SINGULAR MCKEAN-VLASOV PROBLEMS FROM MATHEMATICAL PHYSICS AND FINANCE

SCANDER MUSTAPHA

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for acceptance

by the Program in

Applied and Computational Mathematics

Adviser: Mykhaylo Shkolnikov

May 2024

Abstract

The purpose of this thesis is to develop techniques for analyzing the limiting McKean-Vlasov dynamics of interacting particle systems featuring singularities, and arising in physics and mathematical finance.

We first investigate the asymptotic stability of unidimensional log gases under non-convex confining potentials by establishing new Entropy-Wassertein-Information (HWI) inequalities. Such gases are obtained as the mean-field limit of particles interacting via a repulsive logarithmic potential.

Then, we establish the well-posedness of the supercooled Stefan problem with oscillatory initial condition. This classical problem from mathematical physics is reformulated using a probabilistic description of the free boundary as a cumulative distribution function of the hitting time of a Brownian motion with a jumping drift.

Finally, we study the well-posedness problem of a class of bidimensional stochastic differential equations (SDE), whose coefficients depend on the joint density of the unknown process. This class of local stochastic volatility models is important for the calibration of volatility surfaces. Additionally, we solve the long-standing problem of joint S&P 500/VIX calibration by using SDEs controlled by neural networks.

Acknowledgments

I express my deepest gratitude to Professor Shkolnikov for introducing me to very interesting and original mathematical problems. I am grateful for the time spent helping me solve difficult technical questions.

I am very grateful to Professor Guyon for my internships and the fruitful collaboration.

I am thankful to Professor Tangpi for being a reader of this thesis and to Professor Hubert and Professor Soner for being members of my FPO committee.

I thank Audrey Mainzer, Bernadeta Wysocka and Victoria Beltra of the PACM department for guiding me through the program.

To my father.

Contents

List of Figures

Chapter 1

Introduction

In this dissertation, we are mainly concerned with the analysis of the mean-field limit of unidimensional particle systems $(X_t^{i,N})_{t\geq 0, 1\leq i\leq N}$, interacting with each other via a typical dynamic of the form

$$dX_t^{i,N} = b\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}}\right)dt + \sigma\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}}\right)dB_t^i,\ 1\leq i\leq N, \quad (1.0.1)$$

where $N \geq 1$ is the number of particles, $(B_t^i)_{t\geq 0, 1\leq i\leq N}$ are i.i.d standard Brownian motions. The specificity of such systems is the dependence of the drift coefficient $b : \mathbb{R}_+ \times \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ and the volatility coefficient $\sigma : \mathbb{R}_+ \times \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}_+$ on the averaged field $\frac{1}{N}\sum_{i=1}^{N}\delta_{X_t^{i,N}}$, which models the interaction with other particles. Here $\mathcal{P}(\mathbb{R})$ is the set of probability measures on $\mathbb{R}$. The initial positions of the particles $(X_0^i)_{1\leq i\leq N}$ are assumed to be identically distributed and independent of each other and of the Brownian motions.

The mean-field limit is obtained when $N \to \infty$, and is described by the so-called

McKean-Vlasov stochastic differential equation (SDE)

$$\begin{cases} dX_t = b(t, X_t, \mu_t)dt + \sigma(t, X_t, \mu_t)dB_t, \\ \mathbb{P}\left(X_t \in dx\right) = \mu_t(dx), \\ X_{t=0} = X_0, \end{cases} \tag{1.0.2}$$

where the unknown is $(X_t, \mu_t)_{t\geq 0}$, $(B_t)_{t\geq 0}$ is a real standard Brownian motion and X_0 is a real random variable of law $X_0^{1,1}$. Such equations, in which the dynamic is non-Markovian and depends on the law of the underlying process, have various applications in physics, mathematical finance, economics, biology and neuroscience (see e.g. [88], [87], [66], [17], [34]). The convergence of the particle system (1.0.1) towards the limiting problem (1.0.2) is formalized through the notion of propagation of chaos (see e.g. [68] and [106]). When the propagation of chaos property holds, particles become asymptotically independent and for fixed $k \geq 1$, the law $\mathcal{L}_N$, $N \geq k$, of the process $(X_t^{1,N}, \ldots, X_t^{k,N})_{t\geq 0}$, converges towards the independent product $\mathcal{L} \otimes \cdots \otimes \mathcal{L}$, where $\mathcal{L}$ is the law of a solution $(X_t)_{t\geq 0}$ to (1.0.2).

On the mathematical side, the analysis of SDE (1.0.2) is challenging due to the lack of regularity of the law $(\mu_t)_{t\geq 0}$ of a solution and the non-linear aspect of the equation. Indeed, if we impose that $\mu_t, t \geq 0$ has a density $p(t, \cdot)$ and that $b(t, x, \mu_t)$ and $\sigma(t, x, \mu_t)$ depend on μ_t via $p(t, x)$, then p solves formally the quasi-linear Fokker-Planck equation

$$\begin{cases} \partial_t p(t,x) = \frac{1}{2}\partial_{xx}\left[\sigma^2(t,x,p(t,x))p(t,x)\right] - \partial_x\left[b(t,x,p(t,x))p(t,x)\right], \ (t,x) \in \mathbb{R}_+ \times \mathbb{R}, \\ p(0,x) = P(x), \ x \in \mathbb{R}, \end{cases} \tag{1.0.3}$$

where P denotes the density of the law of X_0. When the coefficients b and σ are Hölder continuous , the analysis of equation (1.0.3) can be made using the classical results of Ladyzhenskaya et al. [77] (see e.g. [65]).

When b and σ are more irregular, the lack of general theory forces us to exploit the specific structure of the equation at hand. In this work, we specifically address the problems of well-posedness and asymptotic stability of equations of type (1.0.1), (1.0.2) and (1.0.3), where the drift or the volatility exhibit singularities, by developing new techniques. More specifically, we consider the problem of asymptotic stability of log gases under non-convex external potential, the well-posedness for the supercooled Stefan problem under oscillatory initial condition and the well-posedness of a local stochastic volatility model. An additional chapter is dedicated to the more practical problem of joint SPX and VIX calibration.

1.1 Asymptotic stability of log-gases

The first chapter is dedicated to the study of the asymptotic stability of the mean-field limit of the generalized Dyson Brownian motions or log gases, defined as the system of $N \geq 1$ interacting particles

$$\begin{cases} dX_t^{i,N} = \sqrt{\frac{2}{\beta N}} dB_t^i + \frac{1}{N} \sum_{1 \leq j \neq i \leq N} \frac{1}{X_t^{i,N} - X_t^{j,N}} dt - \frac{1}{2} V'(X_t^{i,N}) dt, \ 1 \leq i \leq N, \\ X_{t=0}^{i,N} = X_0^{i,N}. \end{cases} \tag{1.1.1}$$

$(X_t^{i,N})_{t>0, 1 \leq i \leq N}$ are the positions of the particles living in $\mathbb{R}$ and starting initially at the i.i.d random variables $(X_0^{i,N})_{1 \leq i \leq N}$, $(B_t^i)_{t \geq 0, i \geq 1}$ is a collection of i.i.d standard Brownian motions, $\beta \geq 1$ is the Dyson index and V is an external confining potential, competing with the repulsive force $W(x-y) := -\log|x-y|$. When $\beta = 1, 2, 4$ and $V = 0$, Dyson proved in [32] that the ordered eigenvalues $(X_t^{1,N} \leq X_t^{2,N} \leq \cdots \leq X_t^{N,N})_{t \geq 0}$ of $N \times N$ Hermitian random matrices with real, complex or quaternion Brownian entries, solve SDE (1.1.1). The present unidimensional model is therefore closely related to the random matrix theory and the study of the Gaussian Orthogonal Ensemble, Gaussian Unitary Ensemble and Gaussian Symplectic Ensemble. The extension of the model to higher dimensions, called the Coulomb gas model, is fundamental in understanding the physical phenomena of vortices in the Ginzburg-Landau model [102].

The existence of particles $(X_t^{i,N})_{t\geq 0, 1\leq i\leq N}$ is a priori not obvious due to the singularity of the interacting Coulomb force W. In a recent paper [82, Theorem 1.1], Li et al. proved that if $V \in C^1(\mathbb{R})$ and under non-restrictive assumptions on V', the collision time $T := \inf\{t \geq 0 : \ \exists\, i \neq j,\ X_t^{i,N} = X_t^{j,N}\}$ is almost-surely infinite for initial conditions $X_0^{1,N} < X_0^{2,N} < \cdots < X_0^{N,N}$ and that the stochastic differential equation (SDE) (1.1.1) has a unique strong solution defined at all times. They established that the mean-field limit of the empirical measure $(\mu_t^N)_{t\geq 0} = \left(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}\right)_{t\geq 0}$ solves the non-linear and non-local Fokker-Planck partial differential equation (PDE)

$$\partial_t \mu_t(x) = \partial_x \left(\mu_t(x)\left(\frac{1}{2}V'(x) - H\mu_t(x)\right)\right), (t,x) \in [0,T]\times\mathbb{R}, \tag{1.1.2}$$

with initial data $\mu_0(dx) = \mathbb{P}\left(X^{1,\infty} \in dx\right)$ and where H is the Hilbert transform. In the case of quadratic potential $V(x) = \frac{x^2}{2}$, this result was proved in [98].

Under mild assumptions on the growth of V (see [12] and [101]), this equation can be seen as the gradient flow of the entropy

$$\Sigma(\mu) := \frac{1}{2}\int_{\mathbb{R}} V(x)\mu(dx) - \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} \log|x-y|\mu(dx)\mu(dy), \tag{1.1.3}$$

for which there exists a unique minimizer μ_V .

A natural question is to study the convergence of $(\mu_t)_{t\geq 0}$ towards the equilibrium measure μ_V. Under strong assumptions on the convexity of V - which are natural in the context of a gradient flow - Li et al. [82, Theorem 1.8] established the exponential convergence of $(\mu_t)_{t\geq 0}$ towards μ_V, with respect to the Wasserstein topology, whose metric is denoted W_2. A similar result, obtained under weaker convexity assumptions, was established in [80, Theorem 2].

Both results rely on the Bakry-Émery strategy, which is a powerful tool to quantify the rate of convergence towards equilibrium by deriving inequalities between the functionals

Σ, $W_2(\cdot, \mu_V)$ and the entropy dissipation, defined by

$$D(\mu) := \int_{\mathbb{R}} \left| \frac{1}{2} V'(x) - H\mu(x) \right|^2 \mu(dx). \tag{1.1.4}$$

In the simpler context of the well-understood overdamped Langevin dynamics, if we assume $\int \exp(-V(x))dx = 1$, then the Gibbs state $\rho_V(dx) := \exp(-V(x))dx$ can be simulated by running for long times the process

$$dX_t = -\frac{1}{2}V'(X_t)dt + dB_t. \tag{1.1.5}$$

This equation can be interpreted as the gradient flow of the relative entropy H with respect to ρ_V

$$H(\mu) := \begin{cases} \frac{1}{2}\int \log \frac{\mu(dx)}{\rho_V(dx)}\mu(dx) \text{ if } \mu \text{ is absolutely continuous with respect to } \rho_V\ , \\ +\infty \text{ otherwise,} \end{cases} \tag{1.1.6}$$

whose minimizer is the Gibbs measure. Here $\frac{\mu(dx)}{\rho_V(dx)}$ denotes the Radon-Nikodym derivative of μ with respect to ρ_V. Under the assumption that $V(x) - \eta x^2$ is convex for some $\eta > 0$, the exponential stability towards ρ_V of the laws $(\rho_t)_{t\geq 0}$ of the solution to SDE (1.1.5) has been established in [7]. This was achieved by proving the logarithmic Sobolev inequality

$$H(\rho_t) \leq c_1 I(\rho_t),\ t \geq 0, \tag{1.1.7}$$

where I is the Fisher information $I(\rho_t) := -\frac{d}{dt}H(\rho_t) = \frac{1}{4}\int |\partial_x \rho_t(x)|^2 \rho_t(dx)$, and the transportation inequality

$$W_2(\rho_t, \rho_V)^2 \leq c_2(H(\rho_t) - H(\rho_V)),\ t \geq 0. \tag{1.1.8}$$

In (1.1.7) and (1.1.8), the constants $c_1, c_2 > 0$ only depend on η.

For the log gases (1.1.2), corresponding logarithmic Sobolev and transportation inequalities were established in [80] and [82], taking into account the technical difficulty of the nonlocal nonlinearity of the interacting potential W. Both results require the convexity of potential V and fail to capture the important case of the double-well potential

$$V_{g,c}(x) := g\frac{x^4}{4} + c\frac{x^2}{2},\ x \in \mathbb{R},\ g > 0,\ c < 0. \tag{1.1.9}$$

Near the origin, $V''_{g,c} < 0$ and the convexity assumptions of [82] and [80] are not satisfied.

The main result of Chapter I is the derivation of a new version of the so-called Entropy-Wasserstein-Information inequality (HWI) that allows establishing exponential stability for the double-well potential $V_{g,c}$ when $-c > 0$ is small enough. It answers partially Conjectures 7.2 and 7.3 [82] that predict exponential stability towards μ_V for all $c \in [-2, 0)$. We also answer positively to the Conjecture 7.3 [12], regarding to the stability for the non-confining potential V (1.1.9) when $g < 0$ and $c > 0$.

1.2 Well-posedness of the supercooled Stefan problem

In the second chapter, we study the well-posedness of the mean-field limit of the system of $N \geq 1$ particles

$$\begin{cases} X_t^{i,N} = X_{0^-}^i + B_t^i - \Lambda_t^N,\ t \geq 0, \\ \tau_{i,N} = \inf\{t \geq 0 : X_t^{i,N} \leq 0\}, \\ \Lambda_t^N = \frac{1}{N}\sum_{i=1}^N 1_{\tau_{i,N} \leq t}, \end{cases} \tag{1.2.1}$$

where $(X_{0^-}^i)_{1\leq i\leq N}$ are i.i.d. real positive random variables and $(B_t^i)_{t\geq 0, 1\leq i\leq N}$ are i.i.d Brownian motions. This model appears, under various forms, in the modeling of interbank risk ([91], [92]), the study of probabilistic growth models [29], and models from computational neuroscience ([26],[25], [100]).

The mean-field limit of (1.2.1) is the McKean-Vlasov SDE

$$\begin{cases} X_t = X_{0^-} + B_t - \Lambda_t, \quad t \geq 0, \\ \tau := \inf\{t \geq 0 : \, X_t \leq 0\}, \\ \Lambda_t = \mathbb{P}\left(\tau \leq t\right), \quad t \geq 0. \end{cases} \tag{1.2.2}$$

A singular aspect of SDE (1.2.2) is the necessity of jump discontinuities in the frontier Λ under generic initial conditions. The solutions minimizing the jump sizes are the so-called physical solutions and satisfy the physicality condition

$$\Lambda_t - \Lambda_{t-} = \inf\left\{x > 0 : \, \mathbb{P}\left(\tau \geq t, \, X_{t-} \in (0, x]\right) < x\right\}, \quad t \geq 0. \tag{1.2.3}$$

After explosion times, the front Λ is typically only $1/2$-Hölder, and therefore the difficulty of analyzing SDE (1.2.2) is dual. Aside of the inherent difficulty of fixed-point problems, due to the McKean-Vlasov nature of the equation, understanding the dynamics of $(X_t)_{t\geq 0}$ and $(\Lambda_t)_{t\geq 0}$ amounts to understanding the hitting time of a highly singular moving boundary by a positive Brownian motion.

Besides the aforementioned applications of such model, SDE (1.2.2) was demonstrated to be a powerful tool to reformulate the supercooled Stefan problem (see [28],[79]). This classical problem from mathematical physics formalizes the dynamic of the freezing of supercooled liquids, and is described by the free-boundary problem

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\partial_{xx} u(t,x), \quad x > \Lambda_t, \; t > 0, \\ u(0,x) = f(x), \; x \geq 0 \quad \text{and} \quad u(t,\Lambda_t) = 0, \; t > 0, \\ \dot{\Lambda}_t = \frac{1}{2}\partial_x u(t, x+)|_{x=\Lambda_t}, \; t \geq 0, \\ \Lambda_0 = 0, \end{cases} \tag{1.2.4}$$

where u is the negative of the temperature of the liquid relative to its equilibrium freezing point, f is the initial profile of temperature and the free boundary Λ describes the location of a liquid-solid frontier over time. Formally, one can check that if X_{0^-} has density f and $(X_t)_{t\geq 0}$ is a solution of SDE (1.2.2), then the law $u(t,x)dx$ of $(X_t+\Lambda_t)1_{\tau\geq t}$ gives rise to a solution of PDE (1.2.4). While the existence of a solution to (1.2.4) has been established only locally (see [69], [104]) - until the explosion time $T \geq 0$ such that $\lim_{t\to T, t<T} \dot{\Lambda}_t = +\infty$ - reformulation (1.2.2) allows the definition of global solutions.

The present boundary problem describes the supercooled Stefan problem on the real line and one phase. A similar formulation to probabilistic formulation (1.2.2) can be written for higher dimensions and multiple phases (see [44],[5]).

If the initial condition X_{0^-} has a density f that changes monotonicity finitely many times over compact intervals, Delarue, Nadtochiy and Shkolnikov proved in [28], the uniqueness of a strong physical solution for SDE (1.2.2) and provided a full analysis of the regularity of the front $(\Lambda_t)_{t\geq 0}$.

The main result of Chapter II is the proof of uniqueness of physical solutions for oscillatory initial densities that violate the monotonicity condition. We use a new contraction argument that replaces the local monotonicity condition of [28] by an averaging condition. We prove that this condition is satisfied by a fairly general class of oscillatory initial densities, which can be described as the almost sure trajectories of stochastic processes. The motivation of this work is the existence of natural oscillatory initial densities arising in continuum limits of interacting particle systems (see [29], [74]).

1.3 Well-posedness of a local stochastic volatility model

In the third chapter, we investigate the well-posedness of the local stochastic volatility (LSV) model

$$\frac{dS_t}{S_t} = \sqrt{\frac{f(Y_t)}{\mathbb{E}\left[f(Y_t)|S_t\right]}}\sigma_{\text{loc}}(t,S_t)dB_t,\ t\geq 0, \tag{1.3.1}$$

in the context of calibration to European call options $(C(t,K))_{t>0,K>0}$. Here $(S_t)_{t\geq 0}$ designs the price process, $(B_t)_{t\geq 0}$ is a Brownian motion, $(Y_t)_{t\geq 0}$ is any adapted Lévy-Itô process, potentially correlated to $(B_t)_{t\geq 0}$, $f : \mathbb{R} \to \mathbb{R}_+$ is a positive smooth function and σ_{loc} is the Dupire's volatility given by

$$\sigma_{\text{loc}}(t,K) := \sqrt{\frac{2\partial_t C(t,K)}{\partial_K^2 C(t,K)}}. \tag{1.3.2}$$

The main motivation behind the particular structure of SDE (1.3.1) is that according to Gÿongy's theorem [57], for a solution $(S_t)_{t\geq 0}$, the fixed-time marginal distributions of $(S_t)_{t\geq 0}$ are independent of f and $(Y_t)_{t\geq 0}$, and are given by the marginal distributions of the local volatility model $dS_t^{\text{loc}} = \sigma_{\text{loc}}(t, S_t^{\text{loc}})S_t^{\text{loc}} dB_t$ with initial data S_0.

This property explains the interest of the quant community in considering SDE (1.3.1) as a powerful class of models to solve calibration problems. Indeed, model (1.3.1) is already calibrated to implied volatility surfaces, and one may use the degree of freedom given by the process $(f(Y_t))_{t\geq 0}$ to jointly calibrate instruments depending on the whole process $(S_t)_{t\geq 0}$ to market data, consistently with the vanilla option market. For example, model (1.3.1) may be - at first sight - a good candidate for jointly calibrating SPX options and VIX options, or generating forward skew coherently with the market implied volatility surface.

Among other methods (see [53]), practitioners solve numerically SDE (1.3.1) by approximating it with the system of $M \geq 1$ interacting particles

$$\frac{dS_t^{i,M}}{S_t^{i,M}} = \sqrt{\frac{f(Y_t^i)\frac{1}{M}\sum_{j=1}^M W_{\delta_N}(S_t^{i,M} - S_t^{j,M})}{\frac{1}{M}\sum_{j=1}^M f(Y_t^j)W_{\delta_M}(S_t^{i,M} - S^{j,M})}}\sigma(t, S_t^{i,M})dB_t^i, \tag{1.3.3}$$

where the particle $(S^{i,M}, Y^{i,M})_{t\geq 0}$, $1 \leq i \leq M$, starts initially at $(S_0^{i,M}, Y_0^{i,M})$ and $(W_{\delta_M})_{M\geq 1} := \left(\frac{1}{\delta_M}W\left(\frac{\cdot}{\delta_M}\right)\right)_{M\geq 1}$ is a family of smoothing kernels, converging - in the sense of distributions - towards the Dirac mass δ_0, when M goes to infinity.

However, on the mathematical level, the well-posedness of SDE (1.3.1) and the propagation of chaos of system (1.3.3) is still an open problem and only partial results were obtained.

Let $X := \log S$ and let's assume that $(Y_t)_{t\geq 0}$ is a diffusion satisfying

$$dY_t = b(t, X_t, Y_t)dt + \sigma_Y(t, X_t, Y_t)dB'_t \tag{1.3.4}$$

where $(B'_t)_{t\geq 0}$ is another standard Brownian motion with correlation $\rho \in [-1, 1]$ with $(B_t)_{t\geq 0}$ and $b : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and $\sigma_Y : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ are the drift and volatility coefficients of Y.

The non-linear and non-local Fokker-Planck PDE solved by the joint density of the process $(X_t, Y_t)_{t\geq 0}$ can be written as

$$\begin{cases} \partial_t p(t,x,y) - L_p p(t,x,y) = 0, \ (t,x,y) \in \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}, \\ p(0,x,y) = P(x,y), \ (x,y) \in \mathbb{R} \times \mathbb{R}, \end{cases} \tag{1.3.5}$$

where L_p is the non-linear elliptic operator of the second order

$$\begin{aligned} L_p u := &\frac{1}{2}\partial_{xx}\left[\frac{f\int p(\cdot,\cdot,z)dz}{\int f(z)p(\cdot,\cdot,z)dz}u\right] + \frac{1}{2}\partial_{yy}\left[\sigma_Y^2 u\right] + \rho\partial_{xy}\left[\frac{f\int p(\cdot,\cdot,z)dz}{\int f(z)p(\cdot,\cdot,z)dz}\sigma_Y\right] \\ &+ \frac{1}{2}\partial_x\left[\frac{f\int p(\cdot,\cdot,z)dz}{\int f(z)p(\cdot,\cdot,z)dz}u\right] - \partial_y\left[bu\right], \ u \in C^{0,2}(\mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}). \end{aligned} \tag{1.3.6}$$

Jourdain and Zhou proved in [67] the existence of a weak solution to SDE (1.3.1) when $(Y_t)_{t\geq 0}$ is a continuous-time Markov chain taking finitely many values and the range of f is small enough. This last condition echoes the result of Abergel and Tachet[2], who proved by a perturbation argument the existence of a strong solution to PDE (1.3.5), under the condition that $\sup_{y\in\mathbb{R}} |f''(y)|$ is small enough. In [53, Section 11.3], Guyon and Henry-Labordère witnessed in their numerical experiments, the divergence of various approximation algorithms of SDE (1.3.1) if the range of f is too large, hinting the non-existence of solutions or numerical instability of the algorithms.

The main objective of the third chapter is to establish the existence of a strong solution to a regularized version of SDE (1.3.1):

$$\begin{cases} dX_t = -\frac{1}{2}\frac{\varepsilon+f(Y)p(t,X_t)}{\varepsilon+\mathbb{E}[f(Y)|X_t]p(t,X_t)}\sigma(t,X_t)^2dt + \sqrt{\frac{\varepsilon+f(Y)p(t,X_t)}{\varepsilon+\mathbb{E}[f(Y)|X_t]p(t,X_t)}}\sigma(t,X_t)dB_t, \\ \mathbb{P}\left(X_t \in dx\right) = p(t,x)dx, \\ X_{t=0} = X_0, \end{cases} \tag{1.3.7}$$

where Y is a discrete random variable taking finitely many values and $\varepsilon > 0$ is a fixed regularization parameter. The main benefit of this regularization is that the calibration property is conserved - fixed-time marginals are still given by the local volatility model. Under the natural condition that the range of f is small enough, we prove the existence of a unique strong solution to SDE problem (1.3.7). In addition, we prove the propagation of chaos for the associated particle system. While weak existence has been established in [67], well-posedness of (1.3.7) and the propagation of chaos result are out of reach due to the lack of regularity of the solution.

1.4 Joint SPX/VIX calibration

In the last chapter, we leave the field of McKean-Vlasov equations and consider a more practical, yet difficult, problem arising in the topic of calibration and address the joint calibration of S&P 500 and VIX.

The VIX being defined as the forward realized variance of the SPX over one month and the market of VIX options - used to hedge against market uncertainty - being liquid, one may be naturally interested in finding a model calibrating jointly both instruments. The use of joint calibration model rules out the possibility of arbitrage between actors trading SPX and VIX options. Moreover, such models are necessary to give the fair price of payoffs involving simultaneously SPX and VIX options. Nevertheless, since the opening of VIX trades in 2006, this problem has been puzzling researchers and practitioners.

In the context of one-factor Markovian LSV, the problem can be stated as follows.

Denote $(X_t)_{t\geq 0}$ the log-price of the SPX and $(Y_t)_{t\geq 0}$ another process driving the volatility of $(X_t)_{t\geq 0}$, and consider the controlled SDE

$$\begin{cases} dX_t = -\frac{1}{2}\sigma_X^\alpha(t, X_t, Y_t)^2 dt + \sigma_X^\alpha(t, X_t, Y_t) dB_t^1, \\ dY_t = \mu_Y^\alpha(t, X_t, Y_t) dt + \sigma_Y^\alpha(t, X_t, Y_t)\big(\rho^\alpha(t, X_t, Y_t) dB_t^1 \\ \qquad\qquad\qquad\qquad\qquad + \sqrt{1-\rho^\alpha(t, X_t, Y_t)^2} dB_t^2\big), \end{cases} \tag{1.4.1}$$

where $(B_t^1)_{t\geq 0}$ and $(B_t^2)_{t\geq 0}$ are two independent Brownian motions, $\alpha \in \mathbb{R}^d$, $d \geq 1$ is the control parameter, σ_X^α is the volatility of the log-price, μ_Y^α is the drift of Y and σ_Y^α the volatility of Y, and ρ^α is the correlation between the two Brownian motions driving the dynamics of X and Y. In this Markov model, with classical notations, the VIX^2 at $t \geq 0$ is given by

$$\mathrm{VIX}_t^2 := \frac{1}{\tau}\int_t^{t+\tau} \mathbb{E}\left[\sigma_X^\alpha(s, X_s, Y_s)^2 | X_t, Y_t\right] ds, \tag{1.4.2}$$

where $\tau = \frac{30}{365}$ (30 days). VIX call options are written as $\mathbb{E}\left[\left(\sqrt{\mathrm{VIX}_t^2} - K\right)_+\right]$, $t, K \geq 0$.

Model (1.4.1) is said to be jointly calibrated to the market if the SPX and VIX smiles, and SPX and VIX futures, computed in the model, match closely given market data at multiple maturities.

The first attempts to solve the problem were using traditional parametric continuous time models (Heston[36] , Bergomi [63], CEV [38]) but failed to mediate the large negative skew of short-term SPX options and the low level of VIX implied volatilities. More general models incorporating jumps were therefore considered [73],[96], [8], [92],[97], but the calibration results were not satisfactory.

In more recent years, different approaches were used to solve the problem. Guyon was the first to calibrate exactly to SPX and VIX smiles by formulating the calibration as a martingale optimal transport problem [46]. The discrete time model of [46] was later extended to continuous time [47] and an efficient calibration method was derived in [51]. A similar but more general approach was used by Guo et al. in [42].

Parametric continuous stochastic volatility models were shown to calibrate jointly

SPX and VIX options. In [62], the authors calibrate a "quintic Ornstein-Uhlenbeck volatility model", with few parameters. In [22], signature based models were used to solve the problem.

Our solution to the problem of joint calibration is part of the continuous modeling approach. We used neural SDEs, which are natural extensions of neural ODEs to the context of modeling of probabilistic physical dynamics. Neural ODEs - introduced in [19] - are ODEs of the form $y' = F(t, y)$ where F is a neural network, and can be seen as the continuous limit of specific neural networks (among which residuals networks, recurrent neural network) whose inner state transformation operations can be written as a step of the Euler method. Neural SDEs are SDEs whose drifts and volatilities are modeled as neural networks.

The key benefits of such models are the expressive power of neural networks, and the efficiency of the available training algorithms [84], [111], [108], [70].

In Chapter IV, the family of models (1.4.1), where $\sigma_X^\alpha, \sigma_Y^\alpha, \mu_Y^\alpha$ and ρ^α are neural networks and with weights controlled by α, is shown to fit jointly market data for maturities spanning over 8 months (and including short maturities). Stochastic gradient descent and backpropagation are used to solve the calibration problem, written as the minimization of a suitable loss.

We observe numerically that the joint calibration model is actually a pure path-dependent volatility model confirming the findings in [45].

Chapter 2

Trend to equilibrium for granular media equations under non-convex potentials and application to log gases

The present chapter studies the extension of stability results proved for the granular media equation

$$\partial_t \mu = \nabla \cdot \left[\mu \nabla \left(\frac{1}{2} V + W * \mu \right) \right] \tag{2.0.1}$$

to non-convex potentials. The unknown μ is a time-dependent probability measure on $\mathbb{R}^d$, $V : \mathbb{R}^d \to \mathbb{R}$ is an external potential and $W : \mathbb{R}^d \to \mathbb{R}$ is an interaction potential. The non-local and non-linear partial differential equation (2.0.1) is the formal gradient flow of the entropy

$$\Sigma(\mu) = \frac{1}{2} \int_{\mathbb{R}^d} V(x) \mu(dx) + \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) \mu(dx) \mu(dy), \tag{2.0.2}$$

whose dissipation is defined as

$$D(\mu) \equiv \int_{\mathbb{R}^d} \left| \nabla \left(\frac{1}{2} V + W * \mu \right) \right|^2 \mu(dx). \tag{2.0.3}$$

Under the two sets of assumptions $D^2V \geq 2\lambda$ and $D^2W \geq 0$ or $D^2V \geq 0$ and $D^2W \geq \lambda$, for some $\lambda > 0$, Carrillo et al. established in [18] the celebrated HWI inequality

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)} W_2(\rho_0, \rho_1) - \frac{\lambda}{2} W_2(\rho_0, \rho_1)^2, \tag{2.0.4}$$

W_2 being the Wasserstein distance and ρ_0, ρ_1 having finite entropy and belongs to $\mathcal{M}_2$, the set of probability measures with finite second-order moment. This inequality implies successively a transportation inequality, a log-Sobolev inequality and exponential stability with respect to the Wasserstein distance towards a minimizer μ_∞ of the entropy

$$W_2(\mu_t, \mu_\infty) \leq \sqrt{\frac{2\left(\Sigma(\mu_0) - \Sigma(\mu_\infty)\right)}{\lambda}} e^{-\lambda t}, \ \forall t \geq 0. \tag{2.0.5}$$

The primary aim of this paper is to extend this method to non-strictly convex potentials. More precisely, we allow V to be non-convex near the origin, according to the following assumptions.

