infty} e^{-2kt} h_k(X_n)h_k(X_m)\right],$$

where t satisfies $h = 1 - e^{-2t}$. Since $X_1, ..., X_N$ are i.i.d. $\mathcal{N}(0,1)$ and the $(h_k)_{k\geq 1}$ are all orthogonal to $h_0 = 1$ in $L^2(\gamma)$,

$$\begin{aligned}\mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^2\right] &= \frac{1}{N^2}\sum_{k=1}^{\infty} e^{-2kt}\Big\{\sum_{n=m}\mathbb{E}\left[(h_k(X_n))^2\right] + \sum_{n\neq m}\mathbb{E}\left[h_k(X_n)\right]\mathbb{E}\left[h_k(X_m)\right]\Big\} \\ &= \frac{1}{N}\left(\sum_{k=1}^{\infty} e^{-2kt}\right) = \frac{1}{N}\left(\frac{e^{-2t}}{1-e^{-2t}}\right) \\ &= \frac{1}{N}\left(\frac{1}{h} - 1\right),\end{aligned}$$

concluding the proof of Equation (4.1.13). □

We end this section by proving the weak convergence result of Corollary 4.1.10. The proof uses the following classical result (see e.g. [Kal02, Lem. 3.2]), which implies in particular that convergence in probability is a topological notion that does not depend on the choice of a metric.

Lemma 4.3.8 (Subsequence criterion)**.** *Let $Y_1, Y_2, ...$ be random elements in a metric space (S, d). Then $Y_n \xrightarrow{\mathbb{P}} Y$ iff for all sub-sequence $(k_n) \subset \mathbb{N}$, there exists a further subsequence $(l_{k_n}) \subset (k_n)$ such that $Y_n \to Y$ a.s. along (l_{k_n}).*

Proof of Corollary 4.1.10. We let $N \to +\infty$ with $\frac{1}{N} \ll h_N \leq h_0 < 1$, and show that in probability, the random measure $\eta^{\star}_{h,N}$ converges weakly to γ.

Let $h_N = 1 - e^{-2t}$. By Lem. 4.3.8, it is enough to prove that from any subsequence of $\eta^{*}_{h_N,N}$, we can extract a further subsequence along which $\eta^{*}_{h_N,N}$ converges in distribution to γ. Let us consider an arbitrary subsequence of $\eta^{*}_{h_N,N}$. By Equation (4.1.13),

$$\mathbb{E}\left[\left\|\eta^{\star}_{h_N,N} - \gamma\right\|_{h_N}^2\right] \leq \frac{1}{N}\left(\frac{1}{h_N} - 1\right) \xrightarrow[N\to\infty]{} 0,$$

so the random variable $\left\|\eta^{\star}_{h_N,N} - \gamma\right\|_{h_N}$ converges in $L^2(\mathbb{P})$ to 0. Convergence in L^2 implies convergence in probability so that by Lemma 4.3.8, there is a further subsequence along which

$$\left\|\eta^{\star}_{h_N,N} - \gamma\right\|_{h_N} \xrightarrow[N\to\infty]{a.s.} 0.$$

Since $h_N \leq h_0$, we get by Theorem 4.3.4, item (4), that along the sub-subsquence,

$$\left\|\eta^{\star}_{h_N,N} - \gamma\right\|_{h_0} \xrightarrow{a.s.} 0.$$

Since $\|\cdot\|_{h_0}$-convergencence implies weak convergence by item (5) of Theorem 4.3.4, $\eta^\star_{h,N} \xrightarrow[(d)]{a.s.} \gamma$ along the sub-subsequence. □

4.3.6 Fast convergence of the weighted measure and a conjecture

In this Section we prove the second part of Theorem 4.1.9: in the $\delta = 0$ case, for h sufficiently large,

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] = o(1/N).$$

4.3.6.1 Strategy of proof

Recall that $\eta^\star_{h,N}$ is defined by minimizing $\|\sum w_n \delta_{X_n} - \gamma\|_h^2$ over all weight vectors. The main difficulty here is that this minimizer is not explicit. However, for any integer K, the spectral decomposition giving (4.3.10) can be used to split the cost function in two terms:

$$\|\eta - \gamma\|_h^2 = \sum_{k=1}^{K} e^{-2kt}\left(\sum_n w_n h_k(X_n)\right)^2 + \sum_{k>K} e^{-2kt}\left(\sum_n w_n h_k(X_n)\right)^2. \tag{4.3.12}$$

Let $w_n^K(\mathbf{X})$ be an optimizer of the *first*, finite dimensional term. If N is large enough with respect to K, then it is reasonable to expect that, with high probability, the value of this finite dimensional problem is zero.

Definition 4.3.9 (K-good vectors)**.** *A vector $\mathbf{x} = (x_1, ..., x_N)$ is said to be* K-good *if there exists a weight vector $(w_1, ..., w_N)$ in the simplex Ω such that*

$$\forall 1 \le k \le K, \quad \sum_n w_n h_k(x_n) = 0.$$

Let G be the "good event" $G = \{\mathbf{X} \text{ is } K\text{-good}\}$. On G we compare $w_n(\mathbf{X})$ to $w_n^K(\mathbf{X})$; on the bad event we simply use the naïve empirical measure:

$$\begin{aligned}\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] &\le \mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2 \mathbf{1}_G\right] + \mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2 \mathbf{1}_{G^c}\right] \\ &\le \mathbb{E}\left[\left\|\sum_n w_n^K(\mathbf{X})\delta_{X_n} - \gamma\right\|_h^2 \mathbf{1}_G\right] + \mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^2 \mathbf{1}_{G^c}\right]\end{aligned}$$

For the first term, on the good event G, we apply (4.3.12): by definition the first term

vanishes and we get

$$\left\|\sum_n w_n^K(\mathbf{X})\delta_{X_n} - \gamma\right\|_h^2 \mathbf{1}_G \leq \sum_{k>K} e^{-2tk}\left(\sum_n w_n^K(\mathbf{X})h_k(X_n)\right)^2 \leq \sum_{k>K}\sum_n e^{-2tk} w_n^K(\mathbf{X})h_k^2(X_n).$$

where we used Jensen's inequality with the weights w_n^K in the last line. We now take the expectation, bounding w_n^K by one, to get

$$\mathbb{E}\left[\left\|\sum_n w_n^g(\mathbf{X})\delta_{X_n} - \gamma\right\|_h^2 \mathbf{1}_G\right] \leq \sum_{k>K}\sum_n e^{-2tk}\mathbb{E}\left[h_k^2(X_n)\right] \leq N\sum_{k>K} e^{-2tk} \leq N\frac{e^{-2t(K+1)}}{1-e^{-2t}}.$$

On the bad event we use Hölder's inequality:

$$\mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_{h_N}^2 \mathbf{1}_{G^c}\right] \leq \mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_{h_N}^4\right]^{1/2}\mathbb{P}\left[G^c\right]^{1/2}.$$

$$\mathbb{E}\left[\left\|\eta_{h,N}^\star - \gamma\right\|_h^2\right] \leq N\frac{e^{-2t(K+1)}}{1-e^{-2t}} + \mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_{h_N}^4\right]^{1/2}\mathbb{P}\left[G^c\right]^{1/2}. \tag{4.3.13}$$

In order to bound the 4th moment of the h-norm, we proceed as follows. Recall that $h = 1 - e^{-2t}$, and suppose that $e^{2t} \geq 3$, so that we can write $t = s + u$ with s satisfying $e^{2s} = 3$. Then

$$\|\overline{\eta}_N - \gamma\|_h^4 = \left\|\frac{d\overline{\eta}_N P_t}{d\gamma} - 1\right\|_2^2 = \left\|P_s\left(\frac{d\overline{\eta}_N P_u}{d\gamma} - 1\right)\right\|_2^4 \leq \left\|P_s\left(\frac{d\overline{\eta}_N P_u}{d\gamma} - 1\right)\right\|_4^2 \leq \left\|\left(\frac{d\overline{\eta}_N P_u}{d\gamma} - 1\right)\right\|_2^2$$

where the last line uses Nelson's theorem, that is, the hypercontractivity of the Ornstein-Uhlenbeck semigroup (see *e.g.* [Gro93]) which here holds true between L^4 and L^2 for time greater than $s = (\ln 3)/2$.

Taking expectations and reusing (4.1.13), we get

$$\mathbb{E}\left[\|\overline{\eta}_N - \gamma\|_h^4\right] \leq \frac{1}{N}\left(\frac{1}{1-e^{-2u}} - 1\right) = \frac{1}{N}\left(\frac{1}{1-3e^{-2t}} - 1\right).$$

Putting everything together, we have for $t \geq (\ln 3)/2$:

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] \leq N\frac{e^{-2t(K+1)}}{1-e^{-2t}} + \frac{1}{\sqrt{N}}\left(\frac{1}{1-3e^{-2t}} - 1\right)^{1/2}\mathbb{P}\left[G^c\right]^{1/2}. \tag{4.3.14}$$

To go forward, the main challenge is to get a bound on the probability of the bad event.

4.3.6.2 Control on the bad event by coupon collecting

Let M be an integer and decompose the real line $\mathbb{R}$ in M segments between the quantiles $(z_i)_{0\leq i\leq M}$, where $F_\gamma(z_i) = \int_{-\infty}^{z_i}\gamma(dx) = i/M$.

Definition 4.3.10 (Well spread vector)**.** *A vector $\mathbf{x} = (x_1, ..., x_N)$ is said to be* M-well-spread *if it visits each of the M quantiles of the Gaussian:*

$$\forall 1 \leq j \leq M, \exists 1 \leq i \leq N \quad x_i \in (z_{i-1}, z_i).$$

The main results of this section are the two following lemmas.

Lemma 4.3.11 (M-well-spread implies K-good)**.** *There exists a universal constant C such that, if $\mathbf{x}$ is M-well-spread, then it is K-good for all K such that*

$$M > CK^{5/2}8^K.$$

Lemma 4.3.12 (Large samples are well-spread)**.** *Suppose that $N > (2p+2)M\ln(M)$. For $\mathbf{X} = (X_1, ..., X_N)$, an iid gaussian sample, the probability that $\mathbf{X}$ is not M-well-spread is small:*

$$\mathbb{P}\left[\exists i, \forall n, X_n \notin (z_{i-1}, z_i)\right] \leq \frac{M}{M-1}\frac{1}{M^{2p+1}}.$$

We start by the short proof of this second lemma.

Proof of Lemma 4.3.12. We interpret the question as a coupon collecting problem. For M coupons, the number of trials T needed to get a complete collection admits the following classical deviation bound, see for example [MR95, Section 3.6.1, p. 58]:

$$\forall l \in \mathbb{N}, \quad \mathbb{P}\left[T > l\right] \leq M(1-1/M)^l \leq M\exp(-l/M),$$

obtained by expressing $\{T > l\}$ as the union of the M events "the k^{th} coupon never appears in the l trials". Thus

$$\forall t, \quad \mathbb{P}\left[T > M\ln(M) + Mt\right] \leq \frac{M}{M-1}\exp(-t),$$

where the $M/(M-1)$ factor comes from the fact that $M\ln(M) + Mt$ might not be an integer. We choose $t = (2p+1)\ln(M)$, and recall that by assumption $(2p+2)M\ln(M) < N$.

This yields a bound on the probability of not being well-spread:

$$\mathbb{P}\left[\exists i, \forall n, X_n \notin (x_{i-1}, x_i)\right] = \mathbb{P}\left[T > N\right] \leq \mathbb{P}\left[T > (2p+2)M\ln(M)\right] \leq \frac{M}{M-1}\frac{1}{M^{2p+1}}.\square$$

The proof of Lemma 4.3.11 is a bit more involved. Let us first state and prove three additional lemmas.

Lemma 4.3.13. $\mathbf{x}$ *is* K*-bad if and only if there exists a polynomial* P *such that* $\deg(P) \leq K$*,* P *is orthogonal to* h_0*, and*

$$\forall 1 \leq n \leq N, \quad P(x_i) > 0.$$

Proof. By definition, $\mathbf{x}$ is K-bad if and only if the origin of $\mathbb{R}^K$ is not in the convex hull of the N points $(h_1(X_n), ..., h_K(X_n))$. If this is the case, then by the hyperplane separation theorem there exists an $\alpha = (\alpha_1, ..., \alpha_K)$ that has a positive scalar products with the N points, that is,

$$\forall 1 \leq n \leq N, \quad \sum_{k=1}^{K} \alpha_k h_k(X_n) > 0.$$

In other words, the polynomial $P = \sum_{k=1}^{K} h_k$ takes positive values on each of the X_n for $1 \leq n \leq N$. Since the (h_k) are orthogonal, P is indeed orthogonal to h_0.

Conversely if such a $P = \sum_{k=1}^{K} \alpha_k$ exists then $\alpha = (\alpha_1, ..., \alpha_K)$ has a (strictly) positive scalar product with the N points $(h_1(X_n), ..., h_K(X_n))_{1 \leq n \leq N}$, so it has a positive scalar product with any convex combination of these points, therefore 0 cannot be in the convex hull of these points. $\square$

Lemma 4.3.14. *There is a universal constant* C *such that, if* $P = \sum_{k=1}^{K} a_k h_k$ *is a polynomial of degree at most* K *orthogonal to* h_0*, then*

$$\mathbb{E}\left[|P(Z)|^3\right]^{1/3} \leq CK^{1/4}2^{K/2}\mathbb{E}\left[P(Z)^2\right]^{1/2}.$$

Proof. Without loss of generality we may assume $\sum_{k=1}^{K} a_k^2 = 1$.

$$\begin{aligned}\mathbb{E}\left[|P(Z)|^3\right]^{1/3} &\leq \sum |a_k|\,\mathbb{E}\left[|h_k(Z)|^3\right]^{1/3} \\ &\leq CK^{-1/4}2^{K/2}\sum |a_k| \\ &\leq CK^{1/4}2^{K/2},\end{aligned}$$

where the second line follows from Theorem 2.1, eq. (2.2) in [LC02], remarking that our h_k are normalized in L^2 instead of monic, and the last line from the bound $\sum |a_k| \leq \sqrt{K}(\sum_k |a_k|^2)^{1/2}$. $\square$

Lemma 4.3.15. *If $X \in L^3$ satisfies $\mathbb{E}[X] = 0$, then*

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X^2]^3}{4\mathbb{E}[X^3]^2}.$$

Proof. Since $\mathbb{E}[X] = 0$, $\mathbb{E}[X_+] = \mathbb{E}[X_-] = \frac{1}{2}\mathbb{E}[|X|]$. Therefore by Hölder's inequality,

$$\frac{1}{4}\mathbb{E}[|X|]^2 = \mathbb{E}[X+]^2 = \mathbb{E}[X\mathbf{1}_{X>0}]^2 \leq \mathbb{E}\left[X^2\right]\mathbb{P}[X > 0].$$

Moreover, another application of Hölder's inequality yields

$$\mathbb{E}\left[X^2\right] \leq \mathbb{E}[|X|]^{1/2}\,\mathbb{E}\left[|X|^3\right]^{1/2}.$$

Putting these two inequalities together, we get

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[|X|]^2}{4\mathbb{E}[X^2]} \geq \frac{\mathbb{E}[X^2]^3}{4\mathbb{E}\left[|X|^3\right]^2}.$$ □

Proof of Lemma 4.3.11. Suppose that $\mathbf{x}$ is M-well-spread but K-bad. By Lemma 4.3.13, there exists a $P_{\mathbf{x}}$ of degree at most K that takes positive values on each of the x_i. This $P_{\mathbf{x}}$ has $L \leq K$ real roots $r_1 \leq \cdots \leq r_L$. To fix ideas, suppose that $P_{\mathbf{x}}$ is negative at $-\infty$. Setting $r_0 = -\infty$ and $r_{L+1} = \infty$, the open set $\{z : P_{\mathbf{x}}(z) < 0\}$ may therefore be written as the union of disjoint, possibly empty, intervals $\bigcup_{m \text{ even}, m \leq L}(r_m, r_{m+1})$. These intervals cannot contain the x_i, so each one is included in the union of two adjacent interquantiles intervals, so that for m even, $m \leq L$,

$$\int_{r_m}^{r_{m+1}} \gamma(dz) \leq \frac{2}{M}.$$

Furhtermore, the number of intervals is at most $\lceil (K+1)/2 \rceil \leq (K+3)/2$. Rewriting the gaussian integral as a probability, we get, for Z a standard Gaussian random variable,

$$\mathbb{P}[P_{\mathbf{x}}(Z) < 0] \leq (K+3)/M.$$

The key point now is that $P_{\mathbf{x}}$ is orthogonal to $h_0 = 1$, that is, in probabilistic terms, $\mathbb{E}[P_{\mathbf{x}}(Z)] = 0$, so that we can use the concentration lemma 4.3.15, to bound the left hand side from below and get:

$$\frac{K+3}{M} \geq \frac{\mathbb{E}[P(Z)^2]^3}{4\mathbb{E}[|P(Z)^3|]^2} \geq \frac{c}{K^{3/2}8^K}.$$

This bound implies the claim. □

4.3.6.3 End of the proof of Theorem 4.1.9

Let us recall the bound (4.3.14) for $t \geq (\ln 3)/2$:

$$\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] \leq N\frac{e^{-2t(K+1)}}{1-e^{-2t}} + \frac{1}{\sqrt{N}}\left(\frac{1}{1-3e^{-2t}} - 1\right)^{1/2}\mathbb{P}\left[G^c\right]^{1/2}. \tag{4.3.15}$$

For each N choose M and K the *largest* possible integers such that

$$N > 4M\ln(M), \qquad\qquad M > CK^{5/2}8^K. \tag{4.3.16}$$

Note that in particular $N = \mathcal{O}^\star(M)$ and $N = O((8+\varepsilon)^K)$ for any $\varepsilon > 0$. This relation between M and K also ensures that, by Lemma 4.3.11, the sample $\mathbf{X}$ is K-good as soon as it is M-well-spread, so that by Lemma 4.3.12,

$$\mathbb{P}\left[G^c\right] = O(1/M^3) = O^\star(1/N^3),$$

and thus the second term in the right hand side of (4.3.15) is $o(1/N)$. For the first term in the right hand side of (4.3.15), the main quantity to be controlled is Ne^{-2tK}. The relation between M, N and K ensures that $N^2e^{-2tK} = o(1)$ if t is large enough ($t > \ln 8$), or in other words $Ne^{-2tK} = o(1/N)$. Putting everything together, we get $\mathbb{E}\left[\left\|\eta^\star_{h,N} - \gamma\right\|_h^2\right] = o(1/N)$, concluding the proof of Equation (4.1.14) and of Theorem 4.1.9.