Assumptions 2.0.1. *There exist $\alpha, \beta, \gamma, r > 0$ such that the C^2 potentials V, W satisfy*

(A1) $D^2V(x) \geq -\beta$ for $x \in \mathbb{R}^d$,

(A2) $D^2V(x) \geq \alpha$ for $|x| \geq r$,

(A3) V is symmetric: $V(x) = V(-x)$ for $x \in \mathbb{R}^d$,

(A4) W is convex and $D^2W(x) \geq \gamma$ for $|x| \leq 2r$,

(A5) W is symmetric: $W(x) = W(-x)$ for $x \in \mathbb{R}^d$.

Under those assumptions, we prove the following

Theorem 2.0.2. *Assume that Assumptions 2.0.1 are satisfied. Let $\rho_0, \rho_1 \in \mathcal{M}_2$ with finite entropy. Define*

$$P_r = \max\left(\int_{|x|>r} \rho_0(dx), \int_{|x|>r} \rho_1(dx)\right). \tag{2.0.6}$$

If ρ_0 and ρ_1 have the same center of mass, then

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)}W_2(\rho_0, \rho_1) - \frac{\lambda}{2}W_2(\rho_0, \rho_1)^2, \tag{2.0.7}$$

where the constant λ is given by

$$\lambda = \frac{\min(\alpha, 2\gamma - \beta)}{2} - 2\gamma P_r. \tag{2.0.8}$$

If ρ_0 and ρ_1 are symmetric, then (2.0.7) holds with the better constant λ:

$$\lambda = \frac{\min\big(\alpha, 2\gamma(1 - 2P_r) - \beta\big)}{2}. \tag{2.0.9}$$

Remark 2.0.3.

(i) The method of establishing HWI inequalities is not the only one to derive stability rates. We can cite the strategy of Bakry-Émery [6], which consists of computing the dissipation of the entropy dissipation or the method of characteristics or even ad-hoc computations similar to what is done in [81]. In any case, the computations turn out to be very similar.

(ii) HWI inequality (2.0.7) is slightly different from usual HWI inequalities, because λ depends on ρ_0 and ρ_1 through their tail probabilities. Therefore, application of our modified HWI inequality requires beforehand uniform bound of those tail probabilities, by uniformly bounding the moments for example.

(iii) Our proof relies heavily on exploiting the convexity of W, which requires the as-

sumptions of a fixed center of mass or a symmetric initial data (see Theorem 2.2 [18] for example or [13] and [14]).

(iv) Consider the internal energy $\mathcal{U}$

$$\mathcal{U}(\mu) = \int_{\mathbb{R}^d} U(\mu(x))\mu(dx), \tag{2.0.10}$$

where $U : \mathbb{R}_+^* \to \mathbb{R}$ *satisfies the dilation condition that* $\lambda \in \mathbb{R}_+^* \to \lambda^d U(\lambda^{-d})$ *is convex and non-increasing. The conclusions of Theorem 2.0.2 hold with the same constants for the entropy*

$$\mu \mapsto \mathcal{U}(\mu) + \Sigma(\mu). \tag{2.0.11}$$

Typically for $U(\rho) = \rho \log \rho$*, our approach can be used to prove exponential stability of McKean-Vlasov diffusions under non-convex external potentials (see [107]).*

In the same fashion of Theorem 2.3 [18], we investigate the case of V convex (but not strictly convex) and W degenerately convex at infinity.

Assumptions 2.0.4. *V and W belong to $C^2(\mathbb{R}^d)$ and $C^2(\mathbb{R}^d - \{0\})$ respectively and there exist positive constants c and η such that*

(B1) $D^2V(x) \geq 0$ *for* $x \in \mathbb{R}^d$*,*

(B2) $D^2W(x) \geq \frac{c}{|x|^\eta}$ *for* $x \in \mathbb{R}^d - \{0\}$*.*

Under those assumptions and additional technical Assumptions 2.3.1, postponed to Section 2.3, we prove the following

Theorem 2.0.5. *Assume that Assumptions 2.0.4 and 2.3.1 are satisfied. Let* $\rho_0, \rho_1 \in \mathcal{M}_2$ *with finite fourth moments and finite entropy. Define*

$$m = \max\left(\int_{\mathbb{R}^d} |x|^4 \rho_0(dx), \int_{\mathbb{R}^d} |x|^4 \rho_1(dx)\right). \tag{2.0.12}$$

Then, there exists a constant $C > 0$ depending only on m, c and η such that the following HWI inequality holds

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)} W_2(\rho_0, \rho_1) - C W_2(\rho_0, \rho_1)^{\eta+2}. \tag{2.0.13}$$

This variant of the HWI inequality (2.0.7) implies algebraic stability, meaning

$$W_2(\mu_t, \mu_\infty) \leq \frac{C}{t^{1/\eta}}, \ \forall t \geq 1. \tag{2.0.14}$$

The second contribution of this paper is the application of the previous results to log gases. Those gases are obtained by taking $W = -\log|\cdot|$ in the unidimensional case $d = 1$. It leads to the Fokker-Planck equation

$$\partial_t \mu_t = \partial_x \left[\mu_t \left(\frac{1}{2} V' - H\mu_t \right) \right], \tag{2.0.15}$$

where H denotes the Hilbert transform. The logarithmic potential has a singularity at the origin and is only convex on the half-lines $\mathbb{R}_+^*$ and $\mathbb{R}_-^*$. Because of these difficulties, the previous theorems do not apply directly and the derivation of the analogue of (2.0.7) is more involved. This is the purpose of Theorem 2.0.6, which extends the application of HWI inequalities to log gases.

Theorem 2.0.6. *Let V be a C^2 symmetric potential satisfying Assumptions (A1)- (A3) and let $W = -\log|\cdot|$. Then the conclusions of Theorem 2.0.2 hold for probability measures with density in $L^\infty(\mathbb{R})$, with $\gamma = \frac{1}{4r^2}$.*

This result and its proof have two main applications. The first concerns log gases under a general convex potential V. Under appropriate growth assumption of V, ensuring the existence and uniqueness of the minimizer of the entropy μ_V, we prove that any solution converges towards μ_V at a square root rate. The definition of a solution to (2.0.15) is given afterwards in Section 2.4.

Theorem 2.0.7. *Let V be a C^2 convex potential satisfying growth Assumptions 2.4.1. Let μ_V be the unique minimizer of the entropy and $(\mu_t)_{t\geq 0}$ be a solution of (2.0.15). Then we have algebraic stability towards μ_V*

$$W_2(\mu_t, \mu_V) \leq \frac{C}{\sqrt{t}}, \ \forall t \geq 1, \tag{2.0.16}$$

for some constant $C > 0$.

To our knowledge, in the case of log gases, the weakest assumptions required for an explicit equilibrium were strict convexity at least away of the origin (see [81]). Our result weakens this assumption, the cost being a slower rate.

The second application of Theorem 2.0.6 concerns quartic potentials V, defined by

Assumptions 2.0.8. *For some constants $g, c \in (-\infty, 0)$, $(C1)$ or $(C2)$ is satisfied*

(C1) $V(x) = \frac{x^4}{4} + c\frac{x^2}{2}$ *for* $x \in \mathbb{R}$,

(C2) $V(x) = g\frac{x^4}{4} + \frac{x^2}{2}$ *for* $x \in \mathbb{R}$.

For $c < 0$, quartic potential (C1) is non-convex and does not fall under the scope of the work of Ledoux and Popescu [81]. After deriving moment estimates (which already imply stability and provide a simple proof to Theorem 1.1 [30]), we apply Theorem 2.0.6 to derive an exponential stability rate towards the unique minimizer μ_V of the entropy, for solutions with a fixed center of mass or with a symmetric initial data.

Theorem 2.0.9. *Let $c \in \left(-\frac{1}{4\sqrt{17}}, 0\right)$ and V be defined by $(C1)$. Let $(\mu_t)_{t\geq 0}$ be a solution of (2.0.15) with a fixed center of mass and finite fourth moments. Assume that the moment condition*

$$\int_{\mathbb{R}} x^2 \mu_0(dx) \leq \frac{-c + \sqrt{c^2+4}}{2}, \tag{2.0.17}$$

is satisfied. Then $(\mu_t)_{t\geq 0}$ is exponentially stable towards μ_V with respect to the Wasserstein distance

$$W_2(\mu_t, \mu_V) \leq \sqrt{\frac{2(\Sigma(\mu_0) - \Sigma(\mu_V))}{\lambda}} e^{-\lambda t}, \ \forall t \geq 0. \tag{2.0.18}$$

The rate λ *is given by*

$$\lambda = \frac{1}{16\left(-c+\sqrt{c^2+4}\right)} + \frac{c}{2} > 0. \tag{2.0.19}$$

Similarly, if $c \in \left(-\frac{1}{\sqrt{6}}, 0\right)$ *and if* $(\mu_t)_{t\geq 0}$ *is a solution of (2.0.15) with finite fourth moments and a symmetric initial data* μ_0*, then under the same moment condition,* $(\mu_t)_{t\geq 0}$ *is exponentially stable towards* μ_V *with respect to the Wasserstein distance with rate* λ *given by*

$$\lambda = \frac{1}{2\left(-c+\sqrt{c^2+4}\right)} + \frac{c}{2} > 0. \tag{2.0.20}$$

The assumption of a fixed center of mass or a symmetric initial data are required to exploit easily the strict convexity of W. It leads to tractable computations and exact numerical values. We are able to relax those assumptions in the following theorem, and assume only lower-bounded second moments for the initial data. We prove exponential stability under weaker assumptions but only for $c \in (c^*, 0)$ with $c^* \sim 10^{-9}$.

Theorem 2.0.10. *There exists* $c^* < 0$*, such that if* $c \in (-c^*, 0)$ *and if* $(\mu_t)_{t\geq 0}$ *is a solution of (2.0.15) with finite sixth moments and initial data satisfying*

$$\left(\frac{2}{-c+\sqrt{c^2+16}}\right)^4 \leq \int x^2 \mu_0(dx) \tag{2.0.21}$$

and

$$\int x^4 \mu_0(dx) \leq \left(\frac{-c+\sqrt{c^2+12}}{2}\right)^2, \tag{2.0.22}$$

then $(\mu_t)_{t\geq 0}$ *is exponentially stable towards the equilibrium* μ_V*.*

For $g < 0$, a new difficulty arises due to the fact that the quartic potential (C2) is neither convex nor confining. Nevertheless, we manage to establish exponential stability for $g \in \left(-\frac{1}{81+36\sqrt{5}}, 0\right)$ and for well-defined solutions. This result is stated in the following theorem and answers a conjecture formulated in [12] (see Conjecture 7.3 [83] as well).

Theorem 2.0.11. *Let* $g \in \left(-\frac{1}{81+36\sqrt{5}}, 0\right)$ *and* $m > 0$ *satisfying*

$$m < \sqrt{-\frac{1}{3g} - \frac{4}{\sqrt{-g}} - 3}. \tag{2.0.23}$$

Let μ_0 *be an initial measure with supp*$(\mu_0) \subset [-m, m]$. *Then any solution* $(\mu_t)_{t\geq 0}$ *of (2.0.15) with quartic* V *(C2) is well-defined and converges exponentially towards a stationary measure* μ_∞

$$W_2(\mu_t, \mu_\infty) \leq \sqrt{\frac{2\left(\Sigma(\mu_0) - \Sigma(\mu_\infty)\right)}{\lambda}} e^{-\lambda t}, \ \forall t \geq 0, \tag{2.0.24}$$

with rate λ *given by*

$$\lambda = \frac{1}{2}\left[1 + 3g\left(m^2 + \frac{4}{\sqrt{-g}} + 3\right)\right] > 0. \tag{2.0.25}$$

Moreover, μ_∞ *is a local minimizer of the entropy*

$$\text{supp}(\mu) \subset \left(-\sqrt{m^2 + \frac{4}{\sqrt{-g}} + 3}, \sqrt{m^2 + \frac{4}{\sqrt{-g}} + 3}\right) \implies \Sigma(\mu_\infty) \leq \Sigma(\mu), \tag{2.0.26}$$

for all $\mu \in \mathcal{M}_2$.

The chapter is structured as follows. In Section 2.1, we recall some preliminary facts collected from [109] and [18], and we introduce notation. Section 2.2 contains the proof of Theorem 2.0.2 and its corollary concerning stability of solutions. In Section 2.3, we establish the proof of Theorem 2.0.5 and its implications. In Section 2.4, we consider log gases and prove Theorems 2.0.6 - 2.0.11. The appendix gathers auxiliary proofs.

This chapter is based on [89].

2.1 Preliminaries and notations

In the whole paper, we denote by $\mathcal{M}$ the set of probability measures on $\mathbb{R}^d$, and we define for $p \geq 1$

$$\mathcal{M}_p = \left\{ \mu \in \mathcal{M} : \int_{x \in \mathbb{R}^d} |x|^p \mu(dx) < \infty \right\}. \tag{2.1.1}$$

We recall the definition of the Wasserstein metric W_2 on $\mathcal{M}_2$

$$W_2(\rho_0, \rho_1) = \left[\inf_{\gamma \in \Gamma(\rho_0, \rho_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \gamma(dx, dy) \right]^{1/2},$$

where $\Gamma(\rho_0, \rho_1)$ denotes the set of couplings between ρ_0 and ρ_1 (see [109]). According to Brenier Theorem, if ρ_0 and ρ_1 have a density, there exists a unique optimal transport map $T = \nabla \phi$, gradient of a convex function $\phi : \mathbb{R}^d \to \mathbb{R}$, ρ_0-almost everywhere, such that ρ_1 is the push-forward measure $\rho_1 = T \# \rho_0$ and such that

$$W_2(\rho_0, \rho_1) = \left[\int_{\mathbb{R}^d} |x - T(x)|^2 \rho_0(dx) \right]^{1/2}. \tag{2.1.2}$$

Recall from [86], that a functional $F : \mathcal{M}_2 \to \mathbb{R}$ is said to be displacement convex if $s \mapsto F(\rho_s)$ is a convex function, where $(\rho_s)_{s \in [0,1]}$ is the geodesic in $(\mathcal{M}_2, W_2)$ joining ρ_0 to ρ_1:

$$(\rho_s)_{s \in [0,1]} = ([(1-s)Id + sT] \# \rho_0)_{s \in [0,1]}. \tag{2.1.3}$$

Strict displacement convexity is a stronger property, which plays a crucial role in establishing equilibrium rates. The key to derive a HWI inequality is to prove that Σ is λ-strictly convex along interpolation (2.1.3):

$$\frac{d^2}{ds^2} \Sigma(\rho_s) \geq \lambda W_2(\rho_0, \rho_1)^2, \; 0 < s < 1.$$

We denote $\mathcal{V}$ and $\mathcal{W}$ the functionals

$$\begin{aligned}\mathcal{V} : \mu \in \mathcal{M} \to \frac{1}{2}\int V(x)\mu(dx),\\ \mathcal{W} : \mu \in \mathcal{M} \to \frac{1}{2}\iint W(x-y)\mu(dx)\mu(dy).\end{aligned} \tag{2.1.4}$$

We define $(\mu_t)_{t\geq 0} \in C(\mathbb{R}_+, \mathcal{M})$ as a solution of (2.0.1) with initial data μ_0 if for all $t \geq 0$, μ_t has a bounded density, that we shall denote $\mu_t(x)$, such that $\nabla W * \mu_t \in L^\infty_{loc}(\mathbb{R}_+ \times \mathbb{R}^d)$ and

$$\int \phi d\mu_t - \int \phi d\mu_0 = -\int_0^t ds \int \nabla\phi \cdot \nabla\left(\frac{1}{2}V + W * \mu_s\right) d\mu_s,\ \forall \phi \in C_0^\infty(\mathbb{R}^d), \tag{2.1.5}$$

where $C_0^\infty(\mathbb{R}^d)$ denotes the space of smooth and compactly supported test functions.

Finally, we end this section by recalling that Proposition 2.1 [18] ensures existence of solutions to (2.0.1), under additional technical assumptions concerning the regularity and the growth at infinity of the C^2 potentials V and W, and establishes the dissipation property

$$\frac{d}{dt}\Sigma(\mu_t) \leq -D(\mu_t),\ \forall t \geq 0, \tag{2.1.6}$$

for $(\mu_t)_{t\geq 0}$ a solution. Moreover, it is straightforward to prove that Σ and D are lower semi-continuous for the weak topology. In Section 2.2, we assume that those technical assumptions are satisfied. However, potentials W, as in Theorem 2.0.5, are not C^2 and Proposition 2.1 [18] cannot be applied. As the purpose of this work is not to prove the dissipation property nor to discuss the existence of solutions, additional assumptions will be therefore assumed to guarantee the validity of Proposition 2.1 [18].

2.2 Proof of Theorem 2.0.2

Proof of Theorem 2.0.2. The main difficulty is to alleviate non-convexity of V near the origin with the strict convexity of W, and conversely to use the strict convexity of V outside a neighborhood of the origin to alleviate non-strict convexity of W near the origin.

Let $\rho_0, \rho_1 \in \mathcal{M}_2$ with finite entropy and the same center of mass. Let T be the optimal transport map from ρ_0 to ρ_1 described by (2.1.2). Consider interpolation (2.1.3) between ρ_0 and ρ_1 given by $(\rho_s)_{s\in[0,1]}$. Set $\theta(x) = T(x) - x$. In these circumstances, we have

$$W_2(\rho_0, \rho_1)^2 = \int_{\mathbb{R}^d} |\theta(x)|^2 \rho_0(dx)$$

and

$$\int_{\mathbb{R}^d} f(x)\rho_s(dx) = \int_{\mathbb{R}^d} f(x + s\theta(x))\rho_0(dx)$$

for all measurable bounded functions f.

The function $t \in [0,1] \to \Sigma(\rho_t)$ is twice differentiable (see [18, Section 4.1]) and Taylor's formula applied between 0 and 1 gives

$$\Sigma(\rho_1) - \Sigma(\rho_0) = \frac{d}{ds}\bigg|_0 \Sigma(\rho_s) + \frac{1}{2}\frac{d^2}{ds^2}\bigg|_{s^*} \Sigma(\rho_s), \tag{2.2.1}$$

for some $s^* \in (0,1)$. Following computations of [18, Section 4.1], we find for all $s \in (0,1)$

$$\frac{d}{ds}\bigg|_0 \Sigma(\rho_s) \geq -\sqrt{D(\rho_0)}W_2(\rho_1, \rho_0) \tag{2.2.2}$$

and

$$\begin{aligned}\frac{d^2}{ds^2}\Sigma(\rho_s) \geq & \underbrace{\frac{1}{2}\int_{\mathbb{R}^d} \left\langle D^2V(x + s\theta(x)) \cdot \theta(x), \theta(x)\right\rangle \rho_0(dx)}_{(2.2.3.1)} \\ & + \underbrace{\frac{1}{2}\int_{\mathbb{R}^{2d}} \left\langle D^2W(x - y + s(\theta(x) - \theta(y)) \cdot (\theta(x) - \theta(y)), \theta(x) - \theta(y)\right\rangle \rho_0(dx)\rho_0(dy)}_{(2.2.3.2)}.\end{aligned} \tag{2.2.3}$$

On the one hand, we have using Assumptions $(A1)$ and $(A2)$

$$(2.2.3.1) \geq \frac{\alpha}{2} \int_{|x+s\theta(x)|\geq r} |\theta(x)|^2 \rho_0(dx) - \frac{\beta}{2} \int_{|x+s\theta(x)|<r} |\theta(x)|^2 \rho_0(dx). \tag{2.2.4}$$

On the other hand, under $(A3)$

$$\begin{aligned}(2.2.3.2) &\geq \frac{1}{2} \int\limits_{\substack{|x+s\theta(x)|\leq r\\ |y+s\theta(y)|\leq r}} \left\langle D^2W(x-y+s(\theta(x)-\theta(y)))\cdot(\theta(x)-\theta(y)), \theta(x)-\theta(y)\right\rangle \rho_0(dx)\rho_0(dx)\\
&\geq \frac{1}{2}\gamma \int\limits_{\substack{|x+s\theta(x)|\leq r\\ |y+s\theta(y)|\leq r}} \left(|\theta(x)|^2 + |\theta(y)|^2 - 2\left\langle \theta(x), \theta(y)\right\rangle\right)\rho_0(dx)\rho_0(dy)\\
&\geq \gamma \int\limits_{|x+s\theta(x)|\leq r} \rho_0(dx) \int\limits_{|x+s\theta(x)|\leq r} |\theta(x)|^2\rho_0(dx) - \gamma \left| \int\limits_{|x+s\theta(x)|\leq r} \theta(x)\rho_0(dx)\right|^2.\end{aligned} \tag{2.2.5}$$

Using the fact that $\int_{\mathbb{R}^d} x\rho_0(dx) = \int_{\mathbb{R}^d} x\rho_1(dx)$, we can write

$$\int\limits_{|x+s\theta(x)|\leq r} \theta(x)\rho_0(dx) = -\int\limits_{|x+s\theta(x)|>r} \theta(x)\rho_0(dx) \tag{2.2.6}$$

and by Cauchy-Schwarz inequality

$$\begin{aligned}\gamma \left| \int\limits_{|x+s\theta(x)|\leq r} \theta(x)\rho_0(dx)\right|^2 &= \gamma \left| \int\limits_{|x+s\theta(x)|>r} \theta(x)\rho_0(dx)\right|^2\\
&\leq \gamma \int\limits_{|x+s\theta(x)|>r} \rho_0(dx) \int\limits_{|x+s\theta(x)|>r} |\theta(x)|^2\rho_0(dx).\end{aligned} \tag{2.2.7}$$

The first integral in the left hand-side is equal to $\int_{|x|>r} \rho_s(dx)$, and therefore combining estimations (2.2.5) and (2.2.7), it follows

$$\begin{aligned}(2.2.3.2) \geq \gamma\Big[\Big(1 - \int\limits_{|x|>r} \rho_s(dx)\Big) \int\limits_{|x+s\theta(x)|\leq r} |\theta(x)|^2 \rho_0(dx) \\ - \int\limits_{|x|>r} \rho_s(dx) \int\limits_{|x+s\theta(x)|>r} |\theta(x)|^2 \rho_0(dx)\Big].\end{aligned} \tag{2.2.8}$$

Noticing that for $s \in [0,1]$

$$\begin{aligned}\int_{|x|>r} \rho_s(dx) &\leq \int_{|x|>r} \rho_0(dx) + \int_{|T(x)|>r} \rho_0(dx) \\ &= \int_{|x|>r} \rho_0(dx) + \int_{|x|>r} \rho_1(dx) \\ &\leq 2P_r,\end{aligned} \tag{2.2.9}$$

we conclude from (2.2.3), (2.2.4), (2.2.8) and (2.2.9), by setting

$$\lambda = \frac{\min(\alpha, 2\gamma - \beta)}{2} - 2\gamma P_r$$

that

$$\frac{d^2}{ds^2}\Big|_{s^*} \Sigma(\rho_s) \geq \lambda \int_{\mathbb{R}^d} |\theta(x)|^2 \rho_0(dx) = \lambda W_2(\rho_1, \rho_0)^2. \tag{2.2.10}$$

Combining (2.2.1), (2.2.2) (2.2.10), yields

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)} W_2(\rho_0, \rho_1) - \frac{\lambda}{2} W_2(\rho_0, \rho_1)^2. \tag{2.2.11}$$

In the special case when ρ_0 and ρ_1 are symmetric, we observe that the quantity (2.2.6) vanishes. Indeed, define

$$A = \{x \in \mathbb{R}^d : |(1-s)x + sT(x)| \leq r\}$$

and

$$c = \int_A (x - T(x))\rho_0(dx) = \int\limits_{|x+s\theta(x)|\le r} \theta(x)\rho_0(dx).$$

From the uniqueness of the optimal transport map T, the symmetry of ρ_0 and ρ_1 and the fact that $W_2(\rho_0, \rho_1)^2 = \int_{\mathbb{R}^d}(x + T(-x))^2\rho_0(dx)$, we see that the map T is odd. Consequently, A is a symmetric domain and $x \mapsto x - T(x)$ is odd. Therefore, $c = 0$ and (2.2.11) holds with

$$\lambda \equiv \frac{\min\left(\alpha, 2\gamma\big(1 - 2P_r\big) - \beta\right)}{2}.$$

□

We derive the following asymptotic behavior and inequalities from (2.0.7).

Corollary 2.2.1. *Assume that Assumptions 2.0.1 are satisfied and that* Σ *is lower-bounded. Let* $(\mu_t)_{t\ge0}$ *be a solution of (2.0.1) with a fixed center of mass. Define*

$$P_r = \sup_{t\ge0} \int_{|x|>r} \mu_t(dx) \tag{2.2.12}$$

and assume that

$$\lambda = \frac{\min\left(\alpha, 2\gamma - \beta\right)}{2} - 2\gamma P_r > 0. \tag{2.2.13}$$

If the family $(\mu_t)_{t\ge0}$ *is tight with respect to the weak topology then* $(\mu_t)_{t\ge0}$ *exponentially converges, with respect to the Wasserstein distance, to the unique minimizer* μ_∞ *of the entropy* Σ *among the class of probability measures* ρ *satisfying*

$$\int_{\mathbb{R}^d} x\rho(dx) = \int_{\mathbb{R}^d} x\mu_0(dx) \quad \text{and} \quad \mathbb{P}_\rho(|x| > r) \le \sup_{t\ge0} \mathbb{P}_{\mu_t}(|x| > r). \tag{2.2.14}$$

Moreover, the following inequalities hold:

(i) Logarithmic Sobolev inequality

$$2\lambda\left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right) \le D(\mu_t), \ \forall t \ge 0. \tag{2.2.15}$$

(ii) Transportation inequality

$$W_2(\mu_t, \mu_\infty) \leq \sqrt{\frac{2\left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right)}{\lambda}}, \ \forall t \geq 0. \tag{2.2.16}$$

(iii) Exponential stability

$$W_2(\mu_t, \mu_\infty) \leq \sqrt{\frac{2\left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right)}{\lambda}} e^{-\lambda t}, \ \forall t \geq 0. \tag{2.2.17}$$

Finally, if μ_0 *is symmetric and*

$$\lambda = \frac{\min\big(\alpha, 2\gamma\big(1 - 2P_r\big) - \beta\big)}{2} > 0, \tag{2.2.18}$$

then the same inequalities hold with this better rate.

Proof. Let μ_∞ be a limit point of $(\mu_t)_{t\geq 0}$ for the weak topology. By lower semi-continuity of D

$$\limsup_{t\to\infty} \frac{d}{dt}\Sigma(\mu_t) \leq -\liminf_{t\to\infty} D(\mu_t) \leq -D(\mu_\infty).$$

If $-D(\mu_\infty) < 0$, then $\limsup_{t\to\infty} \frac{d}{dt}\Sigma(\mu_t) \leq -D(\mu_\infty) < 0$ and there exist t_0 and c such that $\Sigma(\mu_t) \leq -\frac{1}{2}D(\mu_\infty)t + c$ for $t > t_0$. That is $\Sigma(\mu_t) \xrightarrow{t\to\infty} -\infty$. As Σ is bounded below by assumption, $D(\mu_\infty) = 0$ and μ_∞ is a stationary solution.

Moreover, by weak convergence, μ_∞ will satisfy conditions (2.2.14). We can therefore apply Theorem 2.0.2 to $(\rho_0, \rho_1) = (\mu_t, \mu_\infty)$ (notice that in the case of μ_0 symmetric, μ_t stays symmetric at all times for $t > 0$ by symmetry of the potentials). According to the HWI inequality (2.0.7)

$$\Sigma(\mu_t) - \Sigma(\mu_\infty) - \sqrt{D(\mu_t)}W_2(\mu_t, \mu_\infty) + \frac{\lambda}{2}W_2(\mu_t, \mu_\infty)^2 \leq 0, \ \forall t \geq 0.$$

Consequently the following discriminant is non-negative

$$D(\mu_t) - 2\lambda\left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right) \geq 0, \ \forall t \geq 0 \tag{2.2.19}$$

and the log-Sobolev inequality (2.2.15) holds.

For the transportation inequality (2.2.16), take $\rho_1 = \mu_t$ and $\rho_0 = \mu_\infty$. μ_∞ being a stationary solution of (2.0.1), $D(\rho_0) = 0$, which gives

$$W_2(\mu_t, \mu_\infty)^2 \leq \frac{2\left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right)}{\lambda}, \ \forall t \geq 0. \tag{2.2.20}$$

This proves the transportation inequality. Since measure μ_t can be replaced by any measure ρ satisfying (2.2.14) (apply Theorem 2.0.2 to (μ_∞, ρ)), we have

$$\Sigma(\rho) - \Sigma(\mu_\infty) \geq 0 \tag{2.2.21}$$

and if $\Sigma(\rho) = \Sigma(\mu_\infty)$ it follows that $W_2(\rho, \mu_\infty) = 0$. Therefore, μ_∞ is the unique minimizer of the entropy Σ among the class of probability measures ρ satisfying (2.2.14).

Finally, to prove (2.2.17), use successively the log-Sobolev inequality (2.2.15), property (2.1.6), Gronwall's lemma and the transportation inequality (2.2.16). □

2.3 Proof of Theorem 2.0.5

As explained at the end of Section 2.1, the singularity of W at the origin requires additional results, that should be proved for each W of application. Indeed, potentials of interest which satisfy $D^2W(x) \geq \frac{c}{|x|^\eta}$ like $W = -\log|\cdot|$ or $W = |\cdot|^p$, $p \in [0, 2)$, do not satisfy the assumptions of Proposition 2.1 [18]. Therefore, the dissipation property (2.1.6) does not hold necessarily. Moreover, formula (2.2.3) is not justified. To overcome those technical difficulties, we introduce the following additional assumptions.

Assumptions 2.3.1. *$V \in C^2(\mathbb{R}^d)$ and $W \in C^2(\mathbb{R}^d - \{0\})$ satisfy*

(D1) Dissipation property (2.1.6):

$$\frac{d}{dt}\Sigma(\mu_t) \leq -D(\mu_t), \ \forall t \geq 0. \tag{2.3.1}$$

(D2) For any geodesic $(\rho_s)_{0\leq s\leq 1} = ([(1-s)Id + sT]\#\rho_0)_{0\leq s\leq 1}$*,* $\rho_0, \rho_1 \in \mathcal{M}_2$ *with finite entropy, the function* $s \in [0,1] \to \mathcal{W}(\rho_s)$ *is twice differentiable and for all* $s \in (0,1)$

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq \frac{1}{2}\int\limits_{\substack{x,y\in\mathbb{R}^d\\ x\neq y}} \left\langle D^2W(x-y+s(\theta(x)-\theta(y)))\cdot(\theta(x)-\theta(y)), \theta(x)-\theta(y)\right\rangle \rho_0(dx)\rho_0(dy) \tag{2.3.2}$$

where $\theta(x) = x - T(x)$*.*

We are now ready for the proof of Theorem 2.0.5.

Proof of Theorem 2.0.5. Under Assumptions 2.0.4, $\mathcal{V}$ is displacement convex

$$\frac{d^2}{ds^2}\mathcal{V}(\rho_s) \geq 0. \tag{2.3.3}$$

In order to treat $\mathcal{W}$, we follow the proof of Theorem 2.0.2 by fixing some $r > 0$ and taking $\gamma = \frac{1}{(2r)^\eta}$. We have immediately

$$\begin{aligned}\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq \frac{c}{(2r)^\eta}\Big[&(1-\mathbb{P}_{\rho_s}(|x|>r))\int\limits_{|x+s\theta(x)|\leq r}|\theta(x)|^2\rho_0(dx)\\ &-\mathbb{P}_{\rho_s}(|x|>r)\int\limits_{|x+s\theta(x)|\geq r}|\theta(x)|^2\rho_0(dx)\Big]\end{aligned} \tag{2.3.4}$$

and therefore

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq \frac{c}{(2r)^\eta}\mathbb{P}_{\rho_s}(|x|\leq r)\int_{\mathbb{R}^d}|\theta(x)|^2\rho_0(dx) - \frac{c}{(2r)^\eta}\int\limits_{|x+s\theta(x)|\geq r}|\theta(x)|^2\rho_0(dx). \tag{2.3.5}$$

By Cauchy-Schwarz inequality, we have the estimate

$$\int\limits_{|x+s\theta(x)|\geq r}|\theta(x)|^2\rho_0(dx) \leq \sqrt{\int_{\mathbb{R}^d}|\theta(x)|^4\rho_0(dx)}\sqrt{\mathbb{P}_{\rho_s}(|x|\geq r)}. \tag{2.3.6}$$

Using the fact that

$$\int_{\mathbb{R}^d} |\theta(x)|^4 \rho_0(dx) \leq 8m, \tag{2.3.7}$$

we deduce

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq \frac{c}{(2r)^\eta}\mathbb{P}_{\rho_s}(|x| \leq r) W_2^2(\rho_0, \rho_1) - \frac{8^{1/4}m^{1/4}c}{(2r)^\eta}\mathbb{P}_{\rho_s}(|x| \geq r)^{1/4} W_2(\rho_0, \rho_1). \tag{2.3.8}$$

The tail probability $\mathbb{P}_{\rho_s}(|x| \geq r)$ can be estimated by

$$\mathbb{P}_{\rho_s}(|x| \geq r) \leq \frac{1}{r^4}\int_{\mathbb{R}^d} |sx + (1-s)T(x)|^4 \rho_0(dx) \leq \frac{8m}{r^4}. \tag{2.3.9}$$

Moreover, with the simple estimate $W_2(\rho_0, \rho_1)^2 \leq 2\sqrt{m}$, we get

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq \frac{c}{(2r)^\eta} W_2^2(\rho_0, \rho_1) - \frac{\sqrt{8m}c}{2^\eta r^{\eta+1}} W_2(\rho_0, \rho_1) - \frac{4mc}{2^\eta r^{\eta+2}}. \tag{2.3.10}$$

Choosing optimally r yields for some constant $C > 0$ depending only on η, m and c that

$$\frac{d^2}{ds^2}\mathcal{W}(\rho_s) \geq C W_2(\rho_0, \rho_1)^{\eta+2}, \ \forall s \in (0,1), \tag{2.3.11}$$

from which we deduce the desired HWI inequality. □

Corollary 2.3.2. *Assume that Assumptions 2.0.4 and 2.3.1 are satisfied. Let* $(\mu_t)_{t\geq 0}$ *be a solution of (2.0.1) with uniformly bounded fourth-order moments. Set*

$$m = \sup_{t \geq 0}\left(\int_{\mathbb{R}^d} |x|^4 \mu_t(dx)\right). \tag{2.3.12}$$

Then, the following holds for some positive constants δ, C *depending only on* m, c *and* η*:*

(i) Algebraic decay of the entropy

$$\Sigma(\mu_t) - \Sigma(\mu_\infty) \leq \frac{\Sigma(\mu_0) - \Sigma(\mu_\infty)}{\left[1 + \delta\left(\Sigma(\mu_0) - \Sigma(\mu_\infty)\right)^{\eta/(\eta+2)} t\right]^{(\eta+2)/\eta}}, \ \forall t \geq 0. \tag{2.3.13}$$

(ii) Transportation inequality

$$W_2(\mu_t, \mu_\infty) \leq C \left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right)^{\frac{1}{\eta+2}}, \ \forall t \geq 0. \tag{2.3.14}$$

Therefore, we have algebraic stability with respect to W_2 towards the minimizer μ_∞ of the entropy among the class of probability measures ρ satisfying

$$\sup_{t\geq 0} \int_{\mathbb{R}^d} |x|^4 \mu_t(dx) \leq m. \tag{2.3.15}$$

Proof. The proof is similar to the proof of Corollary 2.2.1. The boundedness of moments gives tightness, and we check easily that a limit point μ_∞ (for the weak topology) is a stationary solution. Taking $(\rho_0, \rho_1) = (\mu_\infty, \mu_t)$ in HWI inequality (2.0.13) and noticing that the minimum of power function in $W_2(\mu_t, \mu_\infty)$ (2.0.13) is non-positive, we derive

$$C' \left(\Sigma(\mu_t) - \Sigma(\mu_\infty)\right)^{1+\frac{\eta}{\eta+2}} \leq D(\mu_t) \leq -\frac{d}{dt}\left[\Sigma(\mu_t) - \Sigma(\mu_\infty)\right], \ \forall t \geq 0, \tag{2.3.16}$$

where C' depends only on m, c and η. Integrating leads to the decay estimate (2.3.13). Taking $(\rho_0, \rho_1) = (\mu_t, \mu_\infty)$ in (2.0.13) allows us to derive the transportation inequality (2.3.14). Combining the two inequalities proves result (2.0.14). $\square$

2.4 Application to log gases

This section is devoted to the asymptotic behavior of uni-dimensional log gases. In all the following, V will denote a symmetric external potential and W the logarithmic interaction $W = -\log|\cdot|$. Entropy (2.0.2) is now half of the free entropy introduced by Voiculescu in [110]

$$\Sigma(\mu) = \frac{1}{2}\int_{\mathbb{R}} V(x)\mu(dx) - \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} \log|x-y|\mu(dx)\mu(dy).$$

Define the Hilbert transform of a measure $\mu \in \mathcal{M}$

$$H\mu(x) \equiv -(W * \mu)'(x) = p.v. \int_{\mathbb{R}} \frac{1}{x-y} \mu(dy).$$

where $p.v.$ denotes the principal value.

The granular media equation becomes the Fokker-Planck (2.0.15), with the weak formulation:

$$\frac{d}{dt} \int_{\mathbb{R}} f(x)\mu_t(dx) = -\int_{\mathbb{R}} f'(x) \left(\frac{1}{2} V'(x) - H\mu_t(x) \right) \mu_t(dx), \ \forall f \in \mathcal{C}_0^{\infty}(\mathbb{R}), \tag{2.4.1}$$

or equivalently

$$\begin{aligned} \frac{d}{dt} \int_{\mathbb{R}} f(x)\mu_t(dx) = & \frac{1}{2} \iint_{\mathbb{R}\times\mathbb{R}} \frac{f'(x) - f'(y)}{x-y} \mu_t(dx)\mu_t(dy) \\ & - \frac{1}{2} \int_{\mathbb{R}} V'(x) f'(x) \mu_t(dx), \ \forall f \in \mathcal{C}_0^{\infty}(\mathbb{R}), \end{aligned} \tag{2.4.2}$$

In the rest of the paper, we say that $(\mu_t)_{t\geq 0} \in C(\mathbb{R}_+, \mathcal{M})$, with initial data $\mu_0 \in \mathcal{M}_2 \cap L^{\infty}(\mathbb{R})$ with finite entropy, is a solution of (2.0.15) if (2.4.2) is satisfied and $\mu_t \in L^{\infty}(\mathbb{R})$ for all $t > 0$.

Under the assumption

$$\lim_{|x|\to\infty} V(x) - 2\log|x| = +\infty \tag{2.4.3}$$

the entropy is lower-bounded and there exists a unique minimizer μ_V (see [101])

$$\Sigma(\mu_V) \leq \Sigma(\mu), \ \forall \mu \in \mathcal{M}. \tag{2.4.4}$$

The entropy dissipation (2.0.3) is given by

$$D(\mu) = \int_{\mathbb{R}} \left| \frac{1}{2} V'(x) - H\mu(x) \right|^2 \mu(dx).$$

The existence of solutions and the dissipation property (2.1.6) has been essentially proved by Biane and Speicher in [12] (Theorem 3.1 and Proposition 6.1, see also [83]) under the assumption that V is C^2 and satisfies the growth assumption

$$ax^2 + b \leq \frac{1}{2}xV'(x), \ \forall x \in \mathbb{R}, \tag{2.4.5}$$

for some $a > 0$ and $b \in \mathbb{R}$.

Combining conditions (2.4.3) and (2.4.5), we introduce the following assumptions ensuring the existence of a minimizer and the gradient flow property.

Assumptions 2.4.1. *There exist $a > 0$ and $b \in \mathbb{R}$ such that*

- $\lim_{|x|\to\infty} V(x) - 2\log|x| = +\infty$,
- $ax^2 + b \leq \frac{1}{2}xV'(x), \ \forall x \in \mathbb{R}$.

This section is organized as follows. First, in order to prove Theorem 2.0.6, we establish that $W = -\log|\cdot|$ satisfies a property similar to Assumption 2.3.1 (D2), which allows therefore to apply Theorems 2.0.2 and 2.0.5 to log gases. We provide then in Proposition 2.4.5 uniform bounds for the moments of a solution to (2.0.15). Those estimates are essential to prove tightness with respect to the Wasserstein distance and to bound uniformly the quantities P_r. We state and prove Theorem 2.0.7 that gives algebraic convergence for any convex potential V. Finally, we consider quartic potential V and prove in Theorems 2.0.9 and 2.0.11 exponential stability, when the parameters c and g are small enough in absolute value.

2.4.1 HWI inequality

Despite the singularity of $-\log|\cdot|$ at the origin and its non-convexity, we prove that $\mathcal{W}$ is displacement convex and lower-bound explicitly the second order derivative. The case of log gases will then fall under the scope of application of Theorems 2.0.2 and 2.0.5. Let's first recall the following well-known fact.

Lemma 2.4.2. *Let ρ_0 and ρ_1 be two measures on $\mathbb{R}$ with bounded positive densities. The optimal transport map T carrying ρ_0 to ρ_1 is given by*

$$T = F_{\rho_1}^{-1} \circ F_{\rho_0} \tag{2.4.6}$$

where F_ρ denotes the CDF of measure ρ. Moreover, T is differentiable and for all $r > 0$

$$\sup_{|x|\leq r} T'(x) \leq \frac{\sup_{|x|\leq r} \rho_0(x)}{\inf_{|x|\leq r} \rho_1(T(x))} < \infty. \tag{2.4.7}$$

We now prove Theorem 2.0.6.

Proof of Theorem 2.0.6. We first prove the result for measures with bounded positive densities. We will then extend by approximation to measures with bounded densities. Let $\rho_0, \rho_1 \in \mathcal{M}_2 \cap L^\infty(\mathbb{R})$. Assume either that ρ_0 and ρ_1 have the same center of mass or are both symmetric. Let T be optimal transport map carrying ρ_0 to ρ_1

$$T = F_{\rho_1}^{-1} \circ F_{\rho_0}, \tag{2.4.8}$$

and $(\rho_s)_{0\leq s\leq 1}$ the geodesic from ρ_0 to ρ_1

$$\rho_s = ((1-s)Id + sT)\#\rho_0. \tag{2.4.9}$$

Let $\varepsilon \in \left(0, \frac{1}{2}\right)$, $\eta > 0$ and $B_\eta = \{x \in \mathbb{R} : |x| \leq \eta\}$.

Using the monotonicity of T and Lemma 2.4.2 and setting $R(x,y) \equiv \frac{T(x)-T(y)}{x-y}$, we have

$$0 \leq \sup_{\substack{x,y\in B_\eta \\ x\neq y}} R(x,y) \leq \frac{\sup\limits_{|x|\leq\eta} \rho_0(x)}{\inf\limits_{|x|\leq T(\eta)} \rho_1(x)} < \infty. \tag{2.4.10}$$

Introduce

$$L_\varepsilon(x,y) = \frac{R(x,y)-1}{1-\varepsilon+\varepsilon R(x,y)}, \tag{2.4.11}$$

such that for all $s \in (\varepsilon, 1-\varepsilon)$ and $x \neq y$

$$\left|\frac{(1-s)(x-y)+s(T(x)-T(y))}{(1-\varepsilon)(x-y)+\varepsilon(T(x)-T(y))}\right| = 1+(s-\varepsilon)L_\varepsilon(x,y). \tag{2.4.12}$$

We see from the monotonicity of T and bound (2.4.10) that for all $s \in (\varepsilon, 1-\varepsilon)$

$$1+(s-\varepsilon)L_\varepsilon(x,y) = \frac{1-s+sR(x,y)}{1-\varepsilon+\varepsilon R(x,y)} \geq \frac{\varepsilon}{1-\varepsilon} \geq \varepsilon > 0 \tag{2.4.13}$$

and that for all distinct x and y in B_η

$$1+(s-\varepsilon)L_\varepsilon(x,y) \leq \frac{1+R(x,y)}{1-\varepsilon} \leq 2+2\sup_{\substack{x,y\in B_\eta\\ x\neq y}} R(x,y) < \infty. \tag{2.4.14}$$

Define

$$\omega_\varepsilon = -\frac{1}{2}\iint L_\varepsilon(x,y)\rho_0(dx)\rho_0(dy). \tag{2.4.15}$$

Using the inequality $-\log(1+x)+x \geq 0$ for $x+1>0$, we deduce

$$\begin{aligned}&\mathcal{W}(\rho_s)-\mathcal{W}(\rho_\varepsilon)-(s-\varepsilon)\omega_\varepsilon \geq \\ &\frac{1}{2}\iint\limits_{x,y\in B_\eta}[-\log(1+(s-\varepsilon)L_\varepsilon(x,y))+(s-\varepsilon)L_\varepsilon(x,y)]\rho_0(dx)\rho_0(dy).\end{aligned} \tag{2.4.16}$$

Define the functions

$$G_V : s \in [\varepsilon, 1-\varepsilon] \to \mathcal{V}(\rho_s) = \frac{1}{2}\int V(x+s\theta(x))\rho_0(dx) \tag{2.4.17}$$

and

$$G_W : s \in [\varepsilon, 1-\varepsilon] \to \frac{1}{2}\iint_{x,y\in B_\eta}[-\log(1+(s-\varepsilon)L_\varepsilon(x,y))]\,\rho_0(dx)\rho_0(dy). \tag{2.4.18}$$

By the regularity of V, G_V is twice differentiable with

$$G_V'(\varepsilon) = \frac{1}{2}\int V'(x+\varepsilon\theta(x))\theta(x)\rho_0(dx) \tag{2.4.19}$$

and for all $s \in (\varepsilon, 1-\varepsilon)$

$$G_V''(s) = \frac{1}{2}\int V''(x+s\theta(x))\theta(x)^2\rho_0(dx). \tag{2.4.20}$$

By bounds (2.4.13) and (2.4.14), G_W is twice differentiable as well and its derivatives satisfy

$$G_W'(\varepsilon) = -\frac{1}{2}\iint_{x,y\in B_\eta} L_\varepsilon(x,y)\rho_0(dx)\rho_0(dy) \tag{2.4.21}$$

and for all $s \in (\varepsilon, 1-\varepsilon)$

$$G_W''(s) = \frac{1}{2}\iint_{x,y\in B_\eta} \frac{L_\varepsilon(x,y)^2}{(1+(s-\varepsilon)L_\varepsilon(x,y))^2}\rho_0(dx)\rho_0(dy). \tag{2.4.22}$$

Notice that for all $x \neq y$ and $s \in (\varepsilon, 1-\varepsilon)$

$$\begin{aligned} \frac{L_\varepsilon(x,y)^2}{(1+(s-\varepsilon)L_\varepsilon(x,y))^2} &= \frac{(R-1)^2}{(1-s+sR)^2} \\ &\leq \frac{R^2}{(1-s+sR)^2} + \frac{1}{(1-s+sR)^2} \\ &\leq \frac{2}{\varepsilon^2}. \end{aligned} \tag{2.4.23}$$

Therefore,

$$\iint_{x,y\notin B_\eta} \frac{L_\varepsilon(x,y)^2}{(1+(s-\varepsilon)L_\varepsilon(x,y))^2}\rho_0(dx)\rho_0(dy) \leq \frac{2}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2. \tag{2.4.24}$$

We deduce that for all $s \in (\varepsilon, 1-\varepsilon)$

$$G_W''(s) \geq \frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+s(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) - \frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2. \tag{2.4.25}$$

Using (2.4.16) and Taylor's formula, we deduce for $u \in (\varepsilon, 1-\varepsilon)$ such that

$$\begin{aligned}\Sigma(\rho_{1-\varepsilon}) - \Sigma(\rho_\varepsilon) - (1-2\varepsilon)(G_V'(\varepsilon)+\omega_\varepsilon) &\geq (G_V+G_W)(1-\varepsilon) - (G_V+G_W)(\varepsilon) \\ &\quad - (1-2\varepsilon)(G_V+G_W)'(\varepsilon) \\ &= \frac{1}{2}(G_V+G_W)''(u)(1-2\varepsilon)^2\end{aligned} \tag{2.4.26}$$

where

$$\begin{aligned}(G_V+G_W)''(u) \geq \frac{1}{2}\int & V''(x+u\theta(x))\theta(x)^2\rho_0(dx) \\ &+ \frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+u(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) \\ &- \frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2.\end{aligned} \tag{2.4.27}$$

From there, we follow the computations of the proof of Theorem 2.0.2. Firstly notice that

$$\lim_{\varepsilon\to 0}[G_V'(\varepsilon)+\omega_\varepsilon] = \frac{1}{2}\int V'(x)\theta(x)\rho_0(dx) - \frac{1}{2}\int[R(x,y)-1]\rho_0(dx)\rho_0(dy) \tag{2.4.28}$$

$$= \int\left(\frac{1}{2}V'(x) - H\rho_0(x)\right)\rho_0(dx) \tag{2.4.29}$$

$$\geq -\sqrt{D(\rho_0)}W_2(\rho_0,\rho_1). \tag{2.4.30}$$

Secondly, taking $\gamma = \frac{1}{4r^2}$ and following (2.2.4) - (2.2.11), we obtain

$$\begin{aligned}\frac{1}{2}\int V''(x+u\theta(x))\theta(x)^2\rho_0(dx) + \frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+u(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) \\ \geq \lambda_r W_2(\rho_0,\rho_{1-\varepsilon})^2,\end{aligned} \tag{2.4.31}$$

where λ_r is given either by (2.0.8) or (2.0.9).

Notice that this estimate is independent of $u \in (\varepsilon, 1-\varepsilon)$. Recalling that $W_2(\rho_0, \rho_{1-\varepsilon}) =$

$(1-\varepsilon)W_2(\rho_0,\rho_1)$, we conclude that for all $\eta>0$

$$\Sigma(\rho_{1-\varepsilon})-\Sigma(\rho_\varepsilon)\geq(1-2\varepsilon)[G_V'(\varepsilon)+\omega_\varepsilon]+\lambda_r(1-\varepsilon)^2W_2(\rho_0,\rho_1)^2 \\ -\frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2. \tag{2.4.32}$$

From there, do successively $\eta\to\infty$ and $\varepsilon\to 0$ to derive the result. If ρ_0 and ρ_1 do not have positive densities, apply (2.0.7) for the sequences

$$\rho_i^\delta(dx)=\int\frac{e^{-(x-y)^2/(2\delta)}}{\sqrt{2\pi\delta}}\rho_i(dy),\ \delta>0,\ i=0,1, \tag{2.4.33}$$

and let $\delta\to 0$. Notice that ρ_i^δ has a positive density and belong to $L^\infty(\mathbb{R})$ with $|\rho_i^\delta|_{L^\infty(\mathbb{R})}\leq\frac{1}{\sqrt{2\pi\delta}}$, and that the map $\rho_i\mapsto\rho_i^\delta$ leaves the center of mass or the symmetry of the measure ρ_i invariant. Therefore, using the result for measures with bounded positive densities, we have for all $\delta>0$

$$\Sigma(\rho_1^\delta)-\Sigma(\rho_0^\delta)\geq-\sqrt{D(\rho_0^\delta)}W_2(\rho_0^\delta,\rho_1^\delta)+\lambda_rW_2(\rho_0^\delta,\rho_1^\delta)^2. \tag{2.4.34}$$

We have easily that $\lim_{\delta\to 0}W_2(\rho_0^\delta,\rho_1^\delta)=W_2(\rho^\delta,\rho_1^\delta)$.

Let X and Y be two independent random variables with probability distribution ρ_0 and let Z be a random variable independent with X and Y, and with normal distribution. Then, we can write

$$\Sigma(\rho_0^\delta)=\frac{1}{2}\mathbb{E}\left[V'(X+\delta Z)\right]-\frac{1}{2}\mathbb{E}\left[\log\left(X+\delta Z-Y-\delta Z\right)\right], \tag{2.4.35}$$

and

$$\lim_{\delta\to 0}\Sigma(\rho_0^\delta)=\Sigma(\rho_0). \tag{2.4.36}$$

Next, write

$$
\begin{aligned}
|D(\rho_0) - D(\rho_0^\delta)| \le{}& \frac{1}{4}\left|\int V'^2\rho_0 - \int V'^2\rho_0^\delta\right| \\
&+ \left|\int (H\rho_0)^2\rho_0 - \int (H\rho_0^\delta)^2\rho_0\right| \\
&+ \left|\int (H\rho_0^\delta)^2\rho_0 - \int (H\rho_0^\delta)^2\rho_0^\delta\right| \\
&+ \frac{1}{2}\left|\int V'(H\rho_0)\rho_0 - V'H(\rho_0^\delta)\rho_0^\delta\right|.
\end{aligned} \tag{2.4.37}
$$

Notice that

$$
\lim_{\delta\to 0}\int V'^2\rho_0^\delta = \lim_{\delta\to 0}\mathbb{E}\left[V'(X+\delta Z)^2\right] = \int V'^2\rho_0. \tag{2.4.38}
$$

Upper-bound the second term of (2.4.37)

$$
\begin{aligned}
\left|\int (H\rho_0)^2\rho_0 - \int (H\rho_0^\delta)^2\rho_0^\delta\right| &\le |\rho_0|_{L^\infty}\int |H(\rho_0-\rho_0^\delta)||H(\rho_0+\rho_0^\delta)| \\
&\le |\rho_0|_{L^\infty}\sqrt{\int H(\rho_0-\rho_0^\delta)^2}\sqrt{\int H(\rho_0+\rho_0^\delta)^2} \\
&= \sqrt{2}|\rho_0|_{L^\infty}\sqrt{\int (\rho_0-\rho_0^\delta)^2}\sqrt{\int \rho_0^2+(\rho_0^\delta)^2} \\
&\le 2|\rho_0|_{L^\infty}\sqrt{\int (\rho_0-\rho_0^\delta)^2}\sqrt{\int \rho_0^2},
\end{aligned} \tag{2.4.39}
$$

where we used that ρ_0 and ρ_0^δ belong to $\mathcal{M}(\mathbb{R})\cap L^\infty(\mathbb{R})\subset L^2(\mathbb{R})$, the isometry property of the Hilbert transform and Young's convolution inequality $\int(\rho_0^\delta)^2\le\int\rho_0^2$.

Estimate the third term of (2.4.37) by

$$
\left|\int (H\rho_0^\delta)^2\rho_0 - \int (H\rho_0^\delta)^2\rho_0^\delta\right| \le \left|\int (H\rho_0^\delta)^4\right|^{1/2}\left|\int (\rho_0-\rho_0^\delta)^2\right|^{1/2}. \tag{2.4.40}
$$

Using the bounded-ness of the Hilbert transform for the norm L^4 and the Young's convolution inequality, we deduce

$$
\left|\int (H\rho_0^\delta)^2\rho_0 - \int (H\rho_0^\delta)^2\rho_0^\delta\right| \le C\|\rho_0\|_{L^\infty}^{3/4}\sqrt{\int (\rho_0 - rho_0^\delta)^2} \tag{2.4.41}
$$

for some constant $C > 0$.

We see easily that under the assumption $\rho_0 \in L^2(\mathbb{R})$

$$\lim_{\delta \to 0} \int (\rho_0 - \rho_0^\delta)^2 = 0. \tag{2.4.42}$$

and therefore with (2.4.39) and (2.4.41), we conclude that

$$\lim_{\delta \to 0} \left| \int (H\rho_0^\delta)^2 \rho_0 - \int (H\rho_0^\delta)^2 \rho_0^\delta \right| + \left| \int (H\rho_0)^2 \rho_0 - \int (H\rho_0^\delta)^2 \rho_0^\delta \right| = 0. \tag{2.4.43}$$

In a very similar way, we show that

$$\lim_{\delta \to 0} \left| \int V'(H\rho_0)\rho_0 - V'H(\rho_0^\delta)\rho_0^\delta \right| = 0. \tag{2.4.44}$$

Finally, taking the limit $\delta \to 0$ in (2.4.34) gives the result.

□

Remark 2.4.3. *Our theorem is similar to Theorem 1.4 [83], but we do not assume any assumption regarding the compactness of the support. Therefore, our proof can be extended to gases with a diffusive internal energy (like in [107] for example). In Theorem 5 [81], the authors established a HWI inequality for log gases. Their result can be recovered by considering a simpler version of estimate (2.4.16). Indeed, by applying* $-\log(1+x)+x \geq 0$ *on the whole space, we find*

$$\mathcal{W}(\rho_s) - \mathcal{W}(\rho_0) - (s - \varepsilon)\omega_\varepsilon \geq 0. \tag{2.4.45}$$

Using this crude estimate, (2.4.32) becomes

$$\Sigma(\rho_{1-\varepsilon}) - \Sigma(\rho_\varepsilon) \geq (1 - 2\varepsilon)\left[G_V'(\varepsilon) + \omega_\varepsilon\right] + \frac{1}{2} \inf_{x \in \mathbb{R}} V''(x)(1-\varepsilon)^2 W_2(\rho_0, \rho_1)^2. \tag{2.4.46}$$

Therefore under the assumption that $\inf_{x\in\mathbb{R}} V''(x) \geq 2\lambda > 0$*, letting* $\varepsilon \to 0$*, we find*

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)}W_2(\rho_0, \rho_1) - \frac{\lambda}{2}W_2(\rho_0, \rho_1)^2. \tag{2.4.47}$$

2.4.2 Moment estimates

We establish upper-bounds for the moments of solutions of (2.0.15). Those estimates are useful to prove tightness of a solution and to bound uniformly in t the tail probabilities $\mathbb{P}_{\mu_t}(|x| \geq r)$. The idea is essentially to use dynamics (2.4.2) to get a differential inequation satisfied by the moments.

Let's start with the following lemma, justifying extension of test functions to power functions. We delay the proof to the appendix.

Lemma 2.4.4. *Let* $(\mu_t)_{t\geq 0}$ *be a solution of (2.0.15) such that for all* $t \geq 0$*,* μ_t *has a finite moment of order* $p \geq 1$*. Assume that*

1. $s \mapsto \int_{\mathbb{R}}(1 + |V'(x)|)|x|^p\mu_s(dx) \in L^1_{loc}([0, \infty))$,
2. $x \mapsto |V'(x)||x|^p \in L^1(\mathbb{R}, \mu_s)$ *for all* $s \geq 0$.

Then the power function $x \in \mathbb{R} \to |x|^p$ *is a valid test function in (2.4.2).*

Using this lemma, we prove

Proposition 2.4.5. *Let* $(\mu_t)_{t\geq 0}$ *be a solution of (2.0.15) with finite moments up to order* $p + 2$ *for some* $p \geq 2$*. Then, there exists* $M_p > 0$ *such that*

$$\sup_{t\geq 0} \int_{\mathbb{R}} |x|^p \mu_t(dx) \leq M_p.$$

Proof. Denote $m_p(t) = \int_{\mathbb{R}} |x|^p\mu_t(dx)$ for $t, p \geq 0$. According to Lemma 2.4.4, we can apply (2.4.2) with the test function $f : x \mapsto |x|^p$ to obtain

$$\dot{m}_p(t) \leq \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} \frac{|f'(x) - f'(y)|}{|x - y|}\mu_t(dx)\mu_t(dy) - pam_p(t) + p|b|pm_{p-2}(t). \tag{2.4.48}$$

Without loss of generality, we may assume that $|x| < |y|$. We see easily that

$$\frac{|f'(x) - f'(y)|}{|x-y|} \leq p\frac{|x|x|^{p-2} - y|y|^{p-2}|}{|x-y|} \leq p(|x|^{p-2} + |y|^{p-2}) + p|x|\frac{||x|^{p-2} - |y|^{p-2}||}{||x| - |y||}.$$

We check easily that if $p \geq 3$, then the previous inequality implies

$$\frac{|f'(x) - f'(y)|}{|x-y|} \leq p(|x|^{p-2} + |y|^{p-2}) + p(p-2)(|x|^{p-2} + |y|^{p-2})$$

and if $2 \leq p < 3$

$$\frac{|f'(x) - f'(y)|}{|x-y|} \leq p(|x|^{p-2} + |y|^{p-2}) + p(p-2)|x|^{p-2}.$$

In both cases for all $x, y \in \mathbb{R}$

$$\frac{|f'(x) - f'(y)|}{|x-y|} \leq p(|x|^{p-2} + |y|^{p-2}) + p(p-2)(|x|^{p-2} + |y|^{p-2}). \tag{2.4.49}$$

Using the fact that $m_l(t) \leq m_p^{l/p}(t)$ for $l \leq p$, and inserting (2.4.49) in (2.4.48), we deduce the ordinary differential inequation

$$\dot{m}_p(t) \leq -pam_p(t) + p\,(|b| + p - 1)\,m_p^{(p-2)/(p)}(t),\ t \geq 0. \tag{2.4.50}$$

This inequality can be written as

$$\dot{m}_p(t) \leq Q(m_p(t)),\ t \geq 0 \tag{2.4.51}$$

where $Q(x) = -pax + p(|b| + p - 1)\,x^{(p-2)/p}$ satisfies $Q(x) \underset{+\infty}{\sim} -pax$.