4.3.6.4 Proof of Corollary 4.1.13

Let us denote by $w_n^\delta(\mathbf{X})$ and by $\eta_N^{\delta,\star}$ the optimal weights and the associated weighted empirical distribution obtained by the optimization problem (4.1.8) for a given δ (we drop the subscript h in notation for simplicity). By construction,

$$\begin{aligned}\mathbb{E}\left[\mathrm{dist}\left(\eta_N^{0,\star},\gamma\right)^2\right] + \delta\sum_n w_n^0(\mathbf{X})^2 &\geq \mathbb{E}\left[\mathrm{dist}\left(\eta_N^{\delta,\star},\gamma\right)^2\right] + \delta\sum_n w_n^\delta(\mathbf{X})^2 \\ &\geq \mathbb{E}\left[\mathrm{dist}\left(\eta_N^{0,\star},\gamma\right)^2\right] + \delta\sum_n w_n^\delta(\mathbf{X})^2\end{aligned}$$

so that $\sum_n w_n^\delta(\mathbf{X})^2 \leq \sum_n w_n^0(\mathbf{X})^2$. As a consequence, the MSE obtained with a given δ can be first bounded using (4.1.7) and then using the wieghts $w_n^0(\mathbf{X})$ so that

$$\mathbb{E}\left[\left(\mathbf{\Phi}_W^\delta(\mathbf{X},\mathbf{Y}) - \mathbb{E}\left[\phi(X,Y)\right]\right)^2\right] \leq (v_\phi - \delta m_\phi^2)\mathbb{E}\left[\sum_n w_n^0(\mathbf{X})^2\right] + m_\phi^2\mathbb{E}\left[\mathrm{dist}\left(\eta_N^{0,\star},\gamma\right)^2 + \delta\sum_n w_n^0(\mathbf{X})^2\right].$$

The corallary then follows from Theorem 4.1.9 and Conjecture 4.1.11.

4.4 The Wasserstein method

4.4.1 An exact expression for the optimal weights

The fact that the Wasserstein method is both easier to analyze and faster in practice stems from the fact that the minimization problem can be solved explicitely.

Proposition 4.4.1. *Let $\mathbf{x} = (x_1, ..., x_N)$ be a set of N distinct points in $\mathbb{R}$, let $(x_{(1)} < x_{(2)} \cdots < x_{(N)}$ be their ordered relabelling, and let $(y_n)_{0 \leq n \leq N}$ be the middle points $(1/2)(x_{(n)} + x_{(n+1)})$, with the convention $y_0 = -\infty$ and $y_N = \infty$.*

For $w = (w_1, ..., w_N)$ in the simplex $\Omega = \{(w_1, ..., w_N) \in \mathbb{R}_+^N, \sum_n w_n = 1\}$, let $F(w)$ be the cost

$$F(w) = \mathcal{W}_1 \left(\sum_{n=1}^{N} w_n \delta_{x_n}, \gamma \right).$$

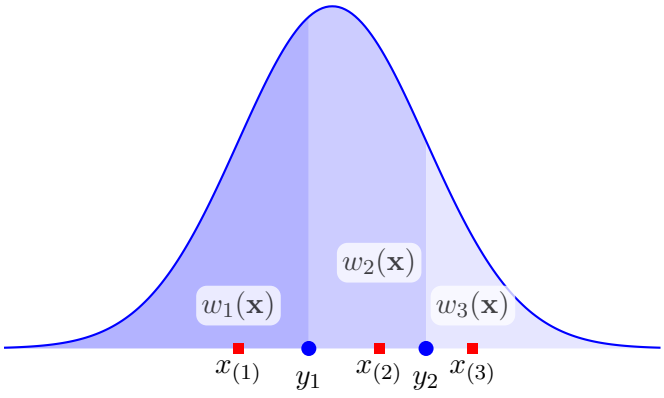
The optimization problem

$$\textit{minimize: } F(w) \quad \textit{subject to: } w \in \Omega$$

has a unique solution $w(\mathbf{x}) = (w_1(\mathbf{x}), ..., w_N(\mathbf{x}))$, given by

$$w_n(\mathbf{x}) = \int_{y_m - 1}^{y_m} \gamma(dz),$$

where m is the unique integer such that $x_n = x_{(m)}$.



Given the sample $(\mathbf{x})$, the optimal Wasserstein weights are obtained by computing the middle points $y_n = (x_{(n)} + x_{(n+1)})/2$, and letting $w_n = \gamma([y_{n-1}, y_n])$.

Figure 4.3: The optimal weights $w(\mathbf{x})$.

Proof. First, note that thanks to the relabelling in the last part of the statement, is enough to prove the result when the (x_n) are already ordered; we assume from now on that $x_1 < \cdots < x_N$.

Let η be any probability measure on $\mathbb{R}$, and recall that γ is the standard Gaussian measure; denote by F_η, F_γ their respective cumulative distribution functions. The Wasserstein distance $\mathcal{W}_1$ between η and γ admits the following classical representation, see for example [Vil03, Remark 2.19 item (iii)] :

$$\mathcal{W}_1(\gamma, \eta) = \int_{\mathbb{R}} |F_\eta(x) - F_\gamma(x)| \, dx.$$

Consider now the discrete measure $\eta(w) = \sum_{n=1}^N w_n \delta_{x_n}$. By cutting the integral at the points x_n and isolating the first and last terms, we get the explicit formula

$$\begin{aligned} F(w) &= \mathcal{W}_1\left(\sum_{n=1}^N w_n \delta_{x_n}, \gamma\right) \\ &= \sum_{n=1}^{N-1} \int_{x_n}^{x_{n+1}} \left| \sum_{m=1}^n w_m - F_\gamma(z) \right| dz + \int_{-\infty}^{x_1} |F_\gamma(z)| \, dz + \int_{x_N}^{\infty} |1 - F_\gamma(z)| \, dz. \end{aligned} \tag{4.4.1}$$

Note that the extremal terms do not depend on the weight vector w. For $1 \leq n \leq N-1$, consider now the n^{th} term in this sum, and write it as $\phi_n(\sum_{m=1}^n w_m)$, where

$$\phi_n(c) = \int_{x_n}^{x_{n+1}} |c - F_\gamma(z)| \, dz.$$

Writing $\phi_n(c) = (x_{n+1} - x_n)\mathbb{E}\left[|c - F_\gamma(U)|\right]$ for U a uniform variable on $[x_n, x_{n+1}]$, we see by classical properties of medians, see e.g. [Str11, p. 43], that ϕ_n attains its minimal value at the unique median of the distribution of $F_\gamma(U)$, that is, at the point p where $\mathbb{P}\left[F_\gamma(U) \leq p\right] = 1/2$. Since

$$\mathbb{P}\left[F_\gamma(U) \leq p\right] = \mathbb{P}\left[U \leq F_\gamma^{-1}(p)\right] = (F_\gamma^{-1}(p) - x_n)/(x_{n+1} - x_n),$$

the minimum of ϕ_n is attained at the unique point $F_\gamma(y_n)$, where we recall that y_n is the midpoint $(x_n + x_{n+1})/2$.

To conclude the proof, it is now enough to remark that letting $w_n = \int_{y_{n-1}}^{y_n} \gamma(dz)$, we get $\sum_{m=1}^n w_m = F_\gamma(y_n)$, so that $(w_1, ..., w_N)$ minimizes all the terms in the sum (4.4.1). □

4.4.2 Probabilistic properties of the optimal weights

Let X_1, ..., X_n be *i.i.d.* $\mathcal{N}(0,1)$. In this section we investigate the behaviour of the $\mathcal{W}_1$ distance $D = D(\mathbf{X}) = \mathcal{W}_1(\sum_n w_n(\mathbf{X})\delta_{X_n}, \gamma)$ between the optimally reweighted sample and the target Gaussian measure. We start by proving the first part of Theorem 4.1.15: for any integer p,

$$\mathbb{E}\left[D^p\right] = \mathcal{O}^\star\left(\frac{1}{N^p}\right)$$

where $\mathcal{O}^\star$ means $\mathcal{O}$ up to logarithmic correction terms.