We will show that this inequality implies the uniform boundedness of moments. Consider M defined by:

$$M = \sup\{x \geq 0 : Q(x) \geq 0\} = \left(\frac{|b| + p - 1}{a}\right)^{p/2}.$$

Notice that m_p is non-increasing for $m_p > M$. Set $M_p \equiv \max(M, m_p(0))$ and $T = \{t \geq 0 : m_p(t) \leq M\}$. Without loss of generality we may assume that $0 \in T$, otherwise m_p is decreasing until $t \in T$. The set $\mathbb{R}_+ \setminus T$ is open (in $\mathbb{R}_+$) so can be written as

$$\mathbb{R}_+ \setminus T = \bigcup_i (s_i, t_i).$$

On the one hand

$$\dot{m}_p(t) < 0 \text{ for } s_i < t < t_i$$

and on the other hand

$$m_p(s_i) = m_p(t_i) = M.$$

Those two contradictory statements imply that $T = \mathbb{R}_+$ and for all $t \geq 0$:

$$m_p(t) \leq M_p.$$

This achieves the proof. □

Remark 2.4.6. *In the next section, we consider quartic potential V, and we derive in the same way Proposition 2.4.7 providing uniform bounds for the second order moments of a solution in a more precise way. Still, we would like to stress the usefulness of the more general Proposition 2.4.5, which allows showing the stability of any solution with finite moments of order $p+2 \geq 2$ towards a stationary measure with respect to the Wasserstein distance of order p. Additionally, we will use this proposition for the proof of Theorem 2.0.7.*

2.4.3 Stability for convex potentials

From the proofs of Theorems 2.0.5 and 2.0.6, we derive Theorem 2.0.7.

Proof of Theorem 2.0.7. Fix $\varepsilon \in (0, 1/2)$ and $r > 0$. Let $\rho_0, \rho_1 \in \mathcal{M}_2$ and T be the optimal transport map from ρ_0 to ρ_1. Set $\theta = T - Id$. Denote $(\rho_s)_{s \in [0,1]}$ the geodesic between ρ_0 and ρ_1. Following to the proof of Theorem 2.0.6, and using that the fact that$D^2V \geq 0$, there exists $u \in (\varepsilon, 1-\varepsilon)$ such that

$$\begin{aligned}\Sigma(\rho_{1-\varepsilon}) - \Sigma(\rho_\varepsilon) \geq (1-2\varepsilon)[G_V'(\varepsilon) + \omega_\varepsilon] \\ + \frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+u(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) \\ - \frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2,\end{aligned} \tag{2.4.52}$$

where ω_ε is defined as in (2.4.15).

From there, follow the proof of Theorem 2.0.5 by taking $c = 1$ and $\eta = 2$. Estimate (2.0.13) becomes

$$\frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+u(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) \geq CW_2(\rho_0,\rho_1)^4. \tag{2.4.53}$$

Inserting this estimate in (2.4.52) and then letting $\eta \to \infty$ and $\varepsilon \to 0$, we get

$$\Sigma(\rho_1) - \Sigma(\rho_0) \geq -\sqrt{D(\rho_0)}W_2(\rho_0,\rho_1) + CW_2(\rho_0,\rho_1)^4. \tag{2.4.54}$$

The result follows by mimicking the proof of Corollary 2.3.2 and using the tightness of $(\mu_t)_{t\geq 0}$ ensured by Proposition 2.4.5. □

2.4.4 Stability for confining quartic potential

In this subsection, we consider the confining quartic potential $V(x) = \frac{x^4}{4} + c\frac{x^2}{2}$. We recall the following results:

- If $c \geq 0$, then V satisfies Ledoux's assumptions [81], which allows proving exponential convergence towards the equilibrium measure μ_V [83], whose density is given by

$$\mu_V(dx) = \frac{1}{\pi}\left(\frac{1}{2}x^2 + b\right)\sqrt{a^2 - x^2}1_{[-a,a]}(x), \tag{2.4.55}$$

where

$$a^2 = \frac{\sqrt{4c^2+48}-2c}{3}, \quad b = \frac{c+\sqrt{\frac{c^2}{4}+3}}{3}.$$

- If $-2 < c < 0$, μ_V is the unique stationary measure, and stability is proved in [30]. The density is again given by (2.4.55).

- If $c < -2$, Donati-Martin et al. have proved in [30] that there could be multiple stationary measures.

The main result of this subsection will follow from Theorem 2.0.6 applied to quartic potential V with c negative and close enough to zero.

The same considerations as in the proof of Proposition 2.4.5 lead to the following estimate of second order moment.

Proposition 2.4.7. *Let $(\mu_t)_{t\geq 0}$ be a solution of (2.0.15) with quartic V $(C1)$. Assume that $(\mu_t)_{t\geq 0}$ have finite fourth moments for all times $t \geq 0$. Then*

$$\sup_{t\geq 0} \int_{\mathbb{R}} x^2 \mu_t(dx) \leq \max\left(\int_{\mathbb{R}} x^2 \mu_0(dx), \frac{-c+\sqrt{c^2+4}}{2}\right). \tag{2.4.56}$$

We are now ready to prove the Theorem 2.0.9.

Proof of Theorem 2.0.9. Let $\mu_0 \in \mathcal{M}$ which satisfies the moment condition (2.0.17) and $(\mu_t)_{t\geq 0}$ be a solution of (2.0.15). According to Proposition 2.4.5, the second order moments of $(\mu_t)_{t\geq 0}$ are uniformly bounded in time. Σ is lower-bounded and D is lower semi-continuous, so similarly to the proof of Corollary 2.2.1, $(\mu_t)_{t\geq 0}$ weakly converges towards a stationary solution μ_∞. According to Proposition 2.7 [30], the unique stationary solution for $c \in [-2, 0)$ is the minimizer of the entropy μ_V. Finally, the uniform bounds of the moments give tightness and convergence with respect to the Wasserstein distance.

Assume that $(\mu_t)_{t\geq 0}$ has a fixed center of mass. According to Theorem 2.0.6, for all measures $\rho_0, \rho_1 \in \mathcal{M}_2$ with finite entropy

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq W_2(\rho_0, \rho_1)\sqrt{D(\rho_0)} - \frac{\lambda}{2} W_2(\rho_0, \rho_1)^2, \tag{2.4.57}$$

with constants (α, β, γ) given by

$$\begin{cases} \alpha = 3r^2 + c, \\ \beta = -c, \\ \gamma = \frac{1}{4r^2}, \end{cases} \tag{2.4.58}$$

and the rate

$$\lambda = \frac{c}{2} + \sup_{r>0} \left[\frac{\min\left(3r^2, \frac{1}{2r^2}\right)}{2} - \frac{P_r}{2r^2} \right]. \tag{2.4.59}$$

Under the assumption that

$$\int_{\mathbb{R}} x^2 \mu_0(dx) \leq \frac{-c + \sqrt{c^2+4}}{2} \tag{2.4.60}$$

we have for $c \in [-2, 0)$

$$P_r \leq \frac{-c + \sqrt{c^2+4}}{2r^2}. \tag{2.4.61}$$

It follows that the constant λ (2.4.59) can be lower-bounded by

$$\lambda \geq \frac{c}{2} + \frac{1}{4} \sup_{r>0} \frac{1}{r^4} \left[\min\left(6r^6, r^2\right) - (-c + \sqrt{c^2+4}) \right]. \tag{2.4.62}$$

We solve this optimization easily and find

$$\lambda \geq \frac{1}{16(-c + \sqrt{c^2+4})} + \frac{c}{2}. \tag{2.4.63}$$

We check that this last quantity is positive for any $c \in (c^*, 0)$ with

$$c^* = -\frac{1}{4\sqrt{17}}. \tag{2.4.64}$$

From there, an obvious variant of the reasoning of Corollary 2.2.1 makes it possible to conclude.

In the case of μ_0 symmetric, the rate λ can be slightly improved. Indeed, we can improve estimate (2.2.9) in the following way. Write for ρ_0 and ρ_1 symmetric

$$((1-s)x+sT(x))^2 = (1-s)^2x^2 + s^2T(x) + 2s(1-s)xT(x). \tag{2.4.65}$$

Now, T being odd and the gradient of a convex function, we have $T(0)=0$ and

$$xT(x) \geq 0. \tag{2.4.66}$$

Therefore

$$((1-s)x+sT(x))^2 \geq (1-s)^2x^2 + s^2T(x). \tag{2.4.67}$$

We deduce that if $s \leq \frac{1}{2}$

$$|(1-s)x+sT(x)| \leq r \implies |x| \leq 2r, \tag{2.4.68}$$

and if $s \geq \frac{1}{2}$

$$|(1-s)x+sT(x)| \leq r \implies |T(x)| \leq 2r. \tag{2.4.69}$$

Therefore

$$\int_{|x|\leq r} \rho_s(dx) \geq 1_{s\leq 1/2} \int_{|x|\leq 2r} \rho_0(dx) + 1_{s\geq 1/2} \int_{|x|\leq 2r} \rho_1(dx) \geq 1 - P_{2r}. \tag{2.4.70}$$

This bound is better because $P_{2r} \leq 2P_r$. Consequently, in the case of μ_0 symmetric, the optimal λ is given by

$$\lambda = \frac{1}{4}\sup_{r>0}\frac{1}{r^2}\min\left(6r^4, 1-P_{2r}\right) + \frac{c}{2}. \tag{2.4.71}$$

Similarly, λ can be lower-bounded by

$$\begin{aligned} \lambda &\geq \frac{1}{4}\sup_{r>0}\frac{1}{r^4}\min\left(6r^6, r^2 - \frac{(-c+\sqrt{c^2+4})}{8}\right) + \frac{c}{2} \\ &= \frac{1}{2(-c+\sqrt{c^2+4})} + \frac{c}{2}. \end{aligned} \tag{2.4.72}$$

This last quantity is positive for any $c \in (c^*, 0)$ with

$$c^* = -\frac{1}{\sqrt{6}}. \tag{2.4.73}$$

The result follows. □

Finally, we end this section by giving the proof of Theorem 2.0.10, which overrides the assumptions of a fixed center of mass or symmetry. The optimal c^* obtained is numerically much smaller than the constant obtained in the previous theorem. Therefore, the value of our result lies in its proof. We first need two technical lemmas. The first one is proved in the appendix, the second proof is omitted and follows exactly the proof of Proposition 2.4.5.

Lemma 2.4.8. *Let $(\rho_s)_{0\le s\le 1} = (((1-s)Id+sT)\#\rho_0)_{0\le s\le 1}$ be a geodesic between $\rho_0 \in \mathcal{M}_4$ and a symmetric measure $\rho_1 \in \mathcal{M}_4$. Then for all $s \in [0,1]$ and $r > 0$*

$$\int\limits_{|x|\le r} x^2\rho_s(dx) \ge \frac{1}{2}\min\left(\int x^2\rho_0(dx), \int x^2\rho_1(dx)\right) - \frac{8}{r^2}\max\left(\int x^4\rho_0(dx), \int x^4\rho_1(dx)\right). \tag{2.4.74}$$

Lemma 2.4.9. *Let $(\mu_t)_{t\ge 0}$ be a solution of (2.0.15) with quartic V. Assume that $(\mu_t)_{t\ge 0}$ have finite sixth moments for all times $t \ge 0$. Then*

$$\inf_{t>0}\int x^2\mu_t(dx) \ge \min\left(\int x^2\mu_0(dx), \left(\frac{2}{-c+\sqrt{c^2+16}}\right)^4\right) \tag{2.4.75}$$

and

$$\sup_{t\ge 0}\int x^4\mu_t(dx) \le \max\left(\int x^4\mu_0(dx), \left(\frac{-c+\sqrt{c^2+12}}{2}\right)^2\right). \tag{2.4.76}$$

Proof of Theorem 2.0.10. As before, tightness is clear. It is enough to show a HWI inequality with rate $\lambda > 0$. We use the notations of the proof of Theorem 2.0.6. In the following ρ_0 and ρ_1 denote respectively μ_t, $t \geq 0$ and μ_V. Start with (2.4.27)

$$
\begin{aligned}
(G_V+G_W)''(u) \geq & \frac{1}{2}\int V''(x+u\theta(x))\theta(x)^2\rho_0(dx) \\
& + \frac{1}{2}\iint \frac{(\theta(x)-\theta(y))^2}{(x-y+u(\theta(x)-\theta(y)))^2}\rho_0(dx)\rho_0(dy) \\
& - \frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2.
\end{aligned}
\tag{2.4.77}
$$

Therefore, denoting $\psi(x) = 3x^2$, plugging $V''(x+u\theta(x)) = \psi(x+u\theta(x)) + c$ and splitting the integrals according to $|x+u\theta(x)| \leq r$, we have

$$
\begin{aligned}
(G_V+G_W)''(u) \geq & \frac{1}{2}\int_{|x+u\theta(x)|\leq r}\psi(x+u\theta(x))\theta(x)^2\rho_0(dx) + \frac{3r^2}{2}\int_{|x+u\theta(x)|>r}\theta(x)^2\rho_0(dx) \\
& + \frac{1}{8r^2}\iint_{\substack{|x+u\theta(x)|\leq r\\ |y+u\theta(y)|\leq r}}(\theta(x)-\theta(y))^2\rho_0(dx)\rho_0(dy) \\
& + \frac{c}{2}W_2(\rho_0,\rho_1)^2 - \frac{1}{\varepsilon^2}\int_{|x|\geq\eta}\rho_0(dx).
\end{aligned}
\tag{2.4.78}
$$

We now adapt an idea used in section 4.5 [18] to our specific integrals. Write

$$
\begin{aligned}
\iint_{|x+u\theta(x)|\leq r\, |y+u\theta(y)|\leq r}(\theta(x)-\theta(y))^2\rho_0(dx)\rho_0(dy) = & \; 2\int_{|x+u\theta(x)|\leq r}\rho_0(dx)\int_{|x+u\theta(x)|\leq r}\theta(x)^2\rho_0(dx) \\
& - 2\left(\int_{|x+u\theta(x)|\leq r}\theta(x)\rho_0(dx)\right)^2.
\end{aligned}
\tag{2.4.79}
$$

From

$$
\begin{aligned}
\left(\int_{|x+u\theta(x)|\leq r}\theta(x)\rho_0(dx)\right)^2 \leq & \int_{|x+u\theta(x)|\leq r}\frac{1}{1+2r^2\psi(x+u\theta(x))}\rho_0(dx) \\
& \times \int_{|x+u\theta(x)|\leq r}(1+2r^2\psi(x+u\theta(x)))\theta(x)^2\rho_0(dx)
\end{aligned}
\tag{2.4.80}
$$

we derive

$$
\begin{aligned}
(G_V+G_W)''(u) \geq \frac{1}{4r^2} & \left(\int\limits_{|x+u\theta(x)|\leq r} \left[1-\frac{1}{1+2r^2\psi(x+u\theta(x))}\right]\rho_0(dx)\right) \\
& \times \int\limits_{|x+u\theta(x)|\leq r} (1+2r^2\psi(x+u\theta(x)))\theta(x)^2\rho_0(dx) \\
& + \frac{3r^2}{2}\int_{|x+u\theta(x)|>r} \theta(x)^2\rho_0(dx) + \frac{c}{2}W_2(\rho_0,\rho_1)^2 \\
& - \frac{1}{\varepsilon^2}\int_{|x|\geq\eta} \rho_0(dx).
\end{aligned} \tag{2.4.81}
$$

The goal is to lower-bound the term

$$
\int\limits_{|x+u\theta(x)|\leq r} \left[1-\frac{1}{1+2r^2\psi(x+u\theta(x))}\right]\rho_0(dx) = \int_{|x|\leq r}\frac{2r^2x^2}{1+2r^2x^2}\rho_u(dx). \tag{2.4.82}
$$

In [18], a similar quantity appears, and the authors prove that it is bounded away from zero by using the internal energy. In absence of any internal energy, we have to exploit the logarithmic repulsive interaction to show that ρ_s - the interpolation between μ_V and μ_t - cannot be concentrated at the origin. This intuition is implemented by looking at the second order moments.

Set $m = \left(\frac{2}{-c+\sqrt{c^2+16}}\right)^4$ and $M = \left(\frac{-c+\sqrt{c^2+12}}{2}\right)^2$. By Lemma 2.4.8 ($\rho_1 = \mu_V$ is indeed symmetric), we have

$$
\int_{|x|\leq r}\frac{2r^2x^2}{1+2r^2x^2}\rho_u(dx) \geq \frac{2r^2}{1+2r^4}\left[\frac{1}{2}m - \frac{8}{r^2}M\right]. \tag{2.4.83}
$$

Use this estimate and $1+2r^2\psi(x+u\theta(x)) \geq 1$ to deduce

$$
(G_V+G_W)''(u) \geq \lambda_r W_2(\rho_0,\rho_1)^2 - \frac{1}{\varepsilon^2}\left[\int_{|x|\geq\eta}\rho_0(dx)\right]^2 \tag{2.4.84}
$$

where λ_r is given by

$$\lambda_r = \frac{c}{2} + \frac{1}{2}\min\left(3r^2, \frac{1}{1+2r^4}\left[\frac{1}{2}m - \frac{8}{r^2}M\right]\right). \tag{2.4.85}$$

From this point, follow the end of the proof of Theorem 2.0.6 to conclude

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq W_2(\rho_0, \rho_1)\sqrt{D(\rho_0)} - \frac{\lambda_r}{2}W_2(\rho_0, \rho_1)^2. \tag{2.4.86}$$

We see clearly that $\sup_{r>0}\lambda_r$ can be made positive for c negative and close enough to zero. The optimal rate can be lower-bounded by

$$\lambda = \sup_{r>0}\lambda_r \geq \frac{c}{2} + \frac{1}{4}\sup_{r^2\geq 16M/m}\left[\frac{1}{r^4}\left(\frac{m}{2} - \frac{8M}{r^2}\right)\right] = \frac{c}{2} + \frac{m^3}{13924M^2}. \tag{2.4.87}$$

Numerically, we find that $\lambda > 0$ for $c > -3.00 \times 10^{-9}$.

$\square$

2.4.5 Stability for non-confining quartic potentials

The difficulty of the non-confining quartic potential

$$V(x) = g\frac{x^4}{4} + \frac{x^2}{2},\ g < 0, \tag{2.4.88}$$

is twofold: how do we define a solution of (2.0.15) and what are the equilibrium and stationary measures? Indeed, solutions of (2.0.15) may explode and Assumption 2.4.1 is not satisfied by this kind of potential, so we do not necessarily have the existence of an equilibrium measure. In [4], authors give insights on how to restart solutions after explosion time. This approach is different from the idea of Biane and Speicher in section 7.1 [12], where the authors explain how to define bounded solutions of (2.0.15), which stay around the origin, and exhibit a good candidate for the equilibrium measure. In the following, we will adopt this last point of view.

Let $(\mathcal{A}, \tau, (\mathcal{A}_t)_{t\geq 0}, (S_t)_{t\geq 0})$ be a filtered non-commutative probability space with a free Brownian motion $(S_t)_{t\geq 0}$. Denote $\mathcal{A}^{op}$ the algebra $\mathcal{A}$ with operation $a \cdot_{\mathcal{A}^{op}} b = b \cdot_{\mathcal{A}} a$ for $a, b \in \mathcal{A}$. See [11], [10] and [12] for more details on free probability and free stochastic processes. For all $t > 0$, denote ρ_t the density of semicircular distribution of mean zero and variance t. In those conditions, S_t has a distribution given by ρ_t.

$$\rho_t(dx) = \frac{2}{\pi t}\sqrt{t - x^2}1_{[-\sqrt{t},\sqrt{t}]}(x)dx, \ \forall t > 0. \tag{2.4.89}$$

In the framework of free probabilities, a solution $(\mu_t)_{t\geq 0}$ of (2.0.15) can be seen as the distribution of the free stochastic process $(X_t)_{t\geq 0}$, solution of the free stochastic differential equation

$$\begin{cases} X_t = X_0 + S_t - \frac{1}{2}\int_0^t V'(X_t)dt \\ Law(X_0) = \mu_0 \end{cases} \tag{2.4.90}$$

We shall denote $\mu_t = Law(X_t)$. We now recall the free Itô formula and the free Burkholder-Davis-Gundy inequality. For any polynomial $\phi = \sum_{n\geq 0} \phi_n X^n$, denote for $X \in \mathcal{A}$ the element $\partial\phi(X)$ of $\mathcal{A} \otimes \mathcal{A}^{op}$ defined by

$$\partial\phi(X) = \sum_{n\geq 0} \phi_n \sum_{k=0}^{n-1} X^k \otimes X^{n-k-1}. \tag{2.4.91}$$

Introduce the operator Δ_t defined by

$$\Delta_t\phi(x) = 2\frac{d}{dx}\left(\int_{\mathbb{R}} \frac{\phi(x) - \phi(y)}{x - y}\rho_t(dy)\right). \tag{2.4.92}$$

The free Itô's formula is then written

$$\phi(X_t) = \phi(X_0) + \int_0^t \partial\phi(X_s)\sharp dX_s + \frac{1}{2}\int_0^t \Delta_s\phi(X_s)ds, \ \forall t \geq 0, \tag{2.4.93}$$

where $\int_0^t \partial\phi(X_s)\sharp dX_s$ denotes the free stochastic integral of the bi-process $(\partial\phi(X_s))_{s\geq 0}$ with respect to the free Itô process $(X_t)_{t\geq 0}$.

The free Burkholder-Davis-Gundy inequality states that

$$\left\| \int_0^t Y_s \sharp dS_s \right\| \le 2\sqrt{2} \sqrt{\int_0^t \|Y_s\|^2 ds}, \ \forall t \ge 0, \tag{2.4.94}$$

for any free Itô process $(Y_t)_{t \ge 0}$.

We are now ready to define solutions to (2.4.90), detailing an idea of Biane and Speicher succinctly presented in [12].

Proposition 2.4.10. *Let $g \in \left(-\frac{1}{81+36\sqrt{5}}, 0\right)$ and μ_0 be an initial with support included in $(-m(g), m(g))$, with $m(g)$ given by*

$$m(g) = \sqrt{-\frac{1}{3g} - \frac{4}{\sqrt{-g}} - 3}. \tag{2.4.95}$$

Then the process $(X_t)_{t \ge 0}$ defined by (2.4.90) exists for all times and the support of μ_t remains in a set of strict convexity of V. More precisely, if for some $m \in [0, m(g))$

$$supp(\mu_0) \subset [-m, m] \tag{2.4.96}$$

then for all $t \ge 0$

$$supp(\mu_t) \subset \left[-\sqrt{m^2 + \frac{4}{\sqrt{-g}} + 3}, \sqrt{m^2 + \frac{4}{\sqrt{-g}} + 3} \right], \tag{2.4.97}$$

and

$$\inf_{x \in supp(\mu_t)} V''(x) = 1 + 3g \left(m^2 + \frac{4}{\sqrt{-g}} + 3 \right) > 0. \tag{2.4.98}$$

Proof. Let $\varepsilon \in (0,1)$. Set $h = \sqrt{\frac{1}{-3g}}$ and $M = \sqrt{1-\varepsilon} h$. With those notations

$$\inf_{|x| \le M} V''(x) = 1 + 3gM^2 = \varepsilon > 0. \tag{2.4.99}$$

The idea of the proof is to show that the stopping time

$$T = \inf\{t \geq 0 : \|X_t\| > M\}$$

is almost surely infinite.

Fix $\delta > 0$. We apply free Itô's formula to $e^{\delta t} G(X_t)$ with $G(x) = x^2$:

$$e^{\delta t} X_t^2 = X_0^2 + \int_0^t e^{\delta s} \delta X_s^2 ds + \int_0^t e^{\delta s} 1_{\mathcal{A}} ds + \int_0^t e^{\delta s} \left(X_s \otimes 1_{\mathcal{A}} + 1_{\mathcal{A}} \otimes X_s \right) \# dS_s \\ - \frac{1}{2} \int_0^t e^{\delta s} X_s V'(X_s) ds, \tag{2.4.100}$$

where we used the fact that $\Delta_s G(x) = 2$, $x \in \mathbb{R}$. Noticing that for $s < T$

$$\delta X_s^2 - \frac{1}{2} X_s V'(X_s) = \frac{X_s^2}{2} [2\delta - g X_s^2 - 1] \leq \frac{X_s^2}{2} [2\delta - g M^2 - 1] = \frac{X_s^2}{2} \left(2\delta - \frac{2+\varepsilon}{3} \right) \tag{2.4.101}$$

as self-adjoint operators, because $\|X_s\| \leq M$ (notice that $-g > 0$). Therefore, with the choice $\delta = \frac{2+\varepsilon}{6}$, we have

$$\delta X_s^2 - \frac{1}{2} X_s V'(X_s) \leq 0. \tag{2.4.102}$$

Observing that $\|X_s \otimes 1_{\mathcal{A}} + 1_{\mathcal{A}} \otimes X_s\| \leq 2\|X_s\| \leq 2M$ for $s < T$, we deduce by using free Burkholder-Davis-Gundy inequality that

$$e^{\delta t} \|X_t^2\| \leq \|X_0^2\| + 4\sqrt{2} M \sqrt{\frac{e^{2\delta t} - 1}{2\delta}} + \frac{e^{\delta t} - 1}{\delta}. \tag{2.4.103}$$

Consequently, for all $t < T$

$$\|X_t\|^2 \leq m^2 + 4M \sqrt{\frac{6}{2+\varepsilon}} + \frac{6}{2+\varepsilon} = m^2 + 4\sqrt{6} \sqrt{\frac{1-\varepsilon}{2+\varepsilon}} h + \frac{6}{2+\varepsilon}. \tag{2.4.104}$$

If the parameters m, h and ε satisfy

$$m^2 + 4\sqrt{6} \sqrt{\frac{1-\varepsilon}{2+\varepsilon}} h + \frac{6}{2+\varepsilon} < M = (1-\varepsilon) h^2, \tag{2.4.105}$$

then we will have, by the continuity of the norm of the free stochastic integral $(X_t)_{t\geq 0}$, that $T = \infty$ and the support of the distribution of X_t will be bounded.

Condition (2.4.105) is satisfied for m and $-g$ small enough. Indeed, we have that

$$(1-\varepsilon)h^2 - 4\sqrt{6}\sqrt{\frac{1-\varepsilon}{2+\varepsilon}}h - \frac{6}{2+\varepsilon} > 0 \tag{2.4.106}$$

as soon as

$$h > \frac{2\sqrt{6}+\sqrt{30}}{\sqrt{(1-\varepsilon)(2+\varepsilon)}}. \tag{2.4.107}$$

We are interested in finding the smallest g^* such that we can define solutions if the initial support is included in a small neighborhood of the origin. To achieve this, we minimize the right hand-side of (2.4.107) by taking $\varepsilon = 0$. Therefore, if

$$-\frac{1}{81+36\sqrt{5}} < g < 0 \tag{2.4.108}$$

and the initial support satisfies

$$m^2 < h^2 - 4\sqrt{3}h - 3 = -\frac{1}{3g} - \frac{4}{\sqrt{-g}} - 3, \tag{2.4.109}$$

then $T = \infty$ and for all $t \geq 0$

$$\|X_t\|^2 \leq m^2 + \frac{4}{\sqrt{-g}} + 3. \tag{2.4.110}$$

This proves (2.4.97). Finally, we have for all $t \geq 0$

$$\inf_{x\in supp(\mu_t)} V''(x) = 1 + 3g\left(m^2 + \frac{4}{\sqrt{-g}} + 3\right) = -3g\left(h^2 - m^2 - \frac{4}{\sqrt{-g}} - 3\right) > 0. \tag{2.4.111}$$

This proves (2.4.98). □

An immediate consequence is the following lemma.

Lemma 2.4.11. *Let* $g \in \left(-\frac{1}{81+36\sqrt{5}}, 0\right)$ *and* $m < m(g)$. *The distributions* $(\mu_t)_{t\geq 0}$ *asso-*

ciated with the process $(X_t)_{t\geq 0}$ have a lower-bounded entropy

$$\inf_{t\geq 0} \Sigma(\mu_t) > -\infty. \tag{2.4.112}$$

Moreover, the entropy dissipation is still given by D

$$\frac{d}{dt}\Sigma(\mu_t) = -D(\mu_t). \tag{2.4.113}$$

Proof. The non-confining potential V (2.4.88) does not satisfy the growth assumption on the whole real line (2.4.5) of Theorem 3.1 and Proposition 6.1 [12]. However, the support of the solution being bounded, it can be still satisfied locally. More precisely, there exists a modified potential $\overline{V}$ satisfying

$$\begin{cases} \overline{V} \in C^2(\mathbb{R}) \\ \overline{V}(x) = V(x), \text{ for } x \in [-M-1, M+1] \\ \overline{V}(x) \geq x^4, \text{ for } x \in [-M-2, M+2]. \end{cases} \tag{2.4.114}$$

$\overline{V}$ will therefore satisfy the growth assumption of Theorem 3.1 [12]. $(\mu_t)_{t\geq 0}$ satisfies the Fokker-Planck equation

$$\partial_t \mu_t = \partial_x \left[\mu_t \left(\frac{1}{2}\overline{V}' - H\mu_t\right)\right]. \tag{2.4.115}$$

The entropy and dissipation of μ_t, $t \geq 0$ are given respectively by

$$\begin{aligned} \Sigma(\mu_t) &= \frac{1}{2}\int_{\mathbb{R}} V(x)\mu_t(dx) - \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} \log|x-y|\mu_t(dx)\mu_t(dy) \\ &= \frac{1}{2}\int_{\mathbb{R}} \overline{V}(x)\mu_t(dx) - \frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} \log|x-y|\mu_t(dx)\mu_t(dy) \end{aligned} \tag{2.4.116}$$

and

$$D(\mu_t) = \int_{\mathbb{R}} \left|\frac{1}{2}\overline{V}'(x) - H\mu(x)\right|^2 \mu_t(dx). \tag{2.4.117}$$

The result follows from Theorem 3.1 and Proposition 6.1 [12] applied to $\overline{V}$. □

We are now ready to prove Theorem 2.0.11.

Proof of Theorem 2.0.11. The measures $(\mu_t)_{t\geq 0}$ have uniformly bounded moments, thanks to the boundedness of the support. Therefore, we have tightness with respect to the Wasserstein distance. Let μ_∞ be a limit point. According to Lemma 2.4.11, this limit point is a stationary solution.

Set $R = \sqrt{m^2 + \frac{4}{\sqrt{-g}} + 3}$. Let $\rho_0, \rho_1 \in \mathcal{M}_2$ with support included in $[-R, R]$. Let T be the optimal transportation map from ρ_0 to ρ_1. We see that for all $s \in [0,1]$

$$(1-s)x + sT(x) \in [-R, R], \ \forall x \in supp(\rho_0). \tag{2.4.118}$$

Using the fact $\inf_{|x|\leq R} V''(x) \geq 2\lambda > 0$, we deduce

$$\frac{1}{2}\int V''((1-s)x + sT(x))(T(x)-x)^2\rho_0(dx) \geq \lambda W_2(\rho_0,\rho_1)^2, \ \forall s \in [0,1]. \tag{2.4.119}$$

Following a similar argument to the Remark 2.4.3, we derive the HWI inequality

$$\Sigma(\rho_0) - \Sigma(\rho_1) \leq \sqrt{D(\rho_0)}W_2(\rho_0,\rho_1) - \frac{\lambda}{2}W_2(\rho_0,\rho_1)^2. \tag{2.4.120}$$

Similarly to Corollary 2.2.1, we deduce exponential stability. To see that μ_∞ is a local minimizer, take $\rho_0 = \mu_\infty$ and $\rho_1 = \mu$ with support included in $[-R, R]$. μ_∞ being a stationary measure, the HWI inequality gives

$$\Sigma(\mu_\infty) - \Sigma(\mu) \leq -\frac{\lambda}{2}W_2(\mu_\infty,\mu)^2 \leq 0. \tag{2.4.121}$$

□

Remark 2.4.12. *This result can be slightly improved by allowing $g < g^*$. Indeed, we did not exploit the logarithmic energy as we did in Theorem 2.0.9. More precisely, we do not need $\inf_{x\in supp(\mu_t)} V''(x) > 0$ for all $t \geq 0$ to deduce (2.4.120). For solutions with a fixed* *center of mass and with a symmetric initial data, it is enough to ensure that*

$$\inf_{x \in supp(\mu_t)} V''(x) + \frac{1}{4 \max_{x \in supp(\mu_t)} |x|^2} > 0, \ \forall t \geq 0. \tag{2.4.122}$$

Appendix

2.A Proof of Lemma 2.4.4

Proof of Lemma 2.4.4. Let $\phi \in \mathcal{C}^\infty(\mathbb{R}, [0,1])$ such that

1. $\phi(0) = 1$ and $\phi(1) = 0$,

2. $\phi^{(k)}(0) = \phi^{(k)}(1) = 0$ for $k \geq 1$.

Set $|\phi^{(k)}|_\infty = \sup_{x\in[0,1]} |\phi^{(k)}(x)|$. Define for $K \geq 1$:

$$\eta_K(x) = \begin{cases} 1 & \text{if } |x| \leq K, \\ \phi(|x| - K) & \text{if } K \leq |x| \leq K+1, \\ 0 & \text{if } |x| > K+1. \end{cases}$$

We check easily that $\eta_K \in \mathcal{C}^\infty(\mathbb{R}, [0,1])$ and that for all $k \geq 1$,

$$|\eta_K^{(k)}|_\infty \leq C_k |\phi^{(k)}|_\infty$$

for some $C_k > 0$.

Setting $f : x \mapsto |x|^p$ and taking $f_K : x \mapsto \eta_K(x) f(x)$ in (2.4.2), gives

$$\underbrace{\int_{\mathbb{R}} f_K(x)\mu_t(dx) - \int_{\mathbb{R}} f_K(x)\mu_0(dx)}_{\equiv A(K)} = \frac{1}{2}\int_0^t \underbrace{\iint_{\mathbb{R}\times\mathbb{R}} \frac{f_K'(x) - f_K'(y)}{x-y}\mu_s(dx)\mu_s(dy)}_{\equiv B(s,K)}\, ds$$
$$- \int_0^t \underbrace{\int_{\mathbb{R}} V'(x) f_K'(x)\mu_s(dx)}_{\equiv C(s,K)}\, ds. \tag{2.A.1}$$

Using that $|f_K| \leq |f|$ and the dominated convergence theorem, we see that the left-hand side $A(K)$ converges towards $\int_{\mathbb{R}} |x|^p \mu_t(dx) - \int_{\mathbb{R}} |x|^p \mu_0(dx)$ when $K \to +\infty$.

Likewise, $C(s,K) = \int_{\mathbb{R}} V'(x) \cdot [|x|^p \eta_K'(x) + p|x|^{p-2}\eta_K(x)x]\mu_s(dx)$ converges towards

$$C(s,\infty) = \int_{\mathbb{R}} pV'(x) \cdot x|x|^{p-2}\mu_s(dx).$$

Moreover

$$C(s,K) \leq \max(|\phi'|_\infty, |\phi|_\infty) \int_{\mathbb{R}} |V'(x)|(|x|^p + p|x|^{p-1})\mu_s(dx) \in L^1([0,t]).$$

According to the dominated convergence theorem

$$\int_0^t C(s,K)ds \xrightarrow{K\to+\infty} \int_0^t \int_{\mathbb{R}} V'(x) \cdot px|x|^{p-2}\mu_s(dx). \tag{2.A.2}$$

We treat $B(s,K)$ similarly

$$\begin{aligned}\frac{|(\eta_K f)'(x) - (\eta_K f)'(y)) \cdot (x-y)|}{|x-y|^2} \leq & |\phi''|_\infty |f(y)| \\ & + |\phi|_\infty \frac{|f'(x) - f'(y)|}{|x-y|} \\ & + |\phi'|_\infty \left[|f'(y)| + \frac{|f(x)-f(y)|}{|x-y|} \right].\end{aligned} \tag{2.A.3}$$

We check that for $f(x) = |x|^p$ and $\int |x|^p \mu_t(dx) < +\infty$, the right-hand side is integrable. Therefore, $B(s,K)$ converges towards

$$B(s,\infty) = \iint_{\mathbb{R}\times\mathbb{R}} \frac{f'(x)-f'(y)}{x-y}\mu_s(dx)\mu_s(dy).$$

Using (2.A.3), we find an integrable function in $L^1([0,t])$ dominating $\iint_{\mathbb{R}} B(s,K)\mu_s(dx)$, and thereupon

$$\int_0^t B(s,K)ds \xrightarrow{K\to+\infty} \int_0^t \iint_{\mathbb{R}\times\mathbb{R}} \frac{f'(x)-f'(y)}{x-y}\mu_s(dx)\mu_s(dy)ds. \tag{2.A.4}$$

Finally, we conclude from (2.A.1), (2.A.2) and (2.A.4), that equation (2.4.2) is satisfied for $f(x) = |x|^p$.

□

2.B Proof of Lemma 2.4.8

Proof of Lemma 2.4.8. Write

$$\int_{|x|\le r} x^2\rho_s(dx) = \int x^2\rho_s(dx) - \int_{|x|>r} x^2\rho_s(dx). \tag{2.B.1}$$

Treat each term independently. First

$$\begin{aligned}\int x^2\rho_s(dx) &= \int((1-s)x+sT(x))^2\rho_0(dx)\\ &\ge \frac{1}{2}\min\left(\int x^2\rho_0(dx), \int x^2\rho_1(dx)\right) + 2s(1-s)\int xT(x)\rho_0(dx).\end{aligned} \tag{2.B.2}$$

Introduce q the median of ρ_0 satisfying $F_{\rho_0}(q) = \frac{1}{2}$. We see immediately by the assumption on ρ_1 that $T(q) = 0$. Therefore, T being the gradient of a convex function

$$\begin{aligned}\int xT(x)\rho_0(dx) &= \int(x-q)(T(x)-T(q))\rho_0(dx) + q\int T(x)\rho_0(dx)\\ &\ge q\int x\rho_1(dx) = 0.\end{aligned} \tag{2.B.3}$$

Consequently

$$\int x^2 \rho_s(dx) \geq \frac{1}{2} \min \left(\int x^2 \rho_0(dx), \int x^2 \rho_1(dx) \right). \tag{2.B.4}$$

For the second term, write

$$\begin{aligned} \int_{|x|>r} x^2 \rho_s(dx) &\leq \frac{1}{r^2} \int x^4 \rho_s(dx) \\ &\leq \frac{8}{r^2} \int \left(\frac{|x| + |T(x)|}{2} \right)^4 \rho_0(dx) \\ &\leq \frac{8}{r^2} \max \left(\int x^4 \rho_0(dx), \int x^4 \rho_1(dx) \right). \end{aligned} \tag{2.B.5}$$

□

Chapter 3

Well-posedness of the supercooled Stefan problem with oscillatory initial conditions

Consider the one-phase one-dimensional supercooled Stefan problem for the heat equation

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\partial_{xx} u(t,x), \quad x > \Lambda_t, \;\; t > 0, \\ u(0,x) = f(x), \;\; x \geq 0 \quad \text{and} \quad u(t,\Lambda_t) = 0, \;\; t > 0, \\ \dot{\Lambda}_t = \frac{1}{2}\partial_x u(t,x+)|_{x=\Lambda_t}, \;\; t \geq 0, \\ \Lambda_0 = 0 \end{cases} \tag{3.0.1}$$

with a non-negative initial condition f. The unknowns are u, the negative of the temperature of a liquid relative to its equilibrium freezing point, as a function of time and space, and the free boundary Λ, which encodes the location of a liquid-solid frontier over time. The temperature is required to solve the heat equation with Dirichlet-type boundary conditions, while the free boundary moves at a speed proportional to the space derivative of the temperature at said boundary ("Stefan condition"). To ease exposition, we normalize the latent heat coefficient, usually denoted by α, to 1.

It turns out that, for generic initial conditions, the frontier Λ can exhibit jump discontinuities (see, e.g., [58, Theorem 1.1]). A way to circumvent this issue is to restate (3.0.1) in a probabilistic form, which allows the definition of global solutions, even in the presence of jump discontinuities. To wit, let X_{0-} be a non-negative random variable with a density f, and let B be an independent standard Brownian motion. The probabilistic reformulation of (3.0.1), first introduced in [25, 26] for a variant of it, is phrased in terms of the McKean-Vlasov problem

$$\begin{cases} X_t = X_{0-} + B_t - \Lambda_t, \quad t \geq 0, \\ \tau := \inf\{t \geq 0 : \; X_t \leq 0\}, \\ \Lambda_t = \mathbb{P}\left(\tau \leq t\right), \quad t \geq 0, \end{cases} \tag{3.0.2}$$

with the unknowns $X = (X_t)_{t\geq 0}$ and $\Lambda = (\Lambda_t)_{t\geq 0}$. When f belongs to the Sobolev space $W_2^1([0,\infty))$ and $f(0) = 0$, a solution (X, Λ) of (3.0.2) such that $\dot{\Lambda} \in L^2([0,T])$ for some $T \in (0,\infty)$ gives rise to a solution $u \in W_2^{1,2}(\{(t,x) \in [0,T] \times [0,\infty) \colon x \geq \Lambda_t\})$ of (3.0.1) on $[0,T]$ by taking $u(t,x)\,\mathrm{d}x$ as the law of $(X_t + \Lambda_t)\,\mathbf{1}_{\{\tau > t\}}$ on (Λ_t, ∞), for $t \in [0,T]$ (cf. [91, proof of Proposition 4.2(b)]).

The probabilistic formulation (3.0.2) brings out the necessary presence of jump discontinuities in the frontier Λ for certain initial data X_{0-} (for example, those with $\mathbb{E}\left[X_{0-}\right] < 1/2$, see [58, Theorem 1.1]), as well as the non-uniqueness of the jump sizes $X_{t-} - X_t := \lim_{s\uparrow t} X_s - X_t = \Lambda_t - \Lambda_{t-}$ at the instants of discontinuity. When extending solutions beyond a discontinuity, one must decide how to choose the jump size, which has led to the introduction of the condition

$$X_{t-} - X_t = \Lambda_t - \Lambda_{t-} = \inf\big\{x > 0 : \; \mathbb{P}\left(\tau \geq t, \, X_{t-} \in (0,x]\right) < x\big\}, \quad t \geq 0. \tag{3.0.3}$$

Solutions of (3.0.2) satisfying (3.0.3) are called physical. It has been shown that (3.0.3) selects the minimal jump sizes a right-continuous solution Λ with left limits can have (see [58, Proposition 1.2]). The global existence of physical solutions is known under natural assumptions on the initial data, see [24], where it is proved under the very mild assumption $\mathbb{E}[X_{0-}] < \infty$, as well as earlier results in [25], [91], [92]. On the other hand, it has been established in [27] that if X_{0-} possesses a density f on $[0, \infty)$ that is bounded and changes monotonicity finitely often on compacts intervals, then the physical solution is unique.

This paper develops new arguments that demonstrate uniqueness for oscillatory initial data, which in particular do not fulfill the monotonicity change assumption of [27], though densities fulfilling the latter assumption are also captured by our main theorem. Oscillatory initial conditions arise frequently when one investigates continuum limits of interacting particle systems. For example, [29, Remark 1.10] and [75, Theorem 1.2] feature initial conditions given by the trajectories of a Brownian motion and a (reparameterized) Brownian bridge, respectively. We also refer to [78, Theorem 4], [33, Theorem 5.4], [20, Theorem 4] where the initial conditions even are distributions rather than functions in general.

Consider the question of short time uniqueness for (3.0.2)–(3.0.3), assume that X_{0-} has a density f, and let F be the cumulative distribution function (CDF) of X_{0-}. If $\operatorname{ess\,lim\,sup}_{x\downarrow 0} f(x) < 1$, there is no jump discontinuity at time 0 (i.e., $\Lambda_0 = 0 =: \Lambda_{0-}$ and $X_0 = X_{0-}$) for any physical solution, and it is straightforward to prove short time uniqueness. If $\operatorname{ess\,lim\,inf}_{x\downarrow 0} f(x) > 1$, any physical solution must have an initial jump of the size $\Lambda_0 = \inf\{x > 0 : F(x) < x\} > 0$, and one can focus on the problem started from $X_0 = X_{0-} - \Lambda_0$, with the density $f(x + \Lambda_0)$. This new density satisfying necessarily $\operatorname{ess\,lim\,inf}_{x\downarrow 0} f(x) \leq 1$, we infer that ultimately one needs to investigate the case $\operatorname{ess\,lim\,inf}_{x\downarrow 0} f(x) \leq 1 \leq \operatorname{ess\,lim\,sup}_{x\downarrow 0} f(x)$.

In [27, Proposition 5.2], short time uniqueness is shown using a contraction argument, based on the fact that for densities satisfying their monotonicity change assumption, there exists a non-decreasing function $h\colon (0, \infty) \to (0, \infty)$, with $h(0+) = 0$, such that for all

$x > 0$ sufficiently small:

$$f(x) \leq 1 - h(x). \tag{3.0.4}$$

The key contribution of this paper is the proof of short time uniqueness for densities oscillating down from 1, and thus violating (3.0.4). Instead, we introduce an averaging condition: There exists a non-decreasing function $g\colon (0,\infty) \to (0,\infty)$ such that

$$f \leq 1, \qquad \int_0^\infty x\, f(x)\,\mathrm{d}x < \infty, \tag{3.0.5a}$$

$$\exists\, \lambda_0 > 0 \quad \forall\, \lambda \in [0,\lambda_0) \quad \forall\, \mu \in [0,1]: \qquad \int_\mu^{\mu+1} f(\lambda x)\,\mathrm{d}x \leq 1 - g\big(\lambda(\mu+1)\big). \tag{3.0.5b}$$

Notice that (3.0.4) implies

$$\begin{aligned} \int_\mu^{\mu+1} f(\lambda x)\,\mathrm{d}x &\leq 1 - \int_\mu^{\mu+1} h(\lambda x)\,\mathrm{d}x \\ &\leq 1 - \int_{(\mu+1)/2}^{\mu+1} h(\lambda x)\,\mathrm{d}x \\ &\leq 1 - \frac{h(\lambda(\mu+1)/2)}{2}, \end{aligned} \tag{3.0.6}$$

i.e., (3.0.5b) with $g(x) := h(x/2)/2$.

We are now ready to state our main result.

Theorem 3.0.1. *Let $X_{0-} \geq 0$ possess a density f that satisfies condition* (3.0.5) *with a continuous function g. Then, the physical solution (X,Λ) of* (3.0.2) *started from X_{0-} is unique.*

Remark 3.0.2. *Our proof of Theorem 3.0.1 (see Section 3.1) shows that the solution (X,Λ) of* (3.0.2) *started from X_{0-} is locally unique even if one weakens the physicality assumption to $\Lambda_0 = 0$.*

In the second part of the article, we provide evidence that condition (3.0.5) is natural and non-restrictive, by establishing that it is fulfilled by many oscillating densities, like ones given by sample paths of certain stochastic processes.

Corollary 3.0.3. *For almost every fixed sample path of a standard Brownian motion* $(W_x)_{x\geq 0}$*, the physical solution* (X,Λ) *of* (3.0.2) *started from* $X_{0-}\geq 0$ *is unique if* X_{0-} *has a density* f *obeying* (3.0.5a) *and such that*

$$f(x)=\big(1+W_x-\sqrt{2x|\log|\log x||}\big)_+\wedge 1,\quad x\in[0,1]. \tag{3.0.7}$$

We also consider deterministically constructed oscillating densities, including the ones in the next corollary.

Corollary 3.0.4. *For* $\alpha>0$*, let* $X_{0-}\geq 0$ *be a random variable with the density*

$$f(x)=\frac{1}{2}\left(1+\sin\frac{1}{x^\alpha}\right),\quad x\in(0,a], \tag{3.0.8}$$

where $a\in(0,\infty)$ *is defined by* $\int_0^a f(x)\,\mathrm{d}x=1$*. Then, the physical solution* (X,Λ) *of (3.0.2) started from* X_{0-} *is unique.*

Remark 3.0.5. *We note that for the densities* f *of Corollary 3.0.4, the decreasing sequence of solutions to* $f(x)=1$ *approaches* 0 *at an arbitrarily high polynomial rate* $n^{-1/\alpha}$*. Such oscillatory densities are termed "pathological" in [79, Figure 3.1] due to the difficulty of showing uniqueness for them.*

The last part the paper exhibits a situation in which it is possible to go beyond condition (3.0.5) and to establish uniqueness for the supercooled Stefan problem via complementary arguments.

Proposition 3.0.6. *Fix a* $T\in(0,\infty)$*, and let* $X_{0-}\geq 0$ *be a random variable with the density*

$$f(x)=\begin{cases}\alpha_1, & x\in\bigcup_{n\geq 1}[a_{2n},a_{2n-1}),\\[2ex] \alpha_2, & x\in\bigcup_{n\geq 1}[a_{2n+1},a_{2n}),\end{cases} \tag{3.0.9}$$

where $0<\alpha_1<1<\alpha_2$*,* $a_{2n-1}=r^{n-1}a_1$*,* $a_{2n}=pr^{n-1}a_1$*, and* $r=pq$*,* $p,q\in(0,1)$*. Then,* *for any $\alpha_2 > 1$ close enough to 1, the physical solution (X, Λ) of (3.0.2) started from X_{0-} is unique on $[0, T]$.*

Remark 3.0.7. *In contrast to the main theorem (Theorem 3.0.1), Proposition 3.0.6 is a local uniqueness result. In particular, we were unable to verify the monotonicity change assumption of [27] at T.*

The rest of the article is structured as follows. In Section 3.1, we introduce notation and prove Theorem 3.0.1. In Subsection 3.2.1, we verify, using functional local laws of the iterated logarithm, that condition (3.0.5) is satisfied by many densities obtained from sample paths of suitable stochastic
processes. In Subsection 3.2.2, we consider oscillating densities constructed from periodic functions. In particular, we deduce Corollaries 3.0.3, 3.0.4 from Theorem 3.0.1 in Subsections 3.2.1, 3.2.2, respectively. Finally, Section 3.3 is devoted to showing Proposition 3.0.6.

This chapter is based on [90].

3.1 Proof of Theorem 3.0.1

Throughout the section, f denotes a density as in Theorem 3.0.1, and we write F for the associated CDF. We also define the continuous strictly increasing function

$$\widetilde{g}:\, [0,\infty) \to [0,\infty), \quad x \mapsto x\, g(x) \tag{3.1.1}$$

and set

$$\psi(\lambda, \mu) = \int_{\mu}^{\mu+1} f(\lambda x)\, \mathrm{d}x, \quad \lambda, \mu \geq 0. \tag{3.1.2}$$

Let (X, Λ) be an arbitrary physical solution of (3.0.2). By [24, Proposition 2.3], there exists a minimal solution $(\underline{X}, \underline{\Lambda})$ of (3.0.2), namely the unique solution of (3.0.2) satisfying

$$\underline{\Lambda}_t \leq \widetilde{\Lambda}_t, \quad t \geq 0, \tag{3.1.3}$$

for any solution $(\widetilde{X}, \widetilde{\Lambda})$ of (3.0.2). The physicality of $(\underline{X}, \underline{\Lambda})$ is ensured by [24, Theorem 6.5]. We further introduce $(Y_t)_{t\geq 0}$, $(Z_t)_{t\geq 0}$ given respectively by

$$Y_t = \sup_{0\leq s\leq t}(-B_s + \underline{\Lambda}_s), \quad t \geq 0, \tag{3.1.4}$$
$$Z_t = \sup_{0\leq s\leq t}(-B_s + \Lambda_s), \quad t \geq 0. \tag{3.1.5}$$

In these terms, the frontiers solve

$$\underline{\Lambda}_t = \mathbb{P}\big(\inf_{0\leq s\leq t}(X_{0-} + B_s - \underline{\Lambda}_s) \leq 0\big) = \mathbb{E}\left[F(Y_t)\right], \quad t \geq 0, \tag{3.1.6}$$
$$\Lambda_t = \mathbb{E}\left[F(Z_t)\right], \quad t \geq 0. \tag{3.1.7}$$

Our starting point is the following continuous upper bound on the frontier Λ.

Lemma 3.1.1. *There exist a $T > 0$ and a strictly increasing continuous function $(\overline{\chi}_t)_{t\geq 0}$, with $\overline{\chi}_0 = 0$, such that*

$$\Lambda_t \leq \overline{\chi}_t, \quad t \in [0, T]. \tag{3.1.8}$$

Proof. For $t \geq 0$, we estimate

$$\begin{aligned}\Lambda_t &= \mathbb{P}\big(X_{0-} \leq \sup_{0\leq s\leq t}(-B_s + \Lambda_s)\big) \\ &\leq \mathbb{P}\left(X_{0-} \leq \Lambda_t\right) + \mathbb{P}\Big(\{\Lambda_t < X_{0-}\} \cap \big\{X_{0-} \leq \sup_{0\leq s\leq t}(-B_s + \Lambda_s)\big\}\Big).\end{aligned} \tag{3.1.9}$$

In view of the upper bound $\sup_{0\leq s\leq t}(-B_s + \Lambda_s) \leq \sup_{0\leq s\leq t}(-B_s) + \Lambda_t$, we find for all $t \geq 0$ that

$$\begin{aligned}\Lambda_t - F(\Lambda_t) &\leq \mathbb{P}\big(\{\Lambda_t < X_{0-}\} \cap \{X_{0-} - \Lambda_t \leq \sqrt{t}\,|\mathcal{N}|\}\big) \\ &= \int_0^\infty \mathbb{P}(x \leq \sqrt{t}\,|\mathcal{N}|)\,\mathbb{P}(X_{0-} - \Lambda_t \in \mathrm{d}x) \\ &\leq \sqrt{2t/\pi},\end{aligned} \tag{3.1.10}$$

where $\mathcal{N}$ is a standard normal random variable, and we have used $f \leq 1$ and $\mathbb{E}\left[|\mathcal{N}|\right] =$

$\sqrt{2/\pi}$.

To conclude we apply (3.0.5b) to obtain

$$\frac{F(\lambda)}{\lambda} = \int_0^1 f(\lambda x)\,\mathrm{d}x \le 1 - g(\lambda), \quad \lambda \in (0, \lambda_0). \tag{3.1.11}$$

Since Λ_t is right-continuous with $\Lambda_0 = 0$, there exists a $T > 0$ such that $\Lambda_t < \lambda_0$, $t \in [0, T]$. Putting this together with (3.1.11) and (3.1.10) we get

$$\widetilde{g}(\Lambda_t) \le \Lambda_t - F(\Lambda_t) \le \mathbb{E}\left[|\mathcal{N}|\right]\sqrt{t}, \quad t \in [0, T]. \tag{3.1.12}$$

The proof is completed by inferring

$$\Lambda_t \le \widetilde{g}^{-1}(\mathbb{E}\left[|\mathcal{N}|\right]\sqrt{t}) =: \overline{\chi}_t, \quad t \in [0, T], \tag{3.1.13}$$

with the continuous $\widetilde{g}^{-1}$ satisfying $\widetilde{g}^{-1}(0) = 0$. □

We also need the next lemma.

Lemma 3.1.2. *Let $(\nu_t)_{t\ge 0}$ be a strictly increasing continuous function, with $\nu_0 = 0$. Then, there exists a positive function $\varphi(t, b)$ of $t > 0$ and $b > 0$ so that*

$$P(t, b) := \mathbb{P}\Big(\sup_{0\le s\le t}(-B_s + \nu_s) \le b\Big) \ge \varphi(t, b). \tag{3.1.14}$$

Proof. Fix $t > 0$ and $b > 0$. If $\nu_t \le b/2$, then

$$P(t, b) \ge \mathbb{P}\Big(\sup_{0\le s\le t}(-B_s) \le b - \nu_t\Big). \tag{3.1.15}$$

Otherwise, $\nu_t > b/2$ and $\tau := \tau(b) := \nu^{-1}(b/2) < t$. Moreover, for any $\tau' \in (0, \tau]$,

$$\begin{aligned} P(t, b) &= \mathbb{P}\Big(\big\{\sup_{0\le s\le \tau'}(-B_s + \nu_s) \le b\big\} \cap \big\{\sup_{\tau' < s\le t}(-B_s + \nu_s) \le b\big\}\Big) \\ &\ge \mathbb{P}\left(\left\{\sup_{0\le s\le \tau'}(-B_s) \le \frac{b}{2}\right\} \cap \big\{\sup_{\tau' < s\le t}(-B_s) \le b - \nu_t\big\}\right). \end{aligned} \tag{3.1.16}$$

We conclude by setting

$$\varphi(t,b) = \mathbb{P}\Big(\Big\{\sup_{0\le s\le \tau(b)\wedge(\nu_t-b/2)_+}(-B_s) \le \frac{b}{2}\Big\} \cap \big\{\sup_{\tau(b)\wedge(\nu_t-b/2)_+ < s\le t}(-B_s) \le b-\nu_t\big\}\Big). \tag{3.1.17}$$

Clearly, φ is positive on $(0,\infty)^2$. □

The following proposition is the key ingredient in our proof of Theorem 3.0.1.

Proposition 3.1.3. *There exists a function* $\Phi\colon (0,T]\times(0,\lambda_0)\to(0,\infty)$ *such that*

$$\mathbb{E}\left[F(Y_t+\lambda)-F(Y_t)\right] \le (1-\Phi(t,\lambda))\,\lambda, \quad (t,\lambda)\in(0,T]\times(0,\lambda_0). \tag{3.1.18}$$

Proof. Let $(t,\lambda)\in(0,T]\times(0,\lambda_0)$. Then,

$$\begin{aligned}\mathbb{E}\left[F(Y_t+\lambda)-F(Y_t)\right] &= \mathbb{E}\left[F\left(\lambda\left(\frac{Y_t}{\lambda}+1\right)\right)-F\left(\lambda\frac{Y_t}{\lambda}\right)\right]\\ &= \mathbb{E}\left[\psi\left(\lambda,\frac{Y_t}{\lambda}\right)\right]\lambda.\end{aligned} \tag{3.1.19}$$

Since $\lambda\in(0,\lambda_0)$, condition (3.0.5b) yields

$$\mathbf{1}_{\{Y_t\le\lambda\}}\,\psi\left(\lambda,\frac{Y_t}{\lambda}\right) \le \mathbf{1}_{\{Y_t\le\lambda\}}\,(1-g(Y_t+\lambda)). \tag{3.1.20}$$

In view of $\psi\le\sup_{x\ge0}f(x)\le 1$,

$$\begin{aligned}\frac{\mathbb{E}\left[F(Y_t+\lambda)-F(Y_t)\right]}{\lambda} &\le \mathbb{E}\left[\mathbf{1}_{\{Y_t\le\lambda\}}\,\psi\left(\lambda,\frac{Y_t}{\lambda}\right)\right]+\mathbb{P}\left(Y_t>\lambda\right)\\ &\le \mathbb{E}\left[\mathbf{1}_{\{Y_t\le\lambda\}}(1-g(Y_t+\lambda))\right]+\mathbb{P}\left(Y_t>\lambda\right)\\ &= 1-\mathbb{E}\left[\mathbf{1}_{\{Y_t\le\lambda\}}\,g(Y_t+\lambda)\right].\end{aligned} \tag{3.1.21}$$

Finally, we use $\sup_{0\le s\le t}(-B_s)\le Y_t=\sup_{0\le s\le t}(-B_s+\underline{\Lambda}_s)$ and $\underline{\Lambda}_t\le\overline{\chi}_t$, $t\in[0,T]$:

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\{Y_t\le\lambda\}}\,g(Y_t+\lambda)\right] &\ge g(\lambda)\,\mathbb{P}(Y_t\le\lambda)\\ &\ge g(\lambda)\,\mathbb{P}\big(\sup_{0\le s\le t}(-B_s+\overline{\chi}_s)\le\lambda\big).\end{aligned} \tag{3.1.22}$$

Thus, thanks to Lemma 3.1.2,

$$\mathbb{E}\left[\mathbf{1}_{\{Y_t \le \lambda\}}\, g(Y_t+\lambda)\right] \ge g(\lambda)\, \varphi(t,\lambda) =: \Phi(t,\lambda). \tag{3.1.23}$$

Inserting this into (3.1.21) we obtain (3.1.18). □

We are now ready for the proof of Theorem 3.0.1.

Proof of Theorem 3.0.1. To start, we fix a $\lambda \in (0,\lambda_0)$ and decrease $T > 0$ to ensure $\Lambda_t - \underline{\Lambda}_t \le \lambda$, $t \in [0,T]$, relying on right-continuity. For a $\lambda' \in (0,\lambda]$, suppose $\{t \in [0,T] : \Lambda_t - \underline{\Lambda}_t \ge \lambda'\} \neq \emptyset$ and consider $t_{\lambda'} := \inf\{t \in [0,T] : \Lambda_t - \underline{\Lambda}_t \ge \lambda'\}$. Then, $t_{\lambda'} > 0$ by right-continuity, and

$$0 < \lambda' \le \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}} = \sup_{0 \le t \le t_{\lambda'}} (\Lambda_t - \underline{\Lambda}_t) \le \lambda < \lambda_0. \tag{3.1.24}$$

Therefore, we have

$$Z_{t_{\lambda'}} = \sup_{0 \le s \le t_{\lambda'}} (-B_s + \Lambda_s) \le Y_{t_{\lambda'}} + \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}}. \tag{3.1.25}$$

Thus, combining (3.1.6), (3.1.7) and Proposition 3.1.3 we infer

$$\begin{aligned} \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}} &= \mathbb{E}[F(Z_{t_{\lambda'}}) - F(Y_{t_{\lambda'}})] \\ &\le \mathbb{E}[F(Y_{t_{\lambda'}} + \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}}) - F(Y_{t_{\lambda'}})] \\ &\le (1 - \Phi(t_{\lambda'}, \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}})) \cdot (\Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}}), \end{aligned} \tag{3.1.26}$$

where $\Phi(t_{\lambda'}, \Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}}) > 0$. Hence, $\Lambda_{t_{\lambda'}} - \underline{\Lambda}_{t_{\lambda'}} = 0$, contradicting (3.1.24). We readily conclude that $\{t \in [0,T] : \Lambda_t - \underline{\Lambda}_t \ge \lambda'\} = \emptyset$, and since $\lambda' \in (0,\lambda]$ was arbitrary, $\Lambda_t \le \underline{\Lambda}_t$, $t \in [0,T]$. Due to the minimality of $\underline{\Lambda}$, it must hold $\Lambda_t = \underline{\Lambda}_t$, $t \in [0,T]$.