Proof of Theorem 4.1.15, first part. Let us first note that, since γ is absolutely continuous with respect to the Lebesgue measure, classical results on optimal transportation in dimension 1 for the usual distance (see for example [Vil03, Theorem 2.18] and the remarks that follow it) imply that the Monge-Kantorovitch problem (4.1.12) defining $\mathcal{W}_1(\eta, \gamma)$ distance has an explicit minimizer, given by the deterministic coupling $(T(Z), Z)$, where $Z \sim \gamma$ and T is the monotone transport map

$$\check{a}T(z) = F_\eta^{-1}(F_\gamma(z)).$$

Therefore, the optimal coupling between a Gaussian random variable X and the optimally reweighted empirical measure $\sum_n w_n(\mathbf{x})\delta_{x_n}$ is given by the piecewise constant transport map that sends each interval $]y_n, y_{n+1}[$ to x_n, so D has the explicit expression

$$D = \int \min_n |x - X_n| \, \gamma(dx).$$

We start by a rough bound: for any $\lambda > 0$, the Laplace transform $\exp\exp(\lambda D)$ may be bounded as follows using Jensen's inequality:

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda D)\right] &= \mathbb{E}\left[\exp\left(\lambda \int \min_n |x - X_n| \, \gamma(dx)\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\lambda \min_n |X - X_n|\right)\right],\end{aligned}$$

where $X \sim \gamma$ is independent of $\mathbf{X} = (X_1, ..., X_n)$. Then

$$\begin{aligned}\mathbb{E}\left[\exp(\lambda D)\right] &\leq \mathbb{E}\left[\exp\left(\lambda |X - X_1|\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\lambda |X|\right)\right]^2.\end{aligned}$$

Since the last expression is finite, we have established

$$\forall \lambda, \exists C_\lambda, \forall N, \quad \mathbb{E}\left[\exp\left(\lambda D\right)\right] \leq C_\lambda. \tag{4.4.2}$$

We now let $M < N$ be an integer and decompose the real line $\mathbb{R}$ in M segments between the quantiles $(x_i)_{0 \leq i \leq M}$, where $F_\gamma(x_i) = \int_{-\infty}^{x_i} \gamma(dx) = i/M$. We let G be the "M-well-spread event" (Definition 4.3.10) that there is at least one of the $(X_n)_{1 \leq n \leq N}$ in each of the M "bins" $(]x_{i-1}, x_i[)_{1 \leq i \leq M}$. We then proceed in three steps.

Step 1: D is small on the well-spread event. Indeed, on G, there exist $N(1), ..., N(M)$

such that $X_{N(i)} \in]x_{i-1}, x_i[$. Therefore

$$\begin{aligned}
D\mathbf{1}_G &= \mathbf{1}_G \int \min_n |x - X_n| \, \gamma(dx) \\
&= \mathbf{1}_G \sum_{i=1}^{M} \int_{x_{i-1}}^{x_i} \min_n |x - X_n| \, \gamma(dx) \\
&\leq \mathbf{1}_G \sum_{i=1}^{M} \int_{x_{i-1}}^{x_i} \left| x - X_{N(i)} \right| \gamma(dx) \\
&\leq \sum_{i=2}^{M-1} |x_i - x_{i-1}| \int_{x_{i-1}}^{x_i} \gamma(dx) + 2 \int_{x_{M-1}}^{\infty} |x - x_{M-1}| \, \gamma(dx) \\
&\leq \frac{2}{M} x_{M-1} + 2 \int_{x_{M-1}}^{\infty} |x - x_{M-1}| \, \gamma(dx) \\
&\leq \frac{2}{M} x_{M-1} + 2 \int_{x_{M-1}}^{\infty} x \gamma(dx) \\
&\leq \frac{2}{M} x_{M-1} + \frac{2}{\sqrt{2\pi}} \exp(-x_{M-1}^2/2).
\end{aligned}$$

From the classical gaussian tail estimate

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{t} - \frac{1}{t^3} \right) \exp(-t^2/2) \leq 1 - F_\gamma(t) \leq \frac{1}{\sqrt{2\pi}} \left(\frac{1}{t} \right), \tag{4.4.3}$$

applied to $t = x_{M-1}$, it is easily seen by taking logarithms that $x_{M-1} \sim \sqrt{2\log(M)}$. Using the first inequality in (4.4.3) again, we get $\exp(-x_{M-1}^2/2) = \mathcal{O}^\star(1/M)$, and finally

$$D\mathbf{1}_G = \mathcal{O}^\star \left(\frac{1}{M} \right).$$

Step 2: the well-spread event is very likely. Assuming from now on that M satisfies $N > (2p+2)M\ln(M)$, we get thanks to Lemma 4.3.12 that

$$\mathbb{P}[G^c] = \mathcal{O}(1/M^{2p+1}).$$

Step 3: conclusion. We decompose $\mathbb{E}[D^p]$ in two parts, depending on whether the sample X is well-spread or not. On G we use the result from Step 1; on G^c we apply Hölder's inequality, the bound on $\mathbb{P}[G^c]$ from step 2, and the *a priori* control on $\mathbb{E}[D^{2p}]$

given by the preliminary bound (4.4.2):

$$\begin{aligned}\mathbb{E}\left[D^p\right] &= \mathbb{E}\left[D^p \mathbf{1}_G\right] + \mathbb{E}\left[D^p \mathbf{1}_{G^c}\right] \\ &\leq \mathcal{O}^\star\left(\frac{1}{M^p}\right) + \sqrt{\mathbb{E}\left[D^{2p}\right]}\sqrt{\mathbb{P}\left[G^c\right]} \\ &\leq \mathcal{O}^\star\left(\frac{1}{M^p}\right) + \mathcal{O}\left(\frac{1}{M^{p+1/2}}\right) \\ &\leq \mathcal{O}^\star\left(\frac{1}{M^p}\right).\end{aligned}$$

Since M may be chosen large enough to guarantee $N = \mathcal{O}^\star(M)$, this implies $\mathbb{E}\left[D^p\right] = \mathcal{O}^\star\left(\frac{1}{N^p}\right)$. □

Proof of Theorem 4.1.15, second part. We now turn to the proof of the control in l^2 of the optimal weights, and show that

$$\mathbb{E}\left[\sum_{n=1}^{N} w_n(\mathbf{X})^2\right] \leq \frac{6}{N}.$$

By definition,

$$w_n(\mathbf{X}) = F_\gamma(Y_{n+1}) - F_\gamma(Y_n),$$

where the Y_n are the middle points of the reordered sample and F_γ is the cdf of the standard Gaussian distribution. By a rough upper bound, for $2 \leq n \leq N-1$,

$$w_n \leq F_\gamma(X_{(n+1)}) - F_\gamma(X_{(n-1)}).$$

The cdf F_γ maps the ordered sample $(X_{(1)}, ..., X_{(N)})$ to an ordered sample $(U_{(1)}, ..., U_{(N)})$ of the uniform distribution on $[0,1]$, so

$$w_n \leq U_{(n+1)} - U_{(n-1)},$$

for $2 \leq n \leq N-1$, $w_1 \leq U_{(2)}$ and $w_n \leq 1 - U_{(N-1)}$.

Let us upper bound $\mathbb{E}\left[w_n^2\right]$ for $2 \leq n \leq N-1$, using known results on order statistics for uniform variables that may be found e.g. in [Das11, Chapter 6, Theorem 6.6]. Conditionnally on $U_{(n+1)} = u$, $U_{(n-1)}$ is distributed like the second largest value in a sample of n uniform variables on $[0,u]$, that is, like

$$uU^{1/n}V^{1/(n-1)}$$

where U and V are iid uniform on $[0,1]$. Therefore

N	Number of samples
M	Number of repetitions
h	Bandwidth

Table 4.1: Notation for the numerical tests

$$\begin{aligned}\mathbb{E}\left[w_n^2\right] &\leq \mathbb{E}\left[U_{(n+1)}^2(1-U^{1/n}V^{1/(n-1)})^2\right] \\ &= \mathbb{E}\left[U_{(n+1)}^2\right]\left(1-2\int u^{1/n}v^{1/(n-1)}dudv+\int u^{2/n}v^{2/(n-1)}dudv)\right) \\ &= \frac{6}{(n+1)(n+2)}\mathbb{E}\left[U_{(n+1)}\right]^2.\end{aligned}$$

Now $U_{(n+1)}$ follow a Beta$(n+1, N-n)$ distribution, so

$$\begin{aligned}\mathbb{E}\left[U_{(n+1)}^2\right] &= \mathrm{Var}(U_{(n+1)}) + \mathbb{E}\left[U_{(n+1)}\right]^2 \\ &= \frac{(n+1)(N-n)}{(N+1)^2(N+2)} + \frac{(n+1)^2}{(N+1)^2} \\ &= \frac{(n+1)(n+2)}{(N+1)(N+2)},\end{aligned}$$

so that $\mathbb{E}\left[w_n^2\right] \leq \frac{6}{(N+1)(N+2)}$, for all $2 \leq n \leq N-1$. One easily checks that this bound also holds for $n = 1$ and N, and by summing we get

$$\mathbb{E}\left[\sum_{n=1}^{N} w_n^2(\mathbf{X})\right] \leq \frac{6N}{(N+1)(N+2)} \leq \frac{6}{N}.$$

□

4.5 Numerical experiments I

In this section we focus on the comparison between the weighted empirical measures $\overline{\eta}_N$, $\eta^\star_{h,N}$ and $\eta^\star_{\mathrm{Wass},N}$.