To derive global uniqueness, we let

$$T' := \inf\{t \ge T :\, \Lambda_t \neq \underline{\Lambda}_t\} \in [T,\infty] \tag{3.1.27}$$

and suppose that $T' < \infty$. By the definition of T',

$$X_{T'-} = X_{0-} + B_{T'} - \Lambda_{T'-} = X_{0-} + B_{T'} - \underline{\Lambda}_{T'-} = \underline{X}_{T'-} \tag{3.1.28}$$

and $\mathbf{1}_{\{\tau \geq T'\}} = \mathbf{1}_{\{\underline{\tau} \geq T'\}}$, so that $\mathbf{1}_{\{\tau \geq T'\}} X_{T'-} = \mathbf{1}_{\{\underline{\tau} \geq T'\}} \underline{X}_{T'-}$. Moreover, for all $0 < a < b$,

$$\mathbb{P}\big(\mathbf{1}_{\{\tau \geq T'\}} X_{T'-} \in [a,b]\big) = \mathbb{P}\big(\tau \geq T',\, X_{T'-} \in [a,b]\big) \leq \mathbb{P}(X_{T'-} \in [a,b]). \tag{3.1.29}$$

Thus, the right essential limit superior of the density of $\mathbf{1}_{\{\tau \geq T'\}} X_{T'-} = \mathbf{1}_{\{\underline{\tau} \geq T'\}} \underline{X}_{T'-}$ at 0 is at most that of $X_{T'-} = \underline{X}_{T'-}$, namely $\mathbb{E}[f(-B_{T'} + \Lambda_{T'-})] = \mathbb{E}[f(-B_{T'} + \underline{\Lambda}_{T'-})]$. Since $f \leq 1$, and $f \equiv 0$ on $(-\infty, 0)$,

$$\mathbb{E}[f(-B_{T'} + \Lambda_{T'-})] = \mathbb{E}[f(-B_{T'} + \underline{\Lambda}_{T'-})] < 1. \tag{3.1.30}$$

Consequently, the condition (3.0.4) is satisfied at $T'-$ and we get $\Lambda \equiv \underline{\Lambda}$ on a non-trivial interval $[T', T'+s]$ by repeating [27, proof of Proposition 5.2]. (Note that the condition (3.0.4) permits us to apply [27, Lemma 5.1].) This is the desired contradiction. □

3.2 Analysis of specific oscillatory initial conditions

3.2.1 Initial conditions constructed from stochastic processes

In the present subsection we illustrate Theorem 3.0.1 on initial conditions obtained from sample paths of stochastic processes. Concretely, we consider initial densities

$$f(x) = \begin{cases} (1 + S_x - \kappa_x)_+ \wedge 1, & x \in [0,1], \\ f_0(x), & x > 1, \end{cases} \tag{3.2.1}$$

where $(S_x)_{x\geq 0}$ is a stochastic process starting at zero, $(\kappa_x)_{x\geq 0}$ is a function with $\kappa_0 = 0$, and the (random) extension $f_0 : (1,\infty) \to [0,1]$ ensures that $\int_0^\infty f(x)\,\mathrm{d}x = 1$ and

$\int_0^\infty x\, f(x)\, \mathrm{d}x < \infty$.

Our interest lies in processes S and functions κ such that, almost surely, f violates the local monotonicity condition (3.0.4) but satisfies condition (3.0.5). Clearly, condition (3.0.4) is violated if $S_x \geq \kappa_x$ for a sequence of x's converging to 0, that is,

$$\limsup_{x \downarrow 0} \frac{S_x}{\kappa_x} \geq 1. \tag{3.2.2}$$

As a guiding example, take S to be a standard Brownian motion and $\kappa_x = \sqrt{2x|\log|\log x||}$. Due to Chung's law of the iterated logarithm (LIL), the resulting f violates condition (3.0.4) almost surely. On the other hand, using the local Strassen's LIL of [105], [37] we prove below that condition (3.0.5) is satisfied almost surely. This result extends to other centered continuous Gaussian processes admitting a local functional LIL as follows.

Let $(S_x)_{x\in[0,1]}$ be a centered continuous Gaussian process with $S_0 = 0$ and a covariance function $\Gamma(x, y) = \mathbb{E}\,[S_x S_y]$ continuous on $[0, 1]^2$ and non-degenerate on $(0, 1]^2$. We write $H(\Gamma)$ for the reproducing kernel Hilbert space associated with Γ. Recall that $H(\Gamma)$ is defined as the completion of the space of finite linear combinations of $\{\Gamma(x, \cdot)\}_{x\in[0,1]}$ under the norm induced by the inner product $\langle \Gamma(x, \cdot), \Gamma(y, \cdot)\rangle := \Gamma(x, y)$. Elements $\phi \in H(\Gamma)$ obey $\phi(x) = \langle \phi, \Gamma(x, \cdot)\rangle$ and are continuous functions. Therefore, $H(\Gamma)$ is a subset of the Banach space $C([0, 1])$. Moreover, the unit ball

$$K = \{\phi \in H(\Gamma) : \langle \phi, \phi\rangle \leq 1\} \tag{3.2.3}$$

is compact in $C([0, 1])$ (see, e.g., [94, Lemma 3]). For technical reasons, we assume throughout that the process S has the scaling property

$$(S_{rx})_{x\in[0,1]} \stackrel{d}{=} (\sqrt{r}^{\,\alpha_2} S_x)_{x\in[0,1]}, \quad r \in (0, 1], \tag{3.2.4}$$

for some $\alpha_2 > 0$. Under (3.2.4), there exists an $\alpha_1 > 0$ such that

$$\gamma(x) \coloneqq \Gamma(x,x) = \alpha_1 x^{\alpha_2}, \quad x \in [0,1]. \tag{3.2.5}$$

For simplicity, we take $\alpha_1 = 1$, i.e., $\mathbb{E}\,[S_1^2] = 1$.

We say that S satisfies a local functional LIL if the following assertion holds for a $\beta \in (0,\infty)$.

Assertion. Almost surely, the set

$$\{(\xi_x^r)_{x\in[0,1]}\}_{r\in(0,1]} \coloneqq \left\{ \left(\frac{S_{rx}}{\beta\sqrt{\gamma(r)\,|\log|\log r||}} \right)_{x\in[0,1]} \right\}_{r\in(0,1]} \tag{3.2.6}$$

is relatively compact in $C([0,1])$, and the set of its limit points as $r \downarrow 0$ is given by K. In particular, for every continuous functional $\mathcal{I}\colon C([0,1]) \to \mathbb{R}$ we have

$$\limsup_{r\downarrow 0} \mathcal{I}(\xi^r) = \sup_{\phi\in K} \mathcal{I}(\phi) \quad \text{almost surely.} \tag{3.2.7}$$

A local functional LIL has been established in the case of a fractional Brownian motion with Hurst exponent $H \in (0,1)$, for which $\Gamma(x,y) = \frac{1}{2}(x^{2H} + y^{2H} - |y-x|^{2H})$ and $\gamma(x) = x^{2H}$ (see [85, Example 4.35]). Further, by taking $\mathcal{I}(\phi) = \phi_1$ one derives the usual LIL, so that the density f in (3.2.1), with $\kappa_x \coloneqq \beta\sqrt{\gamma(x)\,|\log|\log x||}$, violates condition (3.0.4) almost surely. On the other hand, we obtain the next proposition, by observing that $\psi(\lambda,\mu)$ can be estimated in terms of ξ^λ.

Proposition 3.2.1. *Suppose that S satisfies a local functional LIL. Then, almost surely, the density f in* (3.2.1)*, with $\kappa_x \coloneqq \beta\sqrt{\gamma(x)\,|\log|\log x||}$, adheres to condition* (3.0.5)*.*

We start the proof of Proposition 3.2.1 with a technical lemma.

Lemma 3.2.2. *In the context of Proposition 3.2.1*

(i) $\lim_{\lambda\downarrow 0} \sup_{\mu\in[0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \left| \frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)}\sqrt{\gamma(x)}} - \sqrt{\frac{2^{\alpha_2}}{(\mu+1)^{\alpha_2}}} \right|^2 \mathrm{d}x = 0.$

(ii) There exists an $\eta > 0$ such that

$$\limsup_{\lambda \downarrow 0} \sup_{\mu \in [0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\xi_x^\lambda| - \sqrt{\gamma(x)}\, \mathrm{d}x \leq -\eta. \tag{3.2.8}$$

(iii) $\limsup_{\lambda \downarrow 0} \int_0^1 \left| |\xi_x^\lambda| - \sqrt{\gamma(x)} \right|^2 \mathrm{d}x \leq 2.$

Proof of Lemma 3.2.2. For all small enough $\lambda > 0$, all $\mu \in [0,1]$, and all $x \in lI01$,

$$\begin{aligned} \left| \frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)} \sqrt{\gamma(x)}} - \sqrt{\frac{2^{\alpha_2}}{(\mu+1)^{\alpha_2}}} \right| &= \sqrt{\frac{2^{\alpha_2}}{(\mu+1)^{\alpha_2}}} \left| \left| 1 + \frac{\log \left| \frac{\log 2\lambda x}{\log \lambda(\mu+1)} \right|}{\log |\log \lambda(\mu+1)|} \right|^{1/2} - 1 \right| \\ &\leq \sqrt{2^{\alpha_2}} \left| \frac{\log \left| \frac{\log 2\lambda x}{\log \lambda(\mu+1)} \right|}{\log |\log \lambda(\mu+1)|} \right|^{1/2}, \end{aligned} \tag{3.2.9}$$

where we used that $|\sqrt{|1+a|} - 1| \leq \sqrt{|a|}$ for all $a \in \mathbb{R}$. If further $x \leq \frac{\mu+1}{2}$, then $\frac{\log 2x/(\mu+1)}{\log \lambda(\mu+1)} \geq 0$, and therefore

$$\begin{aligned} \left| \log \left| \frac{\log 2\lambda x}{\log \lambda(\mu+1)} \right| \right| &= \left| \log \left| 1 + \frac{\log 2x/(\mu+1)}{\log \lambda(\mu+1)} \right| \right| \\ &= \log \left(1 + \left| \frac{\log 2x/(\mu+1)}{\log \lambda(\mu+1)} \right| \right) \\ &\leq \left| \frac{\log 2x/(\mu+1)}{\log \lambda(\mu+1)} \right|. \end{aligned} \tag{3.2.10}$$

Thus, we deduce

$$\left| \frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)} \sqrt{\gamma(x)}} - \sqrt{\frac{2^{\alpha_2}}{(\mu+1)^{\alpha_2}}} \right| \leq \sqrt{2^{\alpha_2}} \frac{(|\log 2x| + |\log (\mu+1)|)^{1/2}}{|\log \lambda(\mu+1)|^{1/2} |\log |\log \lambda(\mu+1)||^{1/2}}. \tag{3.2.11}$$

Result (i) follows immediately.

The functional

$$\mathcal{I}(\phi) := \sup_{\mu \in [0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\phi(x)| - \sqrt{\gamma(x)}\, \mathrm{d}x \tag{3.2.12}$$

on $C([0,1])$ is continuous, so that the local functional LIL implies

$$\limsup_{\lambda \downarrow 0} \sup_{\mu \in [0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\xi_x^\lambda| - \sqrt{\gamma(x)}\, \mathrm{d}x = \sup_{\phi \in K} \sup_{\mu \in [0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\phi(x)| - \sqrt{\gamma(x)}\, \mathrm{d}x. \tag{3.2.13}$$

Notice that for all $\phi \in K$,

$$|\phi(x)| = |\langle \phi, \Gamma(x, \cdot)\rangle| \leq \langle \phi, \phi\rangle^{1/2} \cdot \langle \Gamma(x, \cdot), \Gamma(x, \cdot)\rangle^{1/2} \leq \sqrt{\gamma(x)}, \quad x \in [0, 1], \tag{3.2.14}$$

thanks to $\langle \Gamma(x, \cdot), \Gamma(x, \cdot)\rangle = \Gamma(x, x) = \gamma(x)$. Moreover, it is enough to prove that

$$\sup_{\phi \in K} \sup_{a \in [0,1/2]} \int_a^{a+1/2} |\phi(x)| - \sqrt{\gamma(x)}\, \mathrm{d}x < 0. \tag{3.2.15}$$

If the supremum in (3.2.15) was zero, then the continuity of the underlying functional on the compact $K \times [0, 1/2]$ would yield the existence of some $\phi \in K$ and some $a \in [0, 1/2]$ such that

$$\int_a^{a+1/2} |\phi(x)| - \sqrt{\gamma(x)}\, \mathrm{d}x = 0, \tag{3.2.16}$$

and thus the Cauchy-Schwarz inequalities in (3.2.14) would hold with equality for Lebesgue almost every $x \in [a, a + 1/2]$. As a consequence, $\{\Gamma(x, \cdot)\}_x$ would be pairwise linearly dependent for these x, in contradiction to the assumed non-degeneracy of Γ. This proves (ii).

To obtain (iii) we apply the local functional LIL to the continuous functional

$$C([0, 1]) \to \mathbb{R}, \quad \phi \mapsto \int_0^1 \left||\phi(x)| - \sqrt{\gamma(x)}\right|^2 \mathrm{d}x, \tag{3.2.17}$$

and use $|\phi(x)| \leq \sqrt{\gamma(x)}$, $x \in [0, 1]$ for $\phi \in K$ to easily get

$$\int_0^1 \left||\phi(x)| - \sqrt{\gamma(x)}\right|^2 \mathrm{d}x \leq 2 \sup_{x \in [0,1]} \gamma(x) = 2 \tag{3.2.18}$$

for all those ϕ. □

We are now ready for the proof of Proposition 3.2.1.

Proof of Proposition 3.2.1. We only need to show (3.0.5b). Throughout the proof we take $\lambda_0 > 0$ to be small enough so that $|\log \lambda x| \geq 1$, $\lambda \in [0, \lambda_0)$, $x \in [0, 2]$; κ is non-decreasing on $[0, 2\lambda_0]$; and $f(x) \le 1 + S_x - \kappa_x$, $x \in [0, 2\lambda_0]$. Then,

$$\begin{aligned} \kappa_{\lambda x} = \beta\sqrt{\gamma(\lambda x)\,\log|\log \lambda x|} &\ge \sqrt{\gamma(x)}\,\kappa_\lambda \left|\frac{\log|-\log\lambda - \log 2|}{\log|\log\lambda|}\right|^{1/2} \\ &= \sqrt{\gamma(x)}\,\kappa_\lambda\, q_\lambda, \end{aligned} \tag{3.2.19}$$

where

$$q_\lambda := \left|1 + \frac{\log\left|1 + \frac{\log 2}{\log \lambda}\right|}{\log|\log\lambda|}\right|^{1/2} \xrightarrow[\lambda\downarrow 0]{} 1. \tag{3.2.20}$$

It follows that, for $\lambda \in [0, \lambda_0)$ and $x \in [0, 2]$,

$$f(\lambda x) \le 1 + \kappa_{\lambda x}\left(\frac{|S_{\lambda x}|}{\kappa_{\lambda x}} - 1\right) \le 1 + \frac{\kappa_{\lambda x}}{q_\lambda}\left(\frac{|\xi_x^\lambda|}{\sqrt{\gamma(x)}} - q_\lambda\right). \tag{3.2.21}$$

Let $\widehat{\xi_x^\lambda} := 2^{-\alpha_2/2}\xi_{2x}^\lambda$. Then, for $\mu \in [0, 1]$,

$$\begin{aligned} \psi(\lambda,\mu) - 1 &\le \frac{1}{q_\lambda}\int_\mu^{\mu+1} \kappa_{\lambda x}\left(\frac{|\xi_x^\lambda|}{\sqrt{\gamma(x)}} - q_\lambda\right)\mathrm{d}x \\ &= \frac{2}{q_\lambda}\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \kappa_{2\lambda x}\left(\frac{|\xi_{2x}^\lambda|}{\sqrt{\gamma(2x)}} - q_\lambda\right)\mathrm{d}x \\ &= \frac{2}{q_\lambda}\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \kappa_{2\lambda x}\left(\frac{|\widehat{\xi_x^\lambda}|}{\sqrt{\gamma(x)}} - q_\lambda\right)\mathrm{d}x \\ &\le \frac{2}{q_\lambda}\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \kappa_{2\lambda x}\left(\frac{|\widehat{\xi_x^\lambda}|}{\sqrt{\gamma(x)}} - 1\right)\mathrm{d}x + \left|1 - \frac{1}{q_\lambda}\right|\,\kappa_{\lambda(\mu+1)}. \end{aligned} \tag{3.2.22}$$

Next, we abbreviate $2^{\alpha_2/2}/(\mu+1)^{\alpha_2/2}$ by $\zeta(\mu)$ and rewrite

$$\begin{aligned} &\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \kappa_{2\lambda x}\left(\frac{|\widehat{\xi_x^\lambda}|}{\sqrt{\gamma(x)}} - 1\right)\mathrm{d}x \\ &= \zeta(\mu)\,\kappa_{\lambda(\mu+1)}\Bigg(\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\widehat{\xi_x^\lambda}| - \sqrt{\gamma(x)}\,\mathrm{d}x \\ &\qquad + \frac{1}{\zeta(\mu)}\int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}}\left(\frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)}\sqrt{\gamma(x)}} - \zeta(\mu)\right)\left(|\widehat{\xi_x^\lambda}| - \sqrt{\gamma(x)}\right)\mathrm{d}x\Bigg). \end{aligned} \tag{3.2.23}$$

Using that $1 \le \zeta(\mu) \le 2^{\alpha_2/2}$, that $(\widehat{\xi_x^\lambda})_{x\in[0,1],\lambda\in(0,1]} \overset{d}{=} (\xi_x^\lambda)_{x\in[0,1],\lambda\in(0,1]}$ by the scaling relation (3.2.4), and the Cauchy-Schwarz inequality in conjunction with Lemma 3.2.2(i),(iii) we obtain

$$\limsup_{\lambda\downarrow 0} \sup_{\mu\in[0,1]} \left| \frac{1}{\zeta(\mu)} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \left(\frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)}\sqrt{\gamma(x)}} - \zeta(\mu) \right) \left(|\widehat{\xi}_x^\lambda| - \sqrt{\gamma(x)} \right) \mathrm{d}x \right| = 0. \tag{3.2.24}$$

In conclusion,

$$\begin{aligned} \frac{\psi(\lambda,\mu)-1}{2^{\alpha_2/2}\,\kappa_{\lambda(\mu+1)}} \leq \frac{\psi(\lambda,\mu)-1}{\zeta(\mu)\,\kappa_{\lambda(\mu+1)}} & \\ \leq \frac{2}{q_\lambda} \sup_{\mu\in[0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} |\widehat{\xi}_x^\lambda| - \sqrt{\gamma(x)}\,\mathrm{d}x & \\ + \frac{2}{q_\lambda} \sup_{\mu\in[0,1]} \left| \frac{1}{\zeta(\mu)} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \left(\frac{\kappa_{2\lambda x}}{\kappa_{\lambda(\mu+1)}\sqrt{\gamma(x)}} - \zeta(\mu) \right) \left(|\widehat{\xi}_x^\lambda| - \sqrt{\gamma(x)} \right) \mathrm{d}x \right| & \\ + \left| 1 - \frac{1}{q_\lambda} \right|, & \end{aligned} \tag{3.2.25}$$

for which (3.2.20), $(\widehat{\xi}_x^\lambda)_{x\in[0,1],\lambda\in(0,1]} \stackrel{d}{=} (\xi_x^\lambda)_{x\in[0,1],\lambda\in(0,1]}$, Lemma 3.2.2 (ii) and (3.2.24) yield the existence of a $\lambda_0 > 0$ such that

$$\psi(\lambda,\mu) \leq 1 - 2^{\alpha_2/2}\,\kappa_{\lambda(\mu+1)}\,\frac{\eta}{2} \tag{3.2.26}$$

for all $\lambda \in [0,\lambda_0)$ and $\mu \in [0,1]$. □

Remark 3.2.3. *(a) Our proof of Proposition 3.2.1 also applies to densities f with the property*

$$f(x) = \left(1 + \widetilde{\kappa}_x \left(\frac{S_x}{\kappa_x} - 1 \right) \right)_+ \wedge 1, \quad x \in [0,1], \tag{3.2.27}$$

for a non-negative non-decreasing function $\widetilde{\kappa}$ obeying

$$\lim_{\lambda\downarrow 0} \sup_{\mu\in[0,1]} \int_{\frac{\mu}{2}}^{\frac{\mu+1}{2}} \left| \frac{\widetilde{\kappa}_{2\lambda x}}{\widetilde{\kappa}_{\lambda(\mu+1)}\sqrt{\gamma(x)}} - \sqrt{\frac{2^{\alpha_2}}{(\mu+1)^{\alpha_2}}} \right|^2 \mathrm{d}x = 0. \tag{3.2.28}$$

In addition, one can cover densities f with

$$f(x) = \frac{|S_x|}{\kappa_x} \wedge 1, \quad x \in [0,1] \tag{3.2.29}$$

by repeating the proof of Lemma 3.2.2(ii) for the final line in (3.2.22) *with* 1 *in place of* κ.

(b) By using a very similar method, we can verify condition (3.0.5) *for densities f such that*

$$f(x) = \frac{|S_{1/x}|}{\kappa_{1/x}} \wedge 1, \quad x \in (0,1], \tag{3.2.30}$$

where S satisfies the scaling property (3.2.4) *and a local functional LIL "at infinity": Almost surely, the family* $\left\{\left(\frac{S_{rx}}{\kappa_r}\right)_{x\in[0,1]}\right\}_{r\geq 3}$ *is relatively compact in $C([0,1])$ with the set of limit points K as above. The local functional LIL at infinity is known for various classes of Gaussian processes S, including fractional Brownian motion (see [85, Example 4.36]), semi-stable Gaussian processes (see [94, Theorem 4]), Gaussian processes that are not necessarily semi-stable (see [95, Theorem 4]) but for which [95, Condition (A-1)] makes an adaptation of our proof possible, and rescalings of Brownian motion (see [16, Theorems 1–3]).*

(c) Another interesting process admitting a local functional LIL at infinity is iterated Brownian motion (see [61, Theorem 1.1]). In this case, our proof can be adjusted as follows. Let $(W_x^1)_{x\in\mathbb{R}}$ and $(W_x^2)_{x\geq 0}$ be two independent standard Brownian motions. Define

$$S_x = W^1_{W^2_x}, \quad x \geq 0 \tag{3.2.31}$$

and

$$\kappa_x = 2^{3/4} x^{1/4} (\log\log x)^{3/4}, \quad x \geq 3. \tag{3.2.32}$$

The relevant compact subset K of $C([0,1])$ is then given by

$$K = \left\{ f \circ g : \; f \in C([-1,1]), g \in C([0,1]), \; f(0) = g(0) = 0, \right.$$
$$\left. \int_{-1}^{1} f'(x)^2 \,\mathrm{d}x + \int_0^1 g'(x)^2 \,\mathrm{d}x \le 1 \right\}. \tag{3.2.33}$$

Indeed, [61, Theorem 1.1] implies that, almost surely,

$$\limsup_{r\to\infty} \mathcal{I}\left(\frac{S_{r\cdot}}{\kappa_r}\right) = \sup_{\phi\in K} \mathcal{I}(\phi), \tag{3.2.34}$$

for any continuous functional $\mathcal{I} : C([0,1]) \to \mathbb{R}$. This allows us to redo the proofs of Lemma 3.2.2 and Proposition 3.2.1. In particular, the inequalities in (3.2.14) *can be replaced by*

$$\begin{aligned} \phi(x) = \int_0^1 f'(y)\, \mathbf{1}_{\{y\le g(x)\}} \,\mathrm{d}y &\le \left(\int_0^1 f'(y)^2 \,\mathrm{d}y\right)^{1/2} \sqrt{g(x)} \\ &\le \sqrt{\left(\int_0^x g'(y)^2 \,\mathrm{d}y\right)^{1/2} x^{1/2}} \\ &\le x^{1/4}, \end{aligned} \tag{3.2.35}$$

for all $x \in [0,1]$ and $\phi = f \circ g \in K$.

3.2.2 Initial conditions constructed from periodic functions

Let $\Psi : [0,\infty) \to [-1,1]$ be a periodic function with $\sup_{x\ge 0} \int_0^x \Psi(y)\,\mathrm{d}y < \infty$ and $\limsup_{x\to\infty} \Psi(x) = 1$. In this subsection, we show that, for any $\alpha > 0$, the oscillating probability density given by

$$f(x) = \frac{1}{2}\left(1 + \Psi\left(\frac{1}{x^\alpha}\right)\right), \quad x \in (0,a] \tag{3.2.36}$$

satisfies condition (3.0.5). The parameter α controls how fast the density oscillates (cf. Remark 3.0.5).

Proposition 3.2.4. *Every probability density f defined by* (3.2.36) *obeys condition* (3.0.5).

Proof. We only need to check (3.0.5b). To this end, for $\lambda \in \left(0, \frac{a}{2}\right)$ and $\mu \in [0,1]$, we compute

$$\psi(\lambda, \mu) - \frac{1}{2} = \frac{1}{2}\int_{\mu}^{\mu+1} \Psi\left(\frac{1}{\lambda^{\alpha}x^{\alpha}}\right) \mathrm{d}x = \frac{1}{2\alpha\lambda}\int_{\frac{1}{\lambda^{\alpha}(\mu+1)^{\alpha}}}^{\frac{1}{\lambda^{\alpha}\mu^{\alpha}}} \frac{\Psi(x)}{x^{\frac{1}{\alpha}+1}} \mathrm{d}x.$$

Integrating by parts, writing $H(x)$ for $\int_0^x \Psi(y)\,\mathrm{d}y$, and using $\mu + 1 \leq 2$ we get

$$\begin{aligned}\psi(\lambda, \mu) - \frac{1}{2} = &\frac{1}{2\alpha\lambda}\left[\lambda^{\alpha+1}\mu^{\alpha+1}H\left(\frac{1}{\lambda^{\alpha}\mu^{\alpha}}\right) - \lambda^{\alpha+1}(\mu+1)^{\alpha+1}H\left(\frac{1}{\lambda^{\alpha}(\mu+1)^{\alpha}}\right)\right]\\ &+ \frac{1}{2\alpha\lambda}\left(\frac{1}{\alpha}+1\right)\int_{\frac{1}{\lambda^{\alpha}(\mu+1)^{\alpha}}}^{\frac{1}{\lambda^{\alpha}\mu^{\alpha}}} \frac{H(x)}{x^{\frac{1}{\alpha}+2}}\,\mathrm{d}x\\ &\leq \sup_{x\geq 0} H(x)\,\frac{\lambda^{\alpha}}{\alpha}\,2^{\alpha+1}.\end{aligned}$$

Therefore, it holds

$$\sup_{\mu\in[0,1]} \psi(\lambda, \mu) < \frac{3}{4} \tag{3.2.37}$$

for all $\lambda \geq 0$ small enough. □

3.3 Refined analysis for some piecewise constant initial conditions

This section is devoted to the well-posedness question for oscillatory and piecewise constant probability densities defined by

$$f(x) = \begin{cases} \alpha_1, & x \in \bigcup_{n\geq 1}[a_{2n}, a_{2n-1}),\\ \alpha_2, & x \in \bigcup_{n\geq 1}[a_{2n+1}, a_{2n}),\end{cases} \tag{3.3.1}$$

where $0 < \alpha_1 < 1 < \alpha_2$, $a_{2n-1} = r^{n-1}a_1$, $a_{2n} = pr^{n-1}a_1$, and $r = pq$, $p, q \in (0,1)$. Such densities are of interest because they can violate both (3.0.5a) and (3.0.5b), thus necessitating additional arguments to prove the uniqueness of the associated physical solution. Note that the CDF F is piecewise linear and oscillates between the half-lines $y = \beta_1 x$ and $y = \beta_2 x$, with $0 < \beta_1 < \beta_2$ given by

$$\beta_1 = \frac{1}{1-pq}\big(\alpha_2 p(1-q) + \alpha_1(1-p)\big), \tag{3.3.2}$$

$$\beta_2 = \frac{1}{1-pq}\big(\alpha_2(1-q) + \alpha_1 q(1-p)\big). \tag{3.3.3}$$

For technical reasons (see Proposition 3.3.5 below), we assume in the following that $\beta_2 < 1$, namely

$$\alpha_2 < 1 + q\frac{1-p}{1-q}(1-\alpha_1). \tag{3.3.4}$$

Condition (3.0.5a) is not satisfied by f. For $q \in (0, 1/2]$, condition (3.0.5b) fails for it as well.

Proposition 3.3.1. *For $q \in (0, 1/2]$, the density f defined by* (3.3.1) *violates condition* (3.0.5b).

Proof. Take $\lambda = \frac{1-q}{q} a_{2n+1}$ for an integer $n \geq 1$ and set $\widetilde{\mu} = \frac{q}{1-q} \in (0, 1]$. Observe that $\lambda\widetilde{\mu} = a_{2n+1}$, whereas $\lambda(\widetilde{\mu} + 1) = a_{2n+1}\big(1 + \frac{1-q}{q}\big) = a_{2n}$. Thus,

$$\int_{\widetilde{\mu}}^{\widetilde{\mu}+1} f(\lambda x)\,\mathrm{d}x = \alpha_2 > 1. \tag{3.3.5}$$

Consequently, also

$$\sup_{\mu\in[0,1]} \int_{\mu}^{\mu+1} f(\lambda x)\,\mathrm{d}x = \alpha_2 > 1. \tag{3.3.6}$$

Hence, condition (3.0.5b) cannot hold. □

Nevertheless, we are able to prove Proposition 3.0.6. Our proof relies on the next proposition, akin to Proposition 3.1.3.

Proposition 3.3.2. *For any $\alpha_2 > 1$ close enough to* 1*,*

$$\sup_{t\in(0,T]} \sup_{h>0} \mathbb{E}\left[\frac{F(Y_t + h) - F(Y_t)}{h}\right] =: \delta_0 < 1. \tag{3.3.7}$$

Once this result is proved, the desired uniqueness on $[0,T]$ can be shown by proceeding as in the proof of Theorem 3.0.1, only with $1-\delta_0$ in place of $\Phi(t,\lambda)$. The strategy of the proof of Proposition 3.3.2, in turn, lies in finding a set $G \subset [0,\infty)$ such that for $\alpha_2 > 1$ close enough to 1,

$$\sup_{y\in G}\ \sup_{h>0:\, y+h\le a_1} \frac{F(y+h)-F(y)}{h} =: L < 1. \tag{3.3.8}$$

Then, estimating the expectation in (3.3.7) according to

$$\mathbb{E}\left[\frac{F(Y_t+h)-F(Y_t)}{h}\right] \le \alpha_2 - (\alpha_2 - L)\,\mathbb{P}\left(Y_t \in G\right) \tag{3.3.9}$$

it remains to check that Y_t falls into G with a sufficiently high probability, namely

$$\inf_{t\in(0,T]} \mathbb{P}\left(Y_t \in G\right) > \frac{\alpha_2 - 1}{\alpha_2 - L}. \tag{3.3.10}$$

The two assertions (3.3.8) and (3.3.10) are the subjects of Subsections 3.3.1 and 3.3.2, respectively.

3.3.1 Proof of (3.3.8)

Lemma 3.3.3. *Let $G = \bigcup_{n\ge1}[a_{2n+2}, \varrho a_{2n+1}] \cup [a_2,\infty)$, where $\varrho := \frac{1+p}{2}$. Then,*

$$\sup_{y\in G}\ \sup_{h>0:\, y+h\le a_1} \frac{F(y+h)-F(y)}{h} = \frac{(1-q)\alpha_2 + q(1-\varrho)\alpha_1}{1-q\varrho} =: L. \tag{3.3.11}$$

Moreover, for $\alpha_2 > 1$ close enough to 1, it holds $L < 1$.

Proof. It suffices to show (3.3.11) with $G\backslash[a_2,\infty)$ in place of G. To this end, fix an $n \ge 1$ and a $y \in [a_{2n+2}, \varrho a_{2n+1}]$. Define the function

$$\theta:\ (0, a_1 - y] \to [0,\infty), \quad h \mapsto \frac{F(y+h)-F(y)}{h}. \tag{3.3.12}$$

By the definition of F, we have for $k = 1, 2, \ldots, n$:

$$\begin{cases} \theta'(h) \geq 0, & y + h \in (a_{2k+1}, a_{2k}), \\ \theta'(h) \leq 0, & y + h \in (a_{2k}, a_{2k-1}). \end{cases} \tag{3.3.13}$$

Therefore,

$$\sup_{h>0:\, y+h \leq a_1} \frac{F(y+h) - F(y)}{h} = \sup_{a_{2k} \geq a_{2n}} \frac{F(a_{2k}) - F(y)}{a_{2k} - y}. \tag{3.3.14}$$

Notice now that the sequence

$$\left(\frac{F(a_{2k}) - F(y)}{a_{2k} - y} \right)_{k=1,\, 2,\, \ldots,\, n} \tag{3.3.15}$$

is non-decreasing. Indeed, for $k = 2, 3, \ldots, n$,

$$\frac{F(a_{2k}) - F(y)}{a_{2k} - y} - \frac{F(a_{2k-2}) - F(y)}{a_{2k-2} - y} = \frac{(y\beta_2 - F(y))(a_{2k-2} - a_{2k})}{(a_{2k} - y)(a_{2k-2} - y)} \geq 0. \tag{3.3.16}$$

We conclude

$$\sup_{h>0:\, y+h \leq a_1} \frac{F(y+h) - F(y)}{h} = \frac{F(a_{2n}) - F(y)}{a_{2n} - y}. \tag{3.3.17}$$

Since the right-hand side is non-decreasing in y on $[a_{2n+2}, \varrho a_{2n+1}]$,

$$\begin{aligned} \sup_{y \in [a_{2n+2}, \varrho a_{2n+1}]} \sup_{h>0:\, y+h \leq a_1} \frac{F(y+h) - F(y)}{h} &= \frac{F(a_{2n}) - F(\varrho a_{2n+1})}{a_{2n} - \varrho a_{2n+1}} \\ &= \frac{(1-q)\alpha_2 + q(1-\varrho)\alpha_1}{1 - q\varrho}. \end{aligned} \tag{3.3.18}$$

This proves the first statement. The second one is straightforward to verify. □

3.3.2 Proof of (3.3.10)

The key step in deriving (3.3.10) is an estimate of the probabilities

$$\mathbb{P}\big(Y_t \in [a\sqrt{t}, b\sqrt{t}]\big), \quad 0 < a < b, \quad t \in (0, T]. \tag{3.3.19}$$

For that purpose, we establish the 1/2–Hölder continuity of the frontier Λ on $[0,T]$. As a preparation for the latter, we introduce for each $t \in [0,T]$ the function

$$F_t : [0,\infty) \to [0,1], \quad x \mapsto \mathbb{P}\,(0 < X_t \le x) = \mathbb{E}\,[F(\Lambda_t - B_t + x) - F(\Lambda_t - B_t)] \tag{3.3.20}$$

and notice immediately that $F_t'(x) \le \alpha_2$. Moreover, we have the following bound.

Lemma 3.3.4. *For all $t \in [0,T]$, it holds*

$$\Lambda_{t+h} - \Lambda_t - F_t(\Lambda_{t+h} - \Lambda_t) \le \alpha_2 \sqrt{\frac{2}{\pi}}\,\sqrt{h}, \quad h > 0. \tag{3.3.21}$$

Proof. We start with the inequalities

$$\begin{aligned}
\Lambda_{t+h} - \Lambda_t &= \mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) < X_{0-} \le \sup_{0\le s\le t+h}(-B_s + \Lambda_s)\big) \\
&= \mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) + B_t - \Lambda_t < X_{0-} + B_t - \Lambda_t \le \sup_{t\le s\le t+h}(-B_s + \Lambda_s) + B_t - \Lambda_t\big) \\
&\le \mathbb{P}\big(0 < X_t \le \sup_{t\le s\le t+h}(B_t - B_s) + \Lambda_{t+h} - \Lambda_t\big) \\
&= F_t(\Lambda_{t+h} - \Lambda_t) \\
&\quad + \mathbb{P}\big(\{\Lambda_{t+h} - \Lambda_t < X_t\} \cap \big\{X_t - (\Lambda_{t+h} - \Lambda_t) \le \sup_{t\le s\le t+h}(B_t - B_s)\big\}\big).
\end{aligned} \tag{3.3.22}$$

Consequently,

$$\begin{aligned}
\Lambda_{t+h} - \Lambda_t - F_t(\Lambda_{t+h} - \Lambda_t) &\le \int_0^\infty \mathbb{P}\big(x \le \sup_{t\le s\le t+h}(B_t - B_s)\big)\,\mathrm{d}F_t(x + \Lambda_{t+h} - \Lambda_t) \\
&\le \alpha_2 \sqrt{\frac{2}{\pi}}\,\sqrt{h},
\end{aligned} \tag{3.3.23}$$

as stated in the lemma. □

As a direct implication, we obtain the square root behavior of the frontier Λ.

Proposition 3.3.5. *For any $\alpha_2 > 1$ close enough to 1, there exist $0 < c_1 \le c_2 < \infty$ such* *that*

$$c_1\sqrt{t} \leq \Lambda_t \leq c_2\sqrt{t}, \quad t \in [0,T]. \tag{3.3.24}$$

Proof. For the lower bound, we notice that $Z_t \geq \sup_{0\leq s\leq t}(-B_s)$, and hence,

$$\Lambda_t = \mathbb{E}\left[F(Z_t)\right] \geq \mathbb{E}\big[F\big(\sup_{0\leq s\leq t}(-B_s)\big)\big] \geq \beta_1\mathbb{E}\big[\sup_{0\leq s\leq t}(-B_s)\big] = \beta_1\sqrt{\frac{2}{\pi}}\sqrt{t}, \quad t \in [0,T]. \tag{3.3.25}$$

For the upper bound, we apply Lemma 3.3.4 with $t = 0$ and get

$$\Lambda_h \leq \frac{\alpha_2}{1-\beta_2}\sqrt{\frac{2}{\pi}}\sqrt{h}, \quad h \in (0,T] \tag{3.3.26}$$

thanks to $\Lambda_0 = 0$ and $F_0(x) = F(x) \leq \beta_2 x$. □

The 1/2–Hölder continuity of Λ on $[0,T]$ is deduced similarly from the next proposition.

Proposition 3.3.6. *For any $\alpha_2 > 1$ close enough to 1, there exists a $\beta \in [0,1)$ such that*

(i) $F_t(x) \leq \beta x$, $x \geq 0$, $t \in [0,T]$, *and*

(ii) $\mathbb{E}\left[f(\Lambda_{t-} + B_t)\right] \leq \beta$, $t \in (0,T]$.

In particular, Λ is continuous on $[0,T]$.

Proof. Fix a $C \in (0,\infty)$ and consider a $t \in (0,T]$. Then, for $x > C\Lambda_t$,

$$\begin{aligned} F_t(x) \leq \mathbb{E}\left[F(\Lambda_t - B_t + x)\right] &\leq \beta_2\mathbb{E}\left[(\Lambda_t - B_t + x)_+\right] \\ &\leq \beta_2\bigg(\frac{1+C}{C}x + \frac{1}{\sqrt{2\pi}}\sqrt{t}\bigg). \end{aligned} \tag{3.3.27}$$

In view of the square root lower bound $\sqrt{t} \leq \frac{\Lambda_t}{c_1}$, we have for $x > C\Lambda_t$,

$$F_t(x) \leq \beta_2\bigg(\frac{1+C}{C} + \frac{1}{Cc_1\sqrt{2\pi}}\bigg)x. \tag{3.3.28}$$

Since $c_1 = \beta_1\sqrt{2/\pi} \geq \alpha_1\sqrt{2/\pi}$, we conclude

$$F_t(x) \leq \beta_2\left(\frac{1+C}{C} + \frac{1}{2\alpha_1 C}\right)x, \quad x > C\Lambda_t. \tag{3.3.29}$$

Next, take $x \leq C\Lambda_t$. By definition,

$$F_t(x) = \mathbb{E}\left[\int_{\Lambda_t - B_t}^{\Lambda_t - B_t + x} f(y)\,\mathrm{d}y\right] = \mathbb{E}\left[\int_{\Lambda_t}^{\Lambda_t + x} f(y - B_t)\,\mathrm{d}y\right] = \int_{\Lambda_t}^{\Lambda_t + x} \mathbb{E}\left[f(y + B_t)\right]\,\mathrm{d}y. \tag{3.3.30}$$

Thus, it suffices to show that for any $\alpha_2 > 1$ close enough to 1, there exists a $\beta \in [0, 1)$ such that

$$\mathbb{E}\left[f(y + B_t)\right] \leq \beta, \quad y \in [\Lambda_t, (1 + C)\Lambda_t]. \tag{3.3.31}$$

Set $H = \bigcup_{k\geq 1} [a_{2k}, a_{2k-1}) \cup [a_1, \infty)$ and estimate $\mathbb{E}\left[f(y + B_t)\right]$ according to

$$\begin{aligned} \mathbb{E}\left[f(y + B_t)\right] &= \mathbb{E}\left[(f\,\mathbf{1}_H)(y + B_t)\right] + \mathbb{E}\left[(f\,\mathbf{1}_{H^c})(y + B_t)\right] \\ &\leq \alpha_2 - (\alpha_2 - \alpha_1)\,\mathbb{P}\left(y + B_t \in H\right). \end{aligned} \tag{3.3.32}$$

Our goal now is to lower bound $\mathbb{P}\left(y + B_t \in H\right)$ for $y \in [\Lambda_t, (1 + C)\Lambda_t]$. We distinguish four cases.

Case 1: $y \in \left[a_{2n+2}, \frac{a_{2n+2}+a_{2n+1}}{2}\right)$ for some $n \geq 0$. In this case, we find

$$\begin{aligned} \mathbb{P}\left(y + B_t \in H\right) &\geq \mathbb{P}\big(B_t \in [a_{2n+2} - y, a_{2n+1} - y)\big) \\ &\geq \mathbb{P}\left(B_t \in \left[0, \frac{a_{2n+1} - a_{2n+2}}{2}\right)\right). \end{aligned} \tag{3.3.33}$$

In view of

$$\frac{a_{2n+1} - a_{2n+2}}{2} = \frac{1-p}{2}a_{2n+1} \geq \frac{1-p}{2}y \geq \frac{1-p}{2}\Lambda_t, \tag{3.3.34}$$

we get

$$\mathbb{P}\left(y + B_t \in H\right) \geq \mathbb{P}\left(B_t \in \left[0, \frac{1-p}{2}\Lambda_t\right)\right). \tag{3.3.35}$$

Case 2: $y \in \left[\frac{a_{2n+2}+a_{2n+1}}{2}, a_{2n+1}\right)$ for some $n \geq 0$. Similarly to the previous case, we have

$$\mathbb{P}\left(y + B_t \in H\right) \geq \mathbb{P}\Big(B_t \in \Big[-\frac{a_{2n+1} - a_{2n+2}}{2}, 0\Big]\Big) \geq \mathbb{P}\Big(B_t \in \Big[0, \frac{1-p}{2}\Lambda_t\Big]\Big). \tag{3.3.36}$$

Case 3: $y \in [a_{2n+1}, a_{2n})$ for some $n \geq 1$. In this case,

$$\begin{aligned} \mathbb{P}\left(y + B_t \in H\right) &\geq \mathbb{P}\left(B_t \in [a_{2n} - y, a_{2n-1} - y]\right) \\ &\geq \frac{a_{2n-1} - a_{2n}}{\sqrt{2\pi t}}\, e^{-\frac{(a_{2n-1}-y)^2}{2t}} \\ &\geq \frac{a_{2n-1} - a_{2n}}{\sqrt{2\pi t}}\, e^{-\frac{(a_{2n-1}-a_{2n+1})^2}{2t}}. \end{aligned} \tag{3.3.37}$$

Using

$$a_{2n-1} - a_{2n} = \frac{1-p}{p} a_{2n} \geq \frac{1-p}{p} y \geq \frac{1-p}{p}\Lambda_t \tag{3.3.38}$$

and

$$a_{2n-1} - a_{2n+1} = \frac{1-pq}{pq} a_{2n+1} \leq \frac{1-pq}{pq} y \leq \frac{1-pq}{pq}(1+C)\Lambda_t \tag{3.3.39}$$

we end up with

$$\mathbb{P}\left(y + B_t \in H\right) \geq \frac{1-p}{p}\frac{\Lambda_t}{\sqrt{2\pi t}} e^{-\frac{(1-pq)^2(1+C)^2}{(pq)^2}\frac{\Lambda_t^2}{2t}}. \tag{3.3.40}$$

Case 4: $y \in [a_1, \infty)$. Here,

$$\mathbb{P}\left(y + B_t \in H\right) \geq \mathbb{P}\big(B_t \in [a_1 - y, \infty)\big) \geq \frac{1}{2}. \tag{3.3.41}$$

Combining (3.3.35), (3.3.36), (3.3.40) and (3.3.41), and employing $c_1 \leq \frac{\Lambda_t}{\sqrt{t}} \leq c_2$, we arrive at

$$\mathbb{P}\left(y + B_t \in H\right) \geq \min\left(\mathbb{P}\Big(\mathcal{N} \in \Big[0, \frac{1-p}{2}c_1\Big]\Big), \frac{1-p}{p}\frac{c_1}{\sqrt{2\pi}} e^{-\frac{(1-pq)^2(1+C)^2}{(pq)^2}\frac{c_2^2}{2}}, \frac{1}{2}\right). \tag{3.3.42}$$

At this point, we choose $C = \frac{(2\alpha_1+1)\beta_2}{\alpha_1(1-\beta_2)}$, so that

$$\beta_2\left(\frac{1+C}{C} + \frac{1}{2\alpha_1 C}\right) = \frac{1+\beta_2}{2} < 1. \tag{3.3.43}$$

Then, the right-hand side in (3.3.42) depends on α_2 via c_1, c_2, and C. For $\alpha_2 \downarrow 1$, the values of β_1, β_2 tend to (distinct) limits in $(0,1)$, hence c_1 stays bounded away from zero, and c_2, C stay bounded away from infinity. Therefore,

$$\begin{aligned}&\liminf_{\alpha_2\downarrow 1} \mathbb{P}\left(y+B_t \in H\right) \geq \\ &\liminf_{\alpha_2\downarrow 1} \min\left(\mathbb{P}\Big(\mathcal{N}\in\Big[0,\frac{1-p}{2}c_1\Big]\Big), \frac{1-p}{p}\frac{c_1}{\sqrt{2\pi}}e^{-\frac{(1-pq)^2(1+C)^2}{(pq)^2}\frac{c_2^2}{2}}, \frac{1}{2}\right) > 0.\end{aligned} \tag{3.3.44}$$

Consequently, for any $\alpha_2 > 1$ close enough to 1,

$$\mathbb{P}\left(y+B_t\in H\right) > \frac{\alpha_2-1}{\alpha_2-\alpha_1}, \tag{3.3.45}$$

yielding by (3.3.32) a $\beta\in[0,1)$ such that

$$\mathbb{E}\left[f(y+B_t)\right] \leq \beta, \quad y\in[\Lambda_t,(1+C)\Lambda_t]. \tag{3.3.46}$$

Together with (3.3.30), (3.3.29) and (3.3.43) this finishes the proof of (i).

Result (ii) can be obtained by noticing that

$$c_1 = c_1 \sup_{0<s<t}\frac{\sqrt{s}}{\sqrt{t}} \leq \sup_{0<s<t}\frac{\Lambda_s}{\sqrt{t}} = \frac{\Lambda_{t-}}{\sqrt{t}}, \tag{3.3.47}$$

and by subsequently repeating (3.3.32)–(3.3.46) mutatis mutandis. Lastly, the final statement in the proposition is immediate from (ii) and the physical jump condition (3.0.3). □

Combining Lemma 3.3.4 and Proposition 3.3.6 we deduce the next proposition.

Proposition 3.3.7. *For any $\alpha_2 > 1$ close enough to 1, there exists a $c_3\in(0,\infty)$ such that*

$$\Lambda_{t+h}-\Lambda_t \leq c_3\sqrt{h}, \quad h\in[0,T-t], \quad t\in[0,T]. \tag{3.3.48}$$

Moreover, c_3 can be chosen according to

$$c_3 = \frac{\alpha_2}{1-\beta}\sqrt{\frac{2}{\pi}}. \tag{3.3.49}$$

We are now ready to estimate the probabilities in (3.3.19).

Lemma 3.3.8. *Let $U \coloneqq \sup_{0\le s\le 1}(B_s + c_3\sqrt{s})$. Then, for any $\alpha_2 > 1$ close enough to 1,*

$$\mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) \in [a\sqrt{t}, b\sqrt{t}]\big) \ge \mathbb{P}\left(|\mathcal{N}| \ge a\right)\,\mathbb{P}\left(U \le b-a\right), \tag{3.3.50}$$

for all $0 < a < b$ and $t \in (0,T]$.

Proof. We fix $0 < a < b$, $t \in (0,T]$, and set

$$\tau_a = \inf\{s > 0 :\ B_s + \Lambda_s \ge a\sqrt{t}\}. \tag{3.3.51}$$

Consider the representation

$$\mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) \in [a\sqrt{t}, b\sqrt{t}]\big) = \mathbb{P}\big(\tau_a \le t,\ \sup_{\tau_a\le s\le t}(B_s + \Lambda_s) \le b\sqrt{t}\big). \tag{3.3.52}$$

By the continuity of Λ,

$$B_{\tau_a} + \Lambda_{\tau_a} = a\sqrt{t}, \tag{3.3.53}$$

and therefore writing W for the Brownian motion $B_{\tau_a+\cdot} - B_{\tau_a}$ we find

$$\mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) \in [a\sqrt{t}, b\sqrt{t}]\big) = \mathbb{P}\big(\tau_a \le t,\ \sup_{0\le s\le t-\tau_a}(W_s + \Lambda_{\tau_a+s} - \Lambda_{\tau_a}) \le (b-a)\sqrt{t}\big). \tag{3.3.54}$$

Next, we use $\Lambda_{\tau_a+s} - \Lambda_{\tau_a} \le c_3\sqrt{s}$ to deduce

$$\mathbb{P}\big(\sup_{0\le s\le t}(-B_s + \Lambda_s) \in [a\sqrt{t}, b\sqrt{t}]\big) \ge \mathbb{P}\big(\tau_a \le t,\ \sup_{0\le s\le t}(W_s + c_3\sqrt{s}) \le (b-a)\sqrt{t}\big). \tag{3.3.55}$$

The trivial lower bound $\Lambda \geq 0$ implies

$$\mathbb{P}\left(\tau_a \leq t\right) = \mathbb{P}\big(\sup_{0 \leq s \leq t} (B_s + \Lambda_s) \geq a\sqrt{t}\big) \geq \mathbb{P}\big(\sup_{0 \leq s \leq t} B_s \geq a\sqrt{t}\big) = \mathbb{P}\left(|\mathcal{N}| \geq a\right). \quad (3.3.56)$$

This and the independence of W from τ_a yield

$$\begin{aligned} \mathbb{P}\big(\sup_{0 \leq s \leq t} (-B_s + \Lambda_s) \in [a\sqrt{t}, b\sqrt{t}]\big) &\geq \mathbb{P}\left(\tau_a \leq t\right) \mathbb{P}\left(U \leq b - a\right) \\ &\geq \mathbb{P}\left(|\mathcal{N}| \geq a\right) \mathbb{P}\left(U \leq b - a\right), \end{aligned} \quad (3.3.57)$$

finishing the proof. □

We conclude with the proof of (3.3.10).

Lemma 3.3.9. *For any $\alpha_2 > 1$ close enough to 1, there exists a $\delta < 1$ such that*

$$\inf_{t \in (0,T]} \mathbb{P}\left(Y_t \in G\right) \geq \frac{\alpha_2 - \delta}{\alpha_2 - L}. \quad (3.3.58)$$

Proof. Fix a $t \in (0, T]$. If $\sqrt{t} < a_3$, let $n \geq 1$ satisfy

$$r^{n+1} a_1 = a_{2n+3} \leq \sqrt{t} < a_{2n+1} = r^n a_1. \quad (3.3.59)$$

Then,

$$\frac{\varrho a_{2n+1} - a_{2n+2}}{\sqrt{t}} \geq \varrho - p \quad (3.3.60)$$

and

$$\frac{a_{2n+2}}{\sqrt{t}} \leq \frac{1}{q}. \quad (3.3.61)$$

Therefore, by Lemma 3.3.8,

$$\mathbb{P}\left(Y_t \in G\right) \geq \begin{cases} \mathbb{P}\left(Y_t \in [a_{2n+2}, \varrho a_{2n+1}]\right) \geq \mathbb{P}\left(|\mathcal{N}| \geq 1/q\right) \mathbb{P}\left(U \leq \varrho - p\right), & \text{if } \sqrt{t} < a_3, \\ \mathbb{P}\left(Y_t \in [a_2, \infty)\right) \geq \mathbb{P}\left(|\mathcal{N}| \geq a_2/\sqrt{t}\right) \geq \mathbb{P}\left(|\mathcal{N}| \geq 1/q\right), & \text{if } \sqrt{t} \geq a_3. \end{cases} \quad (3.3.62)$$

Since c_3 (appearing in the definition of U) stays bounded as $\alpha_2 \downarrow 1$,

$$\liminf_{\alpha_2 \downarrow 1} \mathbb{P}\left(Y_t \in G\right) \geq \mathbb{P}\left(|\mathcal{N}| \geq 1/q\right) \liminf_{\alpha_2 \downarrow 1} \mathbb{P}\left(U \leq \varrho - p\right) =: \iota > 0. \tag{3.3.63}$$

Thus, for any $\alpha_2 > 1$ close enough to 1,

$$\mathbb{P}\left(Y_t \in G\right) \geq \frac{\iota}{2}. \tag{3.3.64}$$

Choosing

$$\delta = \alpha_2 - \frac{\iota}{2}(\alpha_2 - L) \tag{3.3.65}$$

we obtain, for any $\alpha_2 > 1$ close enough to 1,

$$\mathbb{P}\left(Y_t \in G\right) \geq \frac{\iota}{2} = \frac{\alpha_2 - \delta}{\alpha_2 - L} \quad \text{and} \quad \delta < 1, \tag{3.3.66}$$

and hence, (3.3.58). $\square$

Chapter 4

Strong existence and uniqueness of a calibrated stochastic local volatility model

For the family of local stochastic volatility (LSV) models

$$\begin{cases} dS_t = \sigma_{\mathrm{lsv}}(t, S_t) a_t S_t dB_t, \ t \geq 0, \\ S_{t=0} = S_0, \end{cases} \tag{4.0.1}$$

to be calibrated to the market prices of European call options $(C(t, K))_{t>0, K>0}$ it is enough to take the leverage function σ_{lsv} as

$$\sigma_{\mathrm{lsv}}(t, K) = \frac{\sigma_{\mathrm{loc}}(t, K)}{\sqrt{\mathbb{E}\left[a_t^2 | S_t = K\right]}}, \ t > 0, K > 0, \tag{4.0.2}$$

where $(S_t)_{t \geq 0}$ is the price process, $(B_t)_{t \geq 0}$ is a real standard Brownian motion, S_0 is a positive random variable independent of $(B_t)_{t \geq 0}$, $\sigma_{\mathrm{loc}}(t, K) := \sqrt{\frac{2\partial_t C(t,K)}{\partial_K^2 C(t,K)}}, t > 0, K > 0$ is the so-called Dupire volatility [31] and $(a_t)_{t \geq 0}$ is any stochastic process. This is justified by Gyöngy's theorem [57] that asserts that a stochastic process $(S_t)_{t \geq 0}$ solving (4.0.1) and (4.0.2) has fixed-time marginal distributions given by the marginal distributions of the local volatility model $dS_t^{\text{loc}} = \sigma_{\text{loc}}(t, S_t^{\text{loc}})S_t^{\text{loc}}dB_t$ and starting at S_0. In order to make the model tractable, one commonly assumes that $(a_t)_{t\geq 0}$ is given by $(\sqrt{f(Y_t)})_{t\geq 0}$ where $f : \mathbb{R} \to \mathbb{R}_+^*$ is a bounded smooth function and $(Y_t)_{t\geq 0}$ is another real Itô-Lévy process eventually correlated to $(B_t)_{t\geq 0}$.

Formally, the joint density $p(t, s, y)$ of $(S_t, Y_t), (t, s, y) \in \mathbb{R}_+ \times \mathbb{R}_+^* \times \mathbb{R}$, solves a quasi-linear and non-local Fokker-Planck partial differential equation (PDE), whose coefficients depend upon the non-linear term

$$\frac{\int p(\cdot, \cdot, z)dz}{\int f(z)p(\cdot, \cdot, z)dz}. \tag{4.0.3}$$

While the applications are important in calibration to market implied volatility surfaces ([59], [53, Chapter 11], [1], [103]) the existence and uniqueness of solutions to the SDE and PDE problems is still an open problem. Only partial results have been obtained in particular cases.

Abergel and Tachet showed in [2, Theorem 3.1], the existence of a classical solution to the PDE problem in a bounded domain of $\mathbb{R}^2$ and with additional Dirichlet boundary condition - for short time, and for $\sup |f''|$ small enough. In [76, Theorem 1.4], the authors established the existence and uniqueness of a stationary solution to a similar SDE (a drift needs to be added to the dynamic of $(S_t)_{t\geq 0}$ to allow the possibility of a stationary measure). Jourdain and Zhou proved in [67, Theorems 2.4, 2.5], the existence of a weak solution to the SDE. Under the assumption that $(Y_t)_{t\geq 0}$ is a jump process taking finitely many values and after writing the PDE problem as a system of parabolic equations in non-divergence form, the authors are able to write a variational formulation of the problem, for which existence of a solution can be proved - by the Galerkin's method - provided that the range of f is small enough. They are then able to prove the existence of a weak solution to the SDE. Uniqueness and propagation of chaos are out of reach in the approach of [67] because higher regularity of the weak solution is needed. Global regularity of weak solutions to parabolic systems is still an open problem.

The difficulty of the analysis of the McKean-Vlasov SDE and the related non-linear PDE stems from the singularity of the denominator $\int f(z)p(\cdot,\cdot,z)dz$. However, the well-posedness of McKean-Vlasov problems with coefficients depending smoothly on the density p of the unknown process has been studied in various papers [65], [93], [72], [15].

The main contribution of our paper is the proof of well-posedness of a regularized version of SDE (4.0.1) - (4.0.2) when $(a_t)_{t\geq 0}$ is time independent and given by $f(Y)$, where Y is a fixed random variable independent of $(B_t)_{t\geq 0}$ and takes finitely many values in $\{1,\ldots,N\}$, $N\geq 2$ and $f:\{1,\ldots,N\}\to\mathbb{R}_+^*$. The regularization is chosen so that the calibration property is conserved. In other words, the fixed-time marginals of a solution are still given by the marginals of $S_t^{\text{loc}} = S_0 + \int_0^t \sigma_{\text{loc}}(s, S_s^{\text{loc}})S_s^{\text{loc}}dB_s$, $t\geq 0$.

Let $(X_t)_{t\geq 0} = (\log S_t)_{t\geq 0}$ be the log price and set $\sigma(t,x) := \sigma_{\text{loc}}(t,e^x)$, $(t,x)\in\mathbb{R}_+\times\mathbb{R}$. Herein, we assume that Y is a random variable taking finitely many values $\{1,\cdots,N\}$, $N\geq 2$. Let $\varepsilon>0$ and consider the McKean-Vlasov SDE

$$\begin{cases} dX_t = -\frac{1}{2}\frac{\varepsilon+f(Y)p(t,X_t)}{\varepsilon+\mathbb{E}[f(Y)|X_t]p(t,X_t)}\sigma(t,X_t)^2dt + \sqrt{\frac{\varepsilon+f(Y)p(t,X_t)}{\varepsilon+\mathbb{E}[f(Y)|X_t]p(t,X_t)}}\sigma(t,X_t)dB_t, \\ \mathbb{P}(X_t\in dx) = p(t,x)dx, \\ X_{t=0} = X_0, \end{cases} \tag{4.0.4}$$

where X_0 is a real random variable independent of $(B_t)_{t\geq 0}$.

We notice immediately

$$\mathbb{E}\left[\left(\sqrt{\frac{\varepsilon+f(Y)p(t,X_t)}{\varepsilon+\mathbb{E}\left[f(Y)|X_t\right]p(t,X_t)}}\sigma(t,X_t)\right)^2\Big|X_t\right] = \sigma(t,X_t)^2, \tag{4.0.5}$$

and consequently the model is calibrated, meaning $\exp(X_t)$ and S_t^{loc} have the same law for all $t\geq 0$.