4.5.1 Implementation

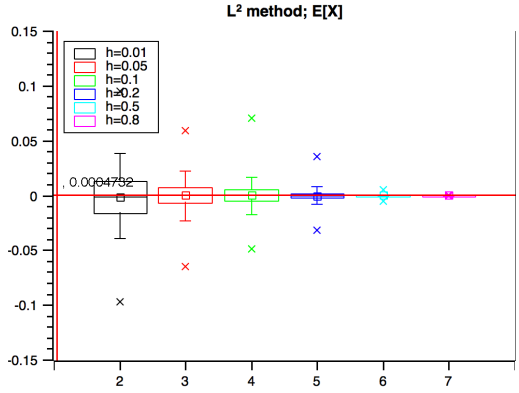
The implementation of the Wasserstein method is straightforward: given $(\mathbf{x})$, we only need to sort it, compute the middle points (y_n) and deduce the weights by applying F_γ.

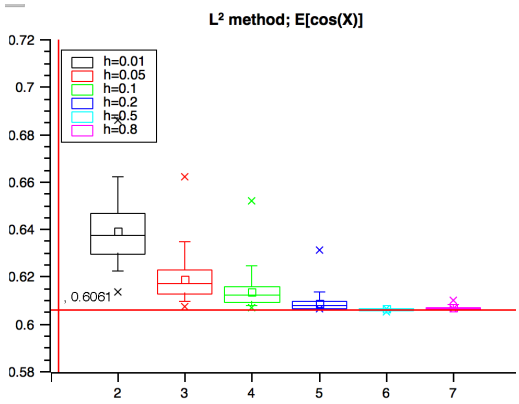
For the L^2 method, the quadratic programming optimization problem 4.1.3 ($\delta \simeq 0$ case) is solved using a standard Scilab library based on the dual iterative method detailed in [GI83].

The methods are then tested by computing estimators for the expected value of three functions of X: $\mathbb{E}[X]$, $\mathbb{E}[\cos(X)]$ and $\mathbb{E}[\mathbf{1}_{X>1}]$. The estimators are computed on samples of size N, and the experiment is repeated M times. We present the results as boxplots representing the quantiles on the M repetitions.

4.5.2 Regularity of the test function and choice of the bandwidth

We first investigate the influence of the bandwidth parameter h on the L^2 method, by testing various values of $h \in \{0.01, 0.05, 0.1, 0.2, 0.5, 0.8\}$ on the three test functions ϕ: a) $x \mapsto x$, b) $x \mapsto \cos(x)$ and c) $x \mapsto \mathbf{1}_{|x|>1}$.





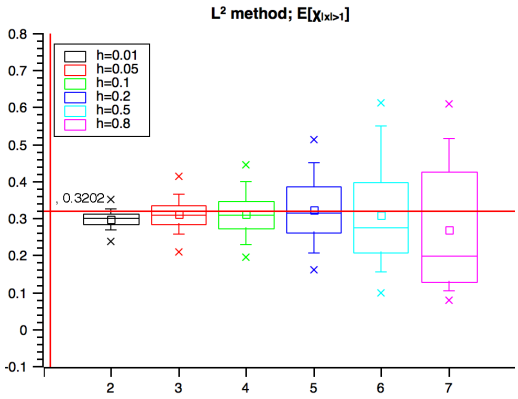


Figure 4.4: $M = 1000$, $N = 100$

Figure ?? corresponds to the test function $x \mapsto x$ which is very specific, the symmetry ensures that the estimator is unbiased, and the method seems to be better the larger h is. Figure ?? corresponds to the test function cos; a bias clearly appears in that case, and the estimator is better when h is quite large, with a trade-off at $h = .5$ ($h = .8$ is not as good). In both cases, the fact that the estimators are better when h is quite large may be linked to two remarks made above:

- Remark 4.3.5 where it is recalled that $x \mapsto x$ and cos are regular test functions that belongs to the image by the Orstein-Uhlenbeck semi-group P_t of an L^2 function on which the optimization is based on;

- Remark 4.1.14 where it is suggested that the more 'regular' this test function is, the larger the optimal h should be.

However, when we apply the method to estimate the expectation of a discontinuous function of X, here $x \mapsto \mathbf{1}_{|x|>1}$, which does not belong to the appropriate class of regularity, the picture is completely different and the best estimator is obtained for a much smaller $h \approx 0.05$, as can be seen in Figure 4.4.

4.5.3 Comparison between naïve, L^2 and Wasserstein

Next, we compare the naïve Monte Carlo method, the L^2 reweighting ($h = .5$ and $h = .05$) and the Wasserstein reweighting.

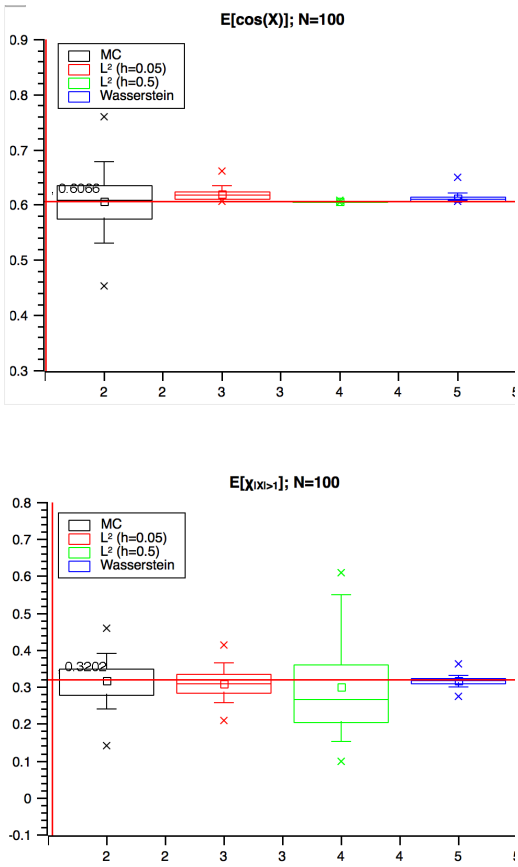


Figure 4.5: $M = 1000$, $N = 100$

Figure ?? corresponds to the cos test function case, and the naïve Monte Carlo approach is outperformed by all reweighting methods, even with sub-optimal tuning (L^2 for $h = .05$). On the contrary, in Figure ?? which corresponds to the step test function the naïve Monte Carlo approach is much better than the L^2 reweighting methods with sub-optimal tuning ($h = .5$), and similar to the L^2 reweighting methods with quasi-optimal tuning ($h = .05$). This is consistent with the fact that the L^2 reweighting method has been derived for regular test functions, which excludes the step function case.

The Wasserstein reweighting is in both cases (cos and step) much better than the naïve Monte Carlo, and better than the L^2 reweighting in the step function case. For the cos case, the Wasserstein reweighting is similar to the L^2 reweighting method with sub-optimal $h = .05$ and much worse than the L^2 reweighting method with quasi-optimal $h = .5$.

4.5.4 Conclusion

The Wasserstein reweighting is more robust (no parameter to tune) than the L^2 reweighting, and outperforms the latter for irregular test functions. However, for sufficiently regular test functions and with well-chosen bandwidth h, the L^2 reweighting is much better.

4.6 Numerical experiments II

In this section, we present numerical results exhibiting the variance reduction obtained with the reweighting method.

For simplicity, and having in mind the various drawbacks of the L^2 method in terms of speed and parameter tuning, we will only focus on weights computed with a Wasserstein distance in the minimization problem (4.1.3) — that is, the minimization problem with $\delta = 0$.

4.6.1 Exchangeable functions of Gaussian vectors

Let $(G_1, \dots, G_N)$ denotes a sequence of N i.i.d. centered Gaussian vectors in $\mathbb{R}^d$ with identity covariance matrix. We consider the problem of reducing the variance of Monte Carlo estimators of the distribution of $F(G)$ where

$$F : \mathbb{R}^d \to \mathbb{R}$$

is a smooth non-linear function, which is invariant by permutation of the d coordinates (exchangeability). We assume for simplicity the following normalization:

$$F(0) = 0, \qquad D_0 F = (1/\sqrt{d}, \dots, 1/\sqrt{d})$$

and set for each $n = 1, \dots, N$:

$$X_n := (D_0 F) \cdot G_n \sim \mathcal{N}(0,1), \qquad Y_n := F(G_n).$$

We are then interested in estimating the cumulant generating function of the distribution of $F(G)$ denoted by

$$k_Y(t) := \log \mathbb{E}\left(\mathrm{e}^{tY}\right) = \log \mathbb{E}\left(\mathrm{e}^{tF(G)}\right),$$

and possibly to compare it to the cumulant generating function of the distribution of the standard Gaussian distribution

$$j_X(t) := \log \mathbb{E}\left(\mathrm{e}^{tX}\right) = \log \mathbb{E}\left(\mathrm{e}^{t(D_0F)\cdot G}\right) = \frac{t^2}{2}.$$

We will consider, compare and combine various estimators. The first two are the naïve and Wasserstein reweighted estimators of k_Y, defined by

$$\mathbf{k}_{\mathrm{MC}}(\mathbf{Y})(t) = \log \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}^{tY_n} \qquad \mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})(t) = \log \sum_{n=1}^{N} w_n(\mathbf{X}) \mathrm{e}^{tY_n},$$

where the weights $w(\mathbf{X})$ are computed with the control variables $\mathbf{X}$ through the minimization problem (4.1.3) associated with the (Euclidean-based) Wasserstein distance.

We define similarly two estimators for j_X,

$$\mathbf{j}_{\mathrm{MC}}(\mathbf{Y})(t) = \log \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}^{tX_n} \qquad \mathbf{j}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})(t) = \log \sum_{n=1}^{N} w_n(\mathbf{X}) \mathrm{e}^{tX_n}.$$

Since $j_X(t)$ is explicit, it is quite natural to try and use e^{tX_n} as a control variate, leading to a new estimator:

$$\mathbf{k}_{\mathrm{CV}}(\mathbf{Y})(t) = \mathbf{k}_{\mathrm{MC}}(\mathbf{Y})(t) - \mathbf{j}_{\mathrm{MC}}(\mathbf{X})(t) + j_X(t),$$

Finally, we combine the reweighting and the control variate idea by defining

$$\mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y})(t) = \mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})(t) - \mathbf{j}_{\mathrm{W}}(\mathbf{X})(t) + j_X(t),$$

Note that the control variate has been defined as the best linear approximation of F around the mean $0 \in \mathbb{R}^d$.

We run our tests with the following particular choice of a non-linear function:

$$F_r(g) = \frac{1}{r} \sin\left(\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \sin(r \times g^i) \right)$$

with the parameters $d = 10$ and $r \in \{0.1, 1\}$. Note that r encodes the strength of the nonlinearity, in the sense that

$$\lim_{r \to 0} F_r(g) = g.$$

In all this section, we have taken samples of size $N = 30$. This choice has been made so that the quantiles of the estimators scales appropriately with the target function k_Y to be estimated.

4.6.1.1 The almost linear case, $r = .1$

This case corresponds to a function F which is close to the identity function. In Fig. 4.6, we have represented the quantiles of the different estimators of $k_Y(t)$ for $t \in [-0.7, 0.7]$.

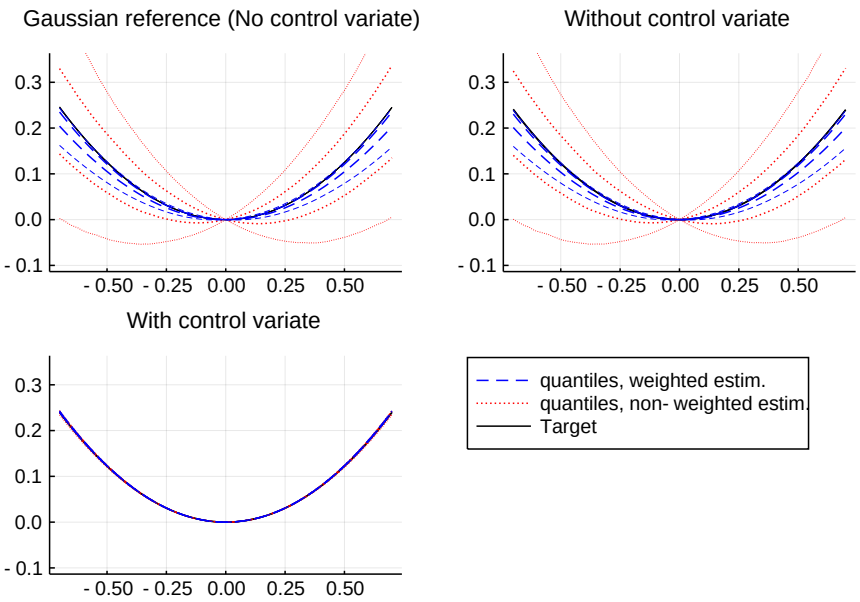


Figure 4.6: **Case** $r = .1$. The figures above represents the $[.05, .25, .75, .95]$-quantiles of the different estimators of the cumulant generating functions for $F(G)$, both without weights (dotted), and with weight (dashed). The figure in the upper left corner represents the estimators of the Gaussian reference $\mathbf{j}_{\mathrm{MC}}(\mathbf{X})$ and $\mathbf{j}_{\mathrm{W}}(\mathbf{X})$. The figure in the upper right corner represents the estimators $\mathbf{k}_{\mathrm{MC}}(\mathbf{Y})$ and $\mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})$ (without control variate). Finally, the figure in the lower left corner represents the estimators of $\mathbf{k}_{\mathrm{CV}}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y})$ (with control variate).

In Fig. 4.7, we zoom in on the figure the lower left corner of Fig. 4.6 where a linear control variate is used, by plotting the difference $\mathbf{k}_{\mathrm{CV}}(\mathbf{X}, \mathbf{Y}) - j_X = \mathbf{k}_{\mathrm{MC}}(\mathbf{Y})(t) - \mathbf{j}_{\mathrm{MC}}(\mathbf{X})(t)$ as well as $\mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y}) - j_X = \mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y}) - \mathbf{j}_{\mathrm{W}}(\mathbf{X})$.

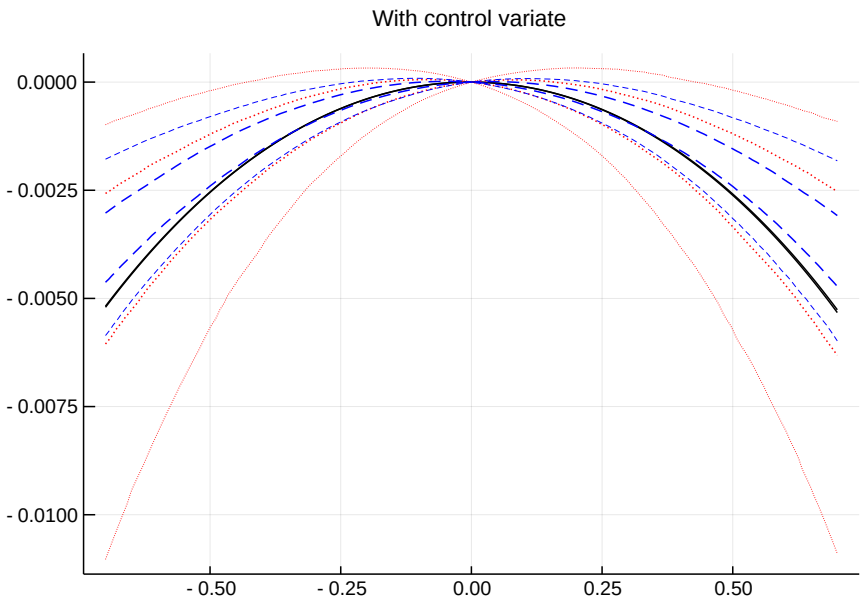


Figure 4.7: **Case** $r = .1$. The figure above represents the $[.05, .25, .75, .95]$-quantiles of $\mathbf{k}_{\mathrm{CV}}(\mathbf{X}, \mathbf{Y}) - j_X$ (dotted) and $\mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y}) - j_X$ (dashed).

4.6.1.2 The nonlinear case, $r = 1$

This case corresponds to a function F with a significant non-linear behavior. In Fig. 4.8, we have represented the quantile envelopes of the different estimators of $k_Y(t)$ for $t \in [-0.7, 0.7]$.

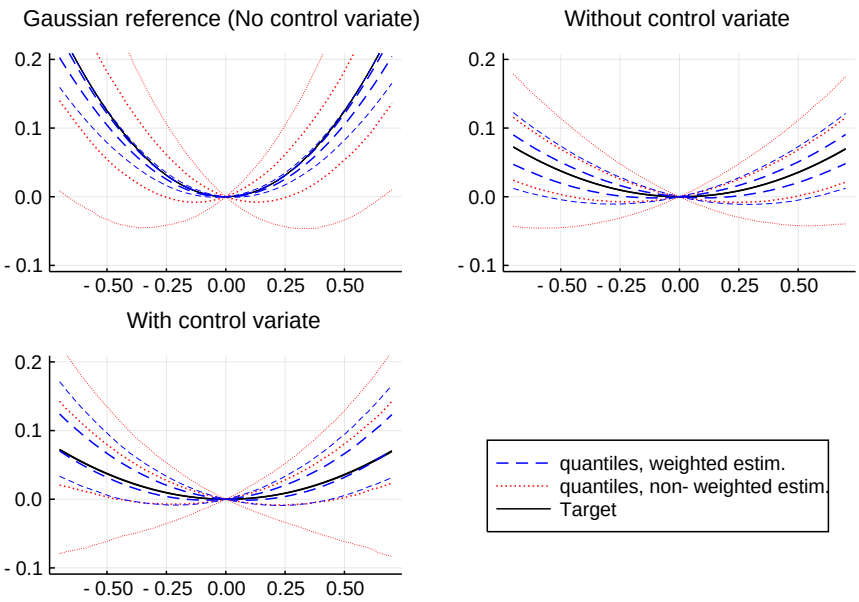


Figure 4.8: **Case** $r = 1$. The figures above represents the $[.05, .25, .75, .95]$-quantiles of the different estimators of the cumulant generating functions for $F(G)$, both without weights (dotted), and with weight (dashed). The figure in the upper left corner represents the estimators of the Gaussian reference $\mathbf{j}_{\mathrm{MC}}(\mathbf{X})$ and $\mathbf{j}_{\mathrm{W}}(\mathbf{X})$. The figure in the upper right corner represents the estimators $\mathbf{k}_{\mathrm{MC}}(\mathbf{Y})$ and $\mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})$ (without control variate). Finally, the figure in the lower left corner represents the estimators of $\mathbf{k}_{\mathrm{CV}}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{k}_{\mathrm{CV+W}}(\mathbf{X}, \mathbf{Y})$ (with control variate).

4.6.1.3 Interpretation

First note that the non-linearity of the function F in the case $r = .1$ has a non-negligible influence on the distribution of $F(G)$, as can be seen in the upper right and lower left figures of Fig. 4.8, where the cumulant generating function $k_Y(t)$ (the 'target', represented with a full line) is substantially different from the Gaussian reference $j_X(t)$ (the 'Gaussian reference', represented with a full thin line), and has a much smaller variance.

We first immediately observe that in all cases (the Gaussian reference, the estimator of k_y without control variate, and the estimator of k_y with control variate) the use of the studied weighting method substantially improve the estimation by:

1. Significantly reducing the spread of the tail distribution (the $\{.1, .9\}$-quantiles) of the estimators.

2. Significantly reducing the statistical error of the typical outcomes (the $\{.25, .75\}$–quantiles) of the estimators.

Then we can observe that as expected, the error reduction due to the weighting method

is slightly better for the Gaussian reference. However, it is clear that the error reduction due to the weighted method is very significant in each case. For instance the typical error (as given by the $\{.25, .75\}$–quantiles) of the estimator $\mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})$ is reduced almost by a factor 2 as compared to $\mathbf{k}_{\mathrm{MC}}(\mathbf{Y})$. As a reference, the typical error on $\mathbf{j}_{\mathrm{W}}(\mathbf{X})$ is reduced by a factor 5 as compared to $\mathbf{j}_{\mathrm{MC}}(\mathbf{X})$.

Finally, it is remarkable to notice that in the case $r = 1$ the control variate method is useless and may even be counterproductive. On the contrary the weighting method behaves well and reduces the error (with or without control variate). It clear from Fig.4.8 that the weighting method outperforms the control variate method which is not useful here.

This experiment demonstrates that the weighting method can then very easily and very efficiently be used to reduce the statistical error caused by non-linear functions, without resorting to *ad hoc* analytic calculations.

4.6.2 A physical toy example

4.6.2.1 Model

In this section, we illustrate the use of the weighting method with a more concrete, physical example. We consider a Langevin stochastic differential equation in $\mathbb{R}^d$

$$\begin{cases} dQ_t = P_t dt \\ dP_t = -Q_t dt + \varepsilon \mathcal{F}(Q_t) dt - P_t dt + \sqrt{2} dW_t \end{cases}$$

which is a toy model for a thermostatted linear mechanical system. The latter is perturbed out of equilibrium by an exterior force field $\mathcal{F} : \mathbb{R}^d \to \mathbb{R}^d$, and we are interested in computing the distribution of the long time stationary back reaction, that is to say the distribution of $\mathcal{F}(Q)$ where $Q \in \mathbb{R}^d$ is distributed according to the invariant distribution of the Langevin process, and this for ε small.

For simplicity we assume that $\mathcal{F}_i(Q) = F(Q) \in \mathbb{R}$ is independent of i with again

$$F(0) = 0, \qquad D_0 F = (1/\sqrt{d}, \ldots, 1/\sqrt{d}).$$

We set $X_n = F(Q_{\tau n})$ where τ is a sufficiently large decorrelation time. We also set $Y_n = D_0 F \cdot \tilde{Q}_{\tau n}$ where $\tilde{Q}$ is solution to the coupled, non perturbed linear system

$$\begin{cases} d\tilde{Q}_t = \tilde{P}_t dt \\ d\tilde{P}_t = -\tilde{Q}_t dt - \tilde{P}_t dt + \sqrt{2} dW_t \end{cases}$$

so that Y_n is a Gaussian sequence of unit standard Gaussian variables that are approximately independent (for large τ). Using elementary calculations (see *e.g.* [BGM10]), one can check that the positive definite quadratic Lyapunov functional

$$D_t := \left|P_t - \tilde{P}_t\right|^2 + \left|Q_t - \tilde{Q}_t\right|^2 + \left(Q_t - \tilde{Q}_t\right) \cdot \left(P_t - \tilde{P}_t\right)$$

satisfies almost surely the following differential inequality:

$$\frac{d}{dt} D_t \leq -\frac{1}{2} D_t + 4 D_t^{1/2} \varepsilon \left|F(Q_t)\right|.$$

Assuming for simplicity that $\|F\|_\infty = 1$, a Gronwall-type integration yields that for any $t \geq 0$

$$\left|P_t - \tilde{P}_t\right|^2 + \left|Q_t - \tilde{Q}_t\right|^2 \leq c D_t \leq \varepsilon + \mathcal{O}\left(\mathrm{e}^{-t/2}\right),$$

for some numerical constant c. Hence (Q_t, P_t) converges when $\varepsilon \to 0$ to the Orstein-Uhlenbeck process $(\tilde{Q}_t, \tilde{P}_t)$ uniformly in time. This coupling calculation thus suggests that $(Q_{\tau n})_{n \geq 1}$ will be close to a i.i.d. Gaussian sequence when $\varepsilon \to 0$ and $\tau \gg 1$.

4.6.2.2 Numerical experiment

We present in Figure 4.9 some numerical results in a test case with the same non-linear function:

$$F(g) = \sqrt{d} \sin\left(\frac{1}{d} \sum_{i=1}^{d} \sin(g^i)\right)$$

with the parameters $(\varepsilon = .01, d = 10, N = 50)$. The methodology is the same as in the last section.

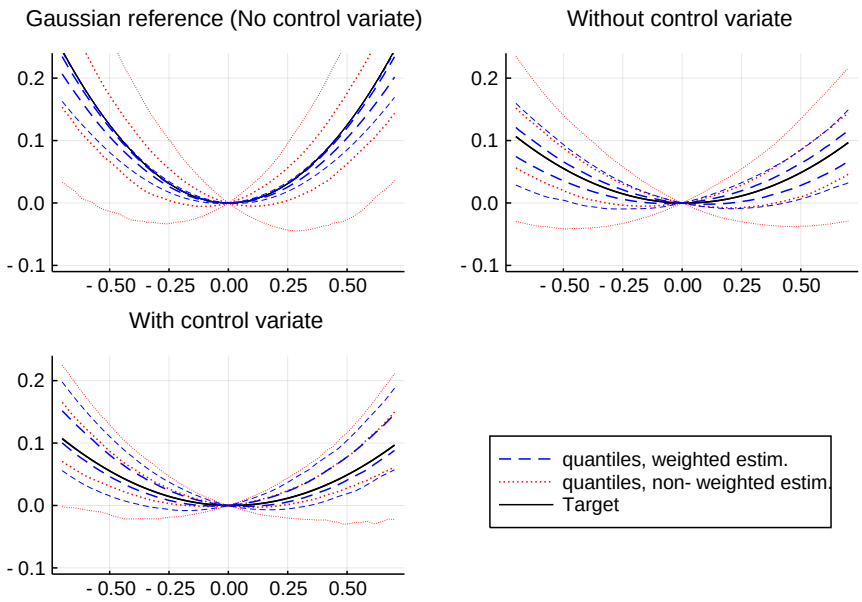


Figure 4.9: The figures above represents the $\{.05, .25, .75, .95\}$-quantiles of the different estimators of the cumulant generating functions for the stationary distribution of the exterior force $F(Q)$, both without weights (dotted), and with weight (dashed). The figure in the upper left corner represents the estimators of the Gaussian reference $\mathbf{k}_{\mathrm{MC}}(\mathbf{X})$ and $\mathbf{k}_{\mathrm{W}}(\mathbf{X})$. The figure in the upper right corner represents the estimators $\mathbf{k}_{\mathrm{MC}}(\mathbf{Y})$ and $\mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})$ (without control variate). Finally, the figure in the upper right corner represents the estimators of $\mathbf{k}_{\mathrm{Cv}}(\mathbf{X}, \mathbf{Y})$ and $\mathbf{k}_{\mathrm{W}}(\mathbf{X}, \mathbf{Y})$ (without control variate)

4.6.2.3 Interpretation

We first remark that the target distribution has an increased variance due to the presence of $\varepsilon \neq 0$. We then remark that the result are very similar as in previous section, except that the statistical error reduction in the case with weights is similar with control variate or without control variate. For any choice of estimator (with or without control variate), we see that the use of weighting substantially improve the statistical error.

4.6.3 Conclusion

In various non-trivial cases where a random quantity is approximated by a Gaussian one dimensional control variate, the Wasserstein reweighting method significantly reduces variance (as compared to a naïve Monte Carlo calculation), and outperforms a control variate variance reduction.

4.6.4 Acknowledgments

We thank P.-M. Samson for suggesting the short proof of Lemma 4.3.15, and J. Bigot for many constructive remarks that led to many clarifications, and a much nicer proof of Proposition 4.4.1. This work was partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement number 614492.

Appendix A

Appendix A

A.1 Stopped martingale problem

Let E be a Polish space. Let L be a linear operator mapping a given space $\mathcal{D} \subset C_b(E)$ into bounded measurable functions. Let μ be a probability distribution on E. Let $U \subset E$ be an open set. A càd-làg process $(X_t)_{t\geq 0}$ with values in E solves the *stopped martingale problem* for the generator L on the space $\mathcal{D}$ with initial measure μ and domain U — in short, X solves $\mathbf{sMP}(L, D(L), \mu, U)$ — if, denoting

$$\tau_U := \inf\left\{t \geq 0 | X_t \notin U \text{ or } X_{t^-} \notin U\right\},$$

(i) $\text{Law}(X_0) = \mu$; (ii) $X_t = X_{t\wedge\tau_U}$; and (iii) if for any $\varphi \in \mathcal{D}$,

$$t \mapsto M_t(\varphi) := \varphi(X_t) - \varphi(X_0) - \int_0^{t\wedge\tau_U} L\varphi(X_s)ds$$

is a martingale with respect to the natural filtration $\left(\mathcal{F}_t^X = \sigma\left(X_s,\, 0 \leq s \leq t\right)\right)_{t\geq 0}$.
Moreover, the stopped martingale problem $\mathbf{sMP}(L, \mathcal{D}, \mu, U)$ is said to be *well-posed* if:

- There exists a probability space and a càd-làg process defined on it that solves the stopped martingale problem (existence);
- whenever two processes solve $\mathbf{sMP}(L, \mathcal{D}, \mu, U)$, then they have the same distribution on $\mathbb{D}_E$ (uniqueness).

The following theorem is a synthesis of the localization technique of Theorem 6.1 and 6.2 of [EK86, Chapter 4]. It gives a simple criteria ensuring equivalence of uniqueness between (i) a global martingale problem, and (ii) local stopped martingale problems.

Theorem A.1.1. *Let $(U_k)_{k\in K}$ be a countable family of open subsets of E such that $\bigcup_{k\in K} U_k = E$. Assume that for any initial ν, there exists a solution to $\boldsymbol{MP}(L, \mathcal{D}, \mu)$.*

Then uniqueness of $\boldsymbol{MP}(L, \mathcal{D}, \mu)$ *for all* μ *is equivalent to uniqueness of* $\boldsymbol{sMP}(L, \mathcal{D}, \mu, U_k)$ *for all* μ *and all* $k \in K$*.*

Bibliography

[AS92] Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions with formulas, graphs, and handbook of mathematical functions with formulas, graphs, and mathematical tables*, Dover Publications, Inc., New York, 1992.

[BGM10] François Bolley, Arnaud Guillin, and Florent Malrieu, *Trend to equilibrium and particle approximation for a weakly selfconsistent vlasov-fokker-planck equation*, ESAIM: Mathematical Modelling and Numerical Analysis **44** (2010), no. 5, 867–884.

[CLS07] E. Cancès, F. Legoll, and G. Stoltz, *Theoretical and numerical comparison of some sampling methods for molecular dynamics*, M2AN Math. Model. Numer. Anal. **41** (2007), no. 2, 351–389.

[Das11] Anirban DasGupta, *Probability for statistics and machine learning*, Springer Texts in Statistics, Springer, New York, 2011, Fundamentals and advanced topics. MR 2807365

[EK86] S. N. Ethier and T. G. Kurtz, *Markov processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1986, Characterization and convergence. MR 838085

[GI83] D. Goldfarb and A. Idnani, *A numerically stable dual method for solving strictly convex quadratic programs*, Mathematical Programming **27** (1983), no. 1, 1–33.

[Gla13] Paul Glasserman, *Monte carlo methods in financial engineering*, vol. 53, Springer Science & Business Media, 2013.

[Gly94] P.W. Glynn, *Efficiency improvement techniques*, Annals of Operations Research **53** (1994), no. 1, 175–197

[Gro93] Leonard Gross, *Logarithmic sobolev inequalities and contractivity properties of semigroups*, Dirichlet forms, Springer, 1993, pp. 54–88.

[GS02] Alison L Gibbs and Francis Edward Su, *On choosing and bounding probability metrics*, International statistical review **70** (2002), no. 3, 419–435.

[Jou09] Benjamin Jourdain, *Adaptive variance reduction techniques in finance*, Advanced Financial Modelling **8** (2009), 205.

[Kal02] Olav Kallenberg, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, 2002.

[Kha66] R. Z. Khas'minskii, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory of Probability & Its Applications **11** (1966), no. 3, 390–406.

[KSTT02] R. Kupferman, A. M. Stuart, J. R. Terry, and P. F. Tupper, *Long-term behaviour of large mechanical systems with random initial data*, Stoch. Dyn. **2** (2002), no. 4, 533–526.

[LC02] Lars Larsson-Cohn, *L p -norms of hermite polynomials and an extremal problem on wiener chaos*, Ark. Mat. **40** (2002), no. 1, 133–144.

[LRS10] T. Lelièvre, M. Rousset, and G. Stoltz, *Free energy computations*, Imperial College Press, London, 2010, A mathematical perspective. MR 2681239

[MR95] Rajeev Motwani and Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, Cambridge, 1995. MR 1344451

[Øks03] B. Øksendal, *Stochastic differential equations*, sixth ed., Universitext, Springer-Verlag, Berlin, 2003, An introduction with applications. MR 2001996

[Owe13] Art B. Owen, *Monte carlo theory, methods and examples*, 2013.

[PK74] G. C. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math. **27** (1974), 641–668. MR 0368142

[PS08] G. Pavliotis and A. Stuart, *Multiscale methods: averaging and homogenization*, Springer Science & Business Media, 2008.

[PS18] François Portier and Johan Segers, *Monte carlo integration with a growing number of control variates*, arXiv preprint arXiv:1801.01797 (2018).

[PSV77] G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan, *Martingale approach to some limit theorems*, ii+120 pp. Duke Univ. Math. Ser., Vol. III. MR 0461684

[PV73] G. C. Papanicolaou and S. R. S. Varadhan, *A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations*, Comm. Pure Appl. Math. **26** (1973), 497–524. MR 0383530

[PV01] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Ann. Probab. **29** (2001), no. 3, 1061–1085. MR 1872736

[PV03] ———, *On Poisson equation and diffusion approximation. II*, Ann. Probab. **31** (2003), no. 3, 1166–1192. MR 1988467

[PV05] ———, *On the Poisson equation and diffusion approximation. III*, Ann. Probab. **33** (2005), no. 3, 1111–1133. MR 2135314

[Ros13] Sheldon M. Ross, *Simulation*, Elsevier/Academic Press, Amsterdam, 2013, Fifth edition [of 1433593]. MR 3294208

[SLS+06] A. Scemama, T. Lelièvre, G. Stoltz, E. Cancès, and M. Caffarel, *An efficient sampling algorithm for variational monte carlo*, The Journal of chemical physics **125** (2006), no. 11, 114105.

[SSMD82] J. M. Sancho, M. San Miguel, and D. Dürr, *Adiabatic elimination for systems of Brownian particles with nonconstant damping coefficients*, J. Statist. Phys. **28** (1982), no. 2, 291–305. MR 666513

[Str63] R. L. Stratonovich, *Topics in the theory of random noise. Vol. I: General theory of random processes. Nonlinear transformations of signals and noise*, Revised English edition. Translated from the Russian by Richard A. Silverman, Gordon and Breach Science Publishers, New York-London, 1963. MR 0158437

[Str11] Daniel W. Stroock, *Probability theory*, second ed., Cambridge University Press, Cambridge, 2011, An analytic view. MR 2760872

[SV07] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Springer, 2007.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003)

[Vil08] ———, *Optimal transport: old and new*, vol. 338 Springer Science & Business Media, 2008.

[Zwa73] R. Zwanzig, *Nonlinear generalized langevin equations*, Journal of Statistical Physics **9** (1973), no. 3, 215–220.