In the general case where $(a_t)_{t\geq 0} = (f(Y_t))_{t\geq 0}$ and Y is an Itô process, the uniform ellipticity of the non-linear PDE - written in divergence form - solved by the joint density $p(t,x,y)$ of (X_t,Y_t), $t\geq 0$, does not hold a priori. This property is a key element in establishing uniqueness in [65, Equation 1.8]. Therefore, we made the restrictive assumption that Y is a time-independent discrete random variable and takes finitely many values. Assuming that the range of f is small enough, we are able to derive a uniform ellipticity property for the PDE problem, in the spirit of what is done in [67].

We assume in the whole chapter that σ is smooth, has bounded derivatives and there exist $0 < \sigma_0 < \sigma_1$ such that $\sigma_0 \leq \sigma(t,x) \leq \sigma_1$ for all $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$. Moreover, we assume that the measure $\mathbb{P}(X_0 \in dx \cap Y = n)$ admits a density $P_n : \mathbb{R} \to \mathbb{R}_+$ - of total mass $\mathbb{P}(Y = n)$ - with bounded first and second derivatives and such that for some $\alpha \in (0,1)$, the second derivative $P_n^{(2)}$ is α-Hölder and

$$\|P_n\|_{C^{2+\alpha}} := \sum_{0 \leq k \leq 2} \|P_n^{(k)}\|_\infty + \sup_{x,y \in \mathbb{R}, x \neq y} \frac{|P_n^{(k)}(x) - P_n^{(k)}(y)|}{|x-y|^\alpha} < \infty. \tag{4.0.6}$$

Denote $f_{\max} = \max_{1 \leq n \leq N} f(n)$, $f_{\min} = \min_{1 \leq n \leq N} f(n)$ and

$$\bar{f} = \frac{1}{N} \sum_{n=1}^{N} f(n). \tag{4.0.7}$$

We introduce the following small range condition on f.

Condition 4.0.1.

$$\frac{1}{2}\left[N + 1 - \max_{1 \leq k \leq N} \sqrt{\sum_{n=1, n\neq k}^{N} f(n) \sum_{n=1, n \neq k}^{N} \frac{1}{f(n)}}\right] \wedge 1 > \frac{1}{f_{\min}} \sqrt{\sum_{n=1}^{N} (f(n) - \bar{f})^2} + \frac{f_{\max} - f_{\min}}{f_{\min}}. \tag{4.0.8}$$

We are now ready to state our main result.

Theorem 4.0.2. *Let $T > 0$. If Condition 4.0.1 is satisfied, P belongs to $(C^{2+\alpha})^N, \alpha \in (0,1)$ and the norm $\sum_{n=1}^N \|P_n\|_{C^{2+\alpha}}$ is small enough, then there exists a unique strong solution to SDE* (4.0.4) *on* $[0,T]$. *Moreover, X admits a smooth density $p \in \left(C^{1+\alpha/4, 2+\alpha/2}\right)^N$.*

We study as well the question of propagation of chaos for SDE (4.0.4). Let $M \geq 1$, $(\delta_M)_{M \geq 1} \in (0,1)^{\mathbb{N}^*}$ be a sequence converging to 0 and $W_1 : \mathbb{R} \to \mathbb{R}_+^*$ be a bounded kernel function with a bounded derivative and satisfying $\int W_1(x)dx = 1$ and $\int xW_1(x)dx = 0$. Denote

$$W_{\delta_M} := \frac{1}{\delta_M} W_1\left(\frac{\cdot}{\delta_M}\right). \tag{4.0.9}$$

Let $(B_t^i)_{t\geq 0, i\geq 1}$ be a collection of i.i.d standard Brownian motions and $(X_0^i, Y^i)_{i\geq 1}$ be a collection of i.i.d random variables of law $\mathbb{P}(Y^i = n \cap X_0^i \in dx) = P_n(x)dx,\ 1 \leq n \leq N$.

For $M \geq 1$, we introduce the approximating system of M particles, already considered in [52] (for $\varepsilon = 0$)

$$\begin{aligned} dX_t^{i,M} =& -\frac{1}{2}\frac{\varepsilon + f(Y^i)\frac{1}{M}\sum_{j=1}^M W_{\delta_M}(X_t^{i,M} - X_t^{j,M})}{\varepsilon + \frac{1}{M}\sum_{j=1}^M f(Y^j) W_{\delta_M}(X_t^{i,M} - X^{j,M})}\sigma(t, X_t^{i,M})^2 dt \\ &+ \sqrt{\frac{\varepsilon + f(Y^i)\frac{1}{M}\sum_{j=1}^M W_{\delta_M}(X_t^{i,M} - X_t^{j,M})}{\varepsilon + \frac{1}{M}\sum_{j=1}^M f(Y^j) W_{\delta_M}(X_t^{i,M} - X^{j,M})}}\sigma(t, X_t^{i,M}) dB_t^i, \end{aligned} \tag{4.0.10}$$

initialized at $X_{t=0}^{i,M} = X_0^i$, $1 \leq i \leq N$.

Assume that the conditions of Theorem 4.0.3 are satisfied. For each $1 \leq i \leq M$, let $(\hat{X}^i)_{t\geq 0}$ be the particle starting at $X_0^{i,M}$ solving (4.0.4) and driven by the Brownian motion $(B_t^i)_{t\geq 0}$ and Y^i

$$\begin{cases} d\hat{X}_t^i = -\dfrac{1}{2}\dfrac{\varepsilon + f(Y^i)\sum_{n=1}^N \hat{p}_n(t, \hat{X}_t^i)}{\varepsilon + \sum_{n=1}^N f(n)\hat{p}_n(t, \hat{X}_t^i)}\sigma(t, \hat{X}_t^i)^2 dt \\ \qquad + \sqrt{\dfrac{\varepsilon + f(Y^i)\sum_{n=1}^N \hat{p}_n(t, \hat{X}_t^i)}{\varepsilon + \sum_{n=1}^N f(n)\hat{p}_n(t, \hat{X}_t^i)}}\sigma(t, \hat{X}_t^i) dB_t^i, \\ \mathbb{P}\left(\hat{X}_t^i \in dx \cap Y^i = n\right) = \hat{p}_n(t, x)dx, \\ 0 \leq t \leq T. \end{cases} \tag{4.0.11}$$

The existence and uniqueness of the process $(X_t^{i,M})_{0\leq t\leq T, 1\leq i\leq N}$ is ensured by the regularity of the drift and diffusion coefficients, given by Theorem 4.0.2.

Under Condition 4.0.1, propagation of chaos holds.

Theorem 4.0.3. *Let $T > 0$, $(X_t^{i,M})_{t\geq 0}$ and $(\hat{X}_t^i)_{t\geq 0}$ be given by (4.0.10) and (4.0.11)* *respectively. Assume that* $(\delta_M)_{M\geq 1}$ *converges slowly enough towards zero such that*

$$\lim_{M\to\infty} \frac{1}{M} \frac{\exp\left(CT\delta_M^{-4}\right)}{\delta_M^2} = 0. \tag{4.0.12}$$

for all constants $C > 0$.

Then for all $1 \leq i \leq N$

$$\lim_{M\to\infty} \mathbb{E}\left[\sup_{0\leq t\leq T} |X_t^i - \hat{X}_t^i|^2\right] = 0. \tag{4.0.13}$$

Consequently, for all $k \geq 1$ *the law of the process* $(X^{1,M}, X^{2,M}, \ldots, X^{k,M})$ *converges weakly towards* $\mu \otimes \mu \otimes \cdots \otimes \mu$, *where* μ *is the law of* $\hat{X}^1$.

The main application of this result is the justification of the particular method for calibrating the LSV model described in [52], [53, Section 11.6.1].

The rest of the article is structured as follows. In Section 4.1 we introduce notation and prove Theorem 4.0.2. Existence is proved in Subsection 4.1.1 and uniqueness in Subsection 4.1.2. Section 4.2 is devoted to establishing propagation of chaos.

4.1 Existence and uniqueness

Throughout the rest of the chapter $T > 0$ is fixed, and

- For any $d \geq 1$ and $X \in \mathbb{R}^d$, we denote $\|X\|_2 = \sqrt{\sum_{i=1}^d X_i^2}$ and $\|X\|_\infty = \max_{1\leq i\leq d} |X_i|$.
- The scalar product is denoted $\langle \cdot, \cdot \rangle$

$$\langle X, Y \rangle = \sum_{i=1}^d X_i Y_i, \;\; X \in \mathbb{R}^d, Y \in \mathbb{R}^d. \tag{4.1.1}$$

- We denote by $L^p, p \geq 1$ the space of measurable real-valued functions ϕ for which $\|\phi\|_{L^p} = \left(\int |\phi(x)|^p dx\right)^{1/p}$ is finite.

- We denote by $C(0,T,L^2)$ the space of measurable functions $\phi:(t,x)\in[0,T]\times\mathbb{R}\to\mathbb{R}$ such that the mapping $t\in[0,T]\to\phi(t,\cdot)$ takes values in $(L^2,\|\|_{L^2})$ and is continuous.

- For $\alpha\in(0,1)$ we denote by $C^{2+\alpha}$ the space of real functions ϕ on $\mathbb{R}$ for which $\|\phi\|_{C^{2+\alpha}}<\infty$. We denote by $C^{(k+\alpha)/2,k+\alpha}$ the space of real functions ϕ on $[0,T]\times\mathbb{R}$ which are continuous together with their derivatives $\partial_t^r\partial_x^l\phi$, $2r+l\leq k$ and admit a finite norm

$$\begin{aligned}\|\phi\|_{C^{(k+\alpha)/2,k+\alpha}} =& \sum_{2r+l\leq k}\sup_{(t,x)\in[0,T]\times\mathbb{R}}|\partial_t^r\partial_x^l\phi(t,x)| \\ &+\sum_{k-1\leq 2r+l\leq k}\sup_{x\in\mathbb{R},t,s\in[0,T]}\frac{|\partial_t^r\partial_x^l\phi(t,x)-\partial_t^r\partial_x^l\phi(s,x)|}{|t-s|^{(k-2r-l+\alpha)/2}} \\ &+\sum_{2r+l=k}\sup_{t\in[0,T],x,y\in\mathbb{R}}\frac{|\partial_t^r\partial_x^l\phi(t,x)-\partial_t^r\partial_x^l\phi(t,y)|}{|x-y|^\alpha},\end{aligned} \tag{4.1.2}$$

where ∂_t and ∂_x are the partial derivatives with respect to t and x.

- S^{++} is the set of symmetric and positive $N\times N$ real matrices.

For each $1\leq n\leq N$ and $t\geq 0$, denote by $p_n(t,x)$ the conditional density of X_t given $Y=n$, multiplied by $\mathbb{P}(Y=n)$. In other words, $p_n(t,x)dx=\mathbb{P}(X_t\in dx\cap Y=n)\,,x\in\mathbb{R},1\leq n\leq N$. Set $P_n(x)dx=\mathbb{P}(X_0\in dx\cap Y=n)$ for $1\leq n\leq N$.

The PDE problem solved by $(p_n)_{1\leq n\leq N}$ can be formulated as a system of N parabolic PDEs

$$\begin{cases}\partial_t p=\frac{1}{2}\partial_{xx}[\sigma^2B^\varepsilon(p)p]+\frac{1}{2}\partial_x\left[\sigma^2B^\varepsilon(p)p\right],\ (t,x)\in[0,T]\times\mathbb{R},\\ p_n/\mathbb{P}(Y=n)\,\text{is a probability density},\\ p_n(0,x)=P_n(x),\ x\in\mathbb{R},1\leq n\leq N,\end{cases} \tag{4.1.3}$$

where $B^\varepsilon(u), u \in (\mathbb{R}_+)^N$ is a diagonal matrix whose diagonal elements are

$$B^\varepsilon_{nn}(u) = B^\varepsilon_n(u) := \frac{\varepsilon + f(n)\sum_{k=1}^N u_k}{\varepsilon + \sum_{k=1}^N f(k)u_k}, \quad u \in \mathbb{R}_+^N, 1 \leq n \leq N. \tag{4.1.4}$$

The proof of Theorem 4.0.2 is a direct application of Propositions 4.1.1 and 4.1.4 below.

4.1.1 Existence

Under the condition that the norm $\sum_{n=1}^N \|P_n\|_{C^{2+\alpha}}$ is small enough, we prove the existence, in the classical sense, of a solution to Problem (4.1.3) and the existence of a stochastic process $(X_t)_{t\geq 0}$ solving SDE (4.0.4). The proof follows the approach of [65, Proof of Proposition 2.2] adapted to our specific McKean-Vlasov equation, whose coefficients depend on the marginal $\sum_{n=1}^N p_n$ and the quantity $\sum_{n=1}^N f(n)p_n$.

Proposition 4.1.1. *There exists a constant C, depending on T, $f_{\max}$, $f_{\min}$ and ε, such that if $\sum_{n=1}^N \|P_n\|_{C^{2+\alpha}} \leq C$, then there exist a solution $(p_n)_{1\leq n\leq N} \in (C^{1+\alpha/2,2+\alpha})^N$ to Problem* (4.1.3)*, and a solution $(X_t)_{t\in[0,T]}$ to SDE* (4.0.4)*.*

Proof. For $1 \leq n \leq N$ and non-negative $u \in (C^{1+\alpha/2,2+\alpha})^N$, define the operator

$$\partial_t v - L_{n,u}v := \partial_t v - \frac{1}{2}\partial_{xx}[\sigma^2 B^\varepsilon_n(u)v] - \frac{1}{2}\partial_x[\sigma^2 B^\varepsilon_n(u)v]. \tag{4.1.5}$$

$L_{n,u}$ is a uniformly parabolic operator of second order with coefficients in $C^{\alpha/2,\alpha}$. According to [65, Proposition 1.1](or [77, Chapter IV, Theorem 5.1]), there exists a unique $v_n \in C^{1+\alpha/2,2+\alpha}$ such that $v_n/\mathbb{P}(Y_t = n)$ is a probability density and solves in $[0,T]\times\mathbb{R}$

$$\begin{cases} \partial_t v_n = L_{n,u}v_n, \\ v_n(0,x) = P_n(x), \ x \in \mathbb{R}. \end{cases} \tag{4.1.6}$$

Moreover, there exists a constant C' depending only on the regularity of the coefficients of $L_{n,u}$ (namely $f_{\max}, f_{\min}, \varepsilon$ and $\sum_{l=1}^{N} \|u_l\|_{1+\alpha/2,2+\alpha}$) such that

$$\|v_n\|_{C^{1+\alpha/2,2+\alpha}} \leq C'\left(f_{\max}, f_{\min}, \varepsilon, \sum_{l=1}^{N} \|u_l\|_{C^{1+\alpha/2,2+\alpha}}\right) \|P_n\|_{C^{2+\alpha}}. \tag{4.1.7}$$

A useful fact about the constant C' is that it is non-decreasing in its last argument.

We construct by induction a sequence of solutions as follows. Start by $p_n^0(t,x) = P_n(x)$, $t \in [0,T], 1 \leq n \leq N$ and $x \in \mathbb{R}$. For $m \geq 1$, let $p_n^m \in C^{1+\alpha/2,2+\alpha}$, for each $1 \leq n \leq N$, be the solution of

$$\begin{cases} \partial_t p_n^m = L_{n,p^{m-1}} p_n^m, \\ p_n^m / \mathbb{P}(Y = n) \text{ is a probability density}, \\ p_n^m(0,x) = P_n(x), \; x \in \mathbb{R}. \end{cases} \tag{4.1.8}$$

Using estimate (4.1.7), we have for each $m \geq 1$

$$\sum_{n=1}^{N} \|p_n^m\|_{C^{1+\alpha/2,2+\alpha}} \leq C'\left(f_{\max}, f_{\min}, \varepsilon, \sum_{l=1}^{N} \|p_l^{m-1}\|_{C^{1+\alpha/2,2+\alpha}}\right) \sum_{n=1}^{N} \|P_n\|_{C^{2+\alpha}}. \tag{4.1.9}$$

Under the assumption $\sum_{n=1}^{N} \|P_n\|_{2+\alpha} \leq C := 1 \wedge \frac{1}{C'(f_{\max}, f_{\min}, \varepsilon, 1)}$, we have by induction that $\sum_{n=1}^{N} \|p_n^m\|_{C^{1+\alpha/2,2+\alpha}} \leq 1$ for each $m \geq 1$. Indeed, the result holds for $m = 0$. If $\sum_{n=1}^{N} \|p_n^{m-1}\|_{C^{1+\alpha/2,2+\alpha}} \leq 1$ then

$$\sum_{n=1}^{N} \|p_n^m\|_{C^{1+\alpha/2,2+\alpha}} \leq C'\left(f_{\max}, f_{\min}, \varepsilon, 1\right) \sum_{n=1}^{N} \|P\|_{C^{2+\alpha}} \leq 1. \tag{4.1.10}$$

We proved that $(p^m)_{m\geq 0}$ is a bounded sequence of elements of $\left(C^{1+\alpha/2,2+\alpha}\right)^N$, and similarly to what is done in [65, Proposition 2.2], we can extract a limit point $(p_n)_{1\leq n\leq N} \in (C^{1+\alpha/4,2+\alpha/2})^N$, such that $p_n/\mathbb{P}(Y = n)$ is a density of probability and $(p_n)_{1\leq n\leq N}$ solves Problem (4.1.3).

Now, for each $1 \leq n \leq N$, $(p_m)_{1\leq m\leq N}$ being regular, there exists a strong solution

$(X_t, Y_t)_{t\in[0,T]}$ to the SDE

$$\begin{aligned} X_t = X_0 - \frac{1}{2}\int_0^t \frac{\varepsilon + f(Y)\sum_{m=1}^N p_m(s, X_s^n)}{\varepsilon + \sum_{m=1}^N f(m)p_m(s, X_s^n)}\sigma(t, X_s)^2 ds \\ + \int_0^t \sqrt{\frac{\varepsilon + f(Y)\sum_{m=1}^N p_m(s, X_s^n)}{\varepsilon + \sum_{m=1}^N f(m)p_m(s, X_s^n)}}\sigma(t, X_s)dB_s. \end{aligned} \tag{4.1.11}$$

Moreover, for each $1 \le n \le N$ and $t \in [0, T]$, the law of $X_t 1_{Y=n}$ admits a density $q_n(t, \cdot)$ solving $\partial_t q_n = L_{n,p} q_n$, with initial data P_n. The solution of this problem being unique and given by p_n, we conclude that $q_n = p_n$ for all $1 \le n \le N$. □

4.1.2 Uniqueness

Uniqueness is proved by first writing system (4.1.3) in divergence form and then proving the uniform ellipticity of the differential operator. For all $u \in C^1(\mathbb{R} \times \mathbb{R}_+)^N$, we can rewrite

$$\partial_x[B^\varepsilon(u)u] = A^\varepsilon(u)\partial_x u, \tag{4.1.12}$$

where $(A^\varepsilon_{nk})_{1\le n,k\le N} : \mathbb{R}_+^N \to \mathbb{R}^{N\times N}$ is defined by

$$A^\varepsilon_{nn}(u) := B^\varepsilon_n(u) + u_n \frac{f(n)\sum_{l=1}^N (f(l) - f(n))u_l}{(\varepsilon + \sum_{l=1}^N f(l)u_l)^2},\ 1 \le n \le N, \tag{4.1.13}$$

$$A^\varepsilon_{nk}(u) := u_n \frac{\varepsilon(f(n) - f(k)) + f(n)\sum_{l=1}^N (f(l) - f(k))u_l}{(\varepsilon + \sum_{l=1}^N f(l)u_l)^2},\ 1 \le n \ne k \le N. \tag{4.1.14}$$

In these terms, PDEs (4.1.3) can be rewritten

$$\begin{cases} \partial_t p = \frac{1}{2}\partial_x[\sigma^2 A^\varepsilon(p)\partial_x p] + \partial_x[\left(\sigma\partial_x\sigma + \frac{1}{2}\sigma^2\right) B^\varepsilon(p)p],\ (t, x) \in [0, T] \times \mathbb{R}, \\ p_n/\mathbb{P}\left(Y = n\right) \text{is a probability density}, \\ p_n(0, x) = P_n(x),\ x \in \mathbb{R}, 1 \le n \le N. \end{cases} \tag{4.1.15}$$

We recall the following result of [67, Proof of Proposition 2.3, Corollary B.3].

Proposition 4.1.2. *Let*

$$\kappa_0 := \frac{1}{2}\left[N+1-\max_{1\le k\le N}\sqrt{\sum_{n=1,n\neq k}^{N} f(i)\sum_{n=1,n\neq k}^{N}\frac{1}{f(n)}}\right]. \tag{4.1.16}$$

Then for all $\delta>0$, $\eta>0$, $u\in(\mathbb{R}_+^*)^N$, $U\in \mathbf{1}\mathbb{R}$ and $V\in\mathbf{1}^\perp$, $\mathbf{1}=(1)_{1\le n\le N}$,

$$\begin{aligned}\langle U+V,(J+\delta I)A^0(u)(U+V)\rangle \ge \frac{N}{2}\left[N-\delta\left(1+\frac{f_{\max}}{f_{\min}}\right)\left(1+\frac{1}{2\eta}\right)\right]\|U\|_2^2 \\ +\delta\left(\kappa_0-N^2\left(1+\frac{f_{\max}}{f_{\min}}\right)\eta\right)\|V\|_2^2,\end{aligned} \tag{4.1.17}$$

where $I=(1_{i=j})_{1\le i,j\le N}$ and $J=(1)_{1\le i,j\le N}$.

We prove the following result.

Proposition 4.1.3. *Under Condition 4.0.1, there exist a matrix $\Gamma\in S^{++}$ and $\kappa>0$ such that for all $u\in\mathbb{R}_+^N$ and $X\in\mathbb{R}^N$*

$$\langle X,\Gamma A^\varepsilon(u)X\rangle \ge \kappa\|X\|_2^2. \tag{4.1.18}$$

Proof. If $\sum_{n=1}^N u_n=0$ then $A^\varepsilon(u)=I$ and the result holds. Assume that $\sum_{n=1}^N u_n>0$ and set

$$\rho := \frac{\sum_{l=1}^N f(l)u_l}{\varepsilon+\sum_{l=1}^N f(l)u_l}\in[0,1]. \tag{4.1.19}$$

Rewrite

$$\begin{aligned}A_{nn}^\varepsilon(u) &= B_n^\varepsilon(u)+u_n\frac{f(n)\sum_{l=1}^N(f(l)-f(n))u_l}{(\varepsilon+\sum_{l=1}^N f(l)u_l)^2}\\ &=1-\rho+\rho B_n^0(u)+\rho^2u_n\frac{f(n)\sum_{l=1}^N(f(l)-f(n))u_l}{(\sum_{l=1}^N f(l)u_l)^2}\\ &=(1-\rho)(1+\rho B_n^0(u))+\rho^2A_{nn}^0(u),\end{aligned} \tag{4.1.20}$$

and

$$
\begin{aligned}
A^{\varepsilon}_{nk}(u) &= u_n \frac{\varepsilon(f(n)-f(k)) + f(n)\sum_{l=1}^{N}(f(l)-f(k))u_l}{(\varepsilon + \sum_{l=1}^{N} f(l)u_l)^2} \\
&= \rho(1-\rho)\frac{(f(n)-f(k))u_n}{\sum_{l=1}^{N} f(l)u_l} + \rho^2 A^0_{nk}(u).
\end{aligned}
\tag{4.1.21}
$$

Define the matrix $D \in \mathbb{R}^{N\times N}$ by

$$
D_{nk} := \frac{(f(n)-f(k))u_n}{\sum_{l=1}^{N} f(l)u_l}, \ 1 \le n \neq k \le N, \tag{4.1.22}
$$

and

$$
D_{nn} := -\sum_{m\neq n} D_{mn} = \frac{\sum_{m=1}^{N}(f(n)-f(m))u_m}{\sum_{l=1}^{N} f(l)u_l} = B^0_n(u) - 1, \ 1 \le n \le N. \tag{4.1.23}
$$

We see immediately that $\sum_{n=1}^{N} D_{nk} = 0$, $JD = 0$, and that

$$
\max_{1\le n,k\le N} |D_{nk}(u)| \le \frac{f_{\max} - f_{\min}}{f_{\min}}. \tag{4.1.24}
$$

In view of

$$
(1-\rho)(1+\rho B^0_n) - D_{nn} = (1-\rho)(1+\rho B^0_n(u)) - \rho(1-\rho)(B^0_m(u)-1) = 1-\rho^2, \tag{4.1.25}
$$

we find that

$$
A^{\varepsilon}(u) = \rho^2 A^0(u) + (1-\rho^2)I + \rho(1-\rho)D. \tag{4.1.26}
$$

Set $M := A^0(u) - I$. For all $1 \le k \le N$, $\sum_{n=1}^{N} M_{nk} = 0$ according to [67, Proof of Lemma 3.13] and therefore $JM = 0$. The previous equation can be written

$$
A^{\varepsilon}(u) = I + \rho^2 M + \rho(1-\rho)D. \tag{4.1.27}
$$

Exploiting the idea of [67, Proof of Proposition 2.3], we are going to show that for

$\delta \in (0,1)$ small enough, there exists $\kappa > 0$ such that for all $X \in \mathbb{R}^N$ and $u \in (\mathbb{R}_+^*)^N$

$$\langle X, (J+\delta I)A^\varepsilon(u)X\rangle \geq \kappa \|X\|_2^2. \tag{4.1.28}$$

Let $\delta \in (0,1)$ and $X \in \mathbb{R}^N$. We decompose

$$(J+\delta I)A^\varepsilon(u) = J + \delta I + \rho^2 \delta M + \rho(1-\rho)\delta D. \tag{4.1.29}$$

If $\langle X, MX\rangle \leq 0$, then using $\rho \in [0,1]$,

$$\begin{aligned}\langle X, (J+\delta I)A^\varepsilon(u)X\rangle &= \langle X, (J+\delta I)X\rangle + \rho^2\delta\langle X, MX\rangle + \rho(1-\rho)\delta\langle X, DX\rangle \\ &\geq \langle X, (J+\delta I)X\rangle + \delta\langle X, MX\rangle + \rho(1-\rho)\delta\langle X, DX\rangle \\ &= \langle X, (J+\delta I)A^0 X\rangle + \rho(1-\rho)\delta\langle X, DX\rangle.\end{aligned} \tag{4.1.30}$$

If $\langle X, MX\rangle > 0$ then

$$\langle X, (J+\delta I)A^\varepsilon(u)X\rangle \geq \langle X, (J+\delta I)X\rangle + \rho(1-\rho)\delta\langle X, DX\rangle. \tag{4.1.31}$$

Write $X = U + V$ where $U \in \mathbf{1}\mathbb{R}$ and $V \in \mathbf{1}^\perp$. Then $DU = 0$ and

$$\langle X, DX\rangle = \langle V, DV\rangle + \langle U, DV\rangle. \tag{4.1.32}$$

On the one hand,

$$\begin{aligned}\langle V, DV\rangle &= \sum_{n=1}^N D_{nn}V_n^2 + \sum_{n,k=1}^N \frac{u_n(f(n)-f(k))}{\sum_{l=1}^N f(l)u_l} V_n V_k \\ &= \sum_{n=1}^N D_{nn}V_n^2 + \sum_{n,k=1}^N \frac{u_n((f(n)-\bar{f})-(f(k)-\bar{f}))}{\sum_{l=1}^N f(l)u_l} V_n V_k \\ &= \sum_{n=1}^N D_{nn}V_n^2 - \frac{\sum_{n,k=1}^N u_n(f(k)-\bar{f})V_n V_k}{\sum_{l=1}^N f(l)u_l},\end{aligned} \tag{4.1.33}$$

where we used that $\sum_{k=1}^{N} V_k = 0$ between the second and third line.

We estimate

$$\left|\sum_{n,k=1}^{N} u_n(f(k)-\bar{f})V_nV_k\right| = \left|\sum_{n=1}^{N} u_nV_n \sum_{k=1}^{N}(f(k)-\bar{f})V_k\right| \le \sqrt{\sum_{n=1}^{N} u_n^2}\sqrt{\sum_{n=1}^{N}(f(n)-\bar{f})^2\|V\|_2^2}. \tag{4.1.34}$$

Using that $u_n > 0$ for $1 \le n \le N$

$$\frac{\sqrt{\sum_{n=1}^{N} u_n^2}}{\sum_{l=1}^{N} f(l)u_l} \le \frac{1}{f_{\min}}\sqrt{\frac{\sum_{n=1}^{N} u_n^2}{\left(\sum_{l=1}^{N} u_l\right)^2}} \le \frac{1}{f_{\min}}, \tag{4.1.35}$$

we readily conclude that

$$\rho(1-\rho)\delta\langle V, DV\rangle \ge -\delta\beta(f)\|V\|_2^2. \tag{4.1.36}$$

where $\beta(f) := \left(\frac{f_{\max}-f_{\min}}{f_{\min}} + \frac{1}{f_{\min}}\sqrt{\sum_{n=1}^{N}(f(n)-\bar{f})^2}\right)$,

On the second hand, for all $\eta > 0$

$$\begin{aligned}\rho(1-\rho)\delta\langle U, DV\rangle &\ge -\delta \max_{1\le n,k\le N} |D_{nk}|N\|U\|_2\|V\|_2 \\ &\ge -\delta\beta(f)\sqrt{N}|\|U\|_2\|V\|_2 \\ &\ge -\delta\frac{\beta(f)N}{\eta}\|U\|_2^2 - \delta\beta(f)N^2\eta\|V\|_2^2.\end{aligned} \tag{4.1.37}$$

Thus combining (4.1.36), (4.1.37) and Proposition 4.1.2, we infer in the case $\langle X, MX\rangle \le 0$

$$\begin{aligned}\langle X, (J+\delta I)A^{\varepsilon}(u)X\rangle \ge \frac{N}{2}&\left[N - \delta\left[\left(1+\frac{f_{\max}}{f_{\min}}\right)\left(1+\frac{1}{2\eta}\right)+\frac{2\beta(f)}{\eta}\right]\right]\|U\|_2^2 \\ &+\delta\left(\kappa_0 - \beta(f) - N^2\left(1+\frac{f_{\max}}{f_{\min}}+\beta(f)\right)\eta\right)\|V\|_2^2.\end{aligned} \tag{4.1.38}$$

Condition 4.0.1 yields $\kappa_0 > \beta(f)$ and we can choose

$$\eta < \eta_- := \frac{1}{2}\frac{(\kappa_0 - \beta(f))}{N^2}\left(1 + \frac{f_{\max}}{f_{\min}} + \beta(f)\right)^{-1}, \tag{4.1.39}$$

and

$$\delta < \delta_-(\delta) := \frac{N}{\left[\left(1 + \frac{f_{\max}}{f_{\min}}\right)\left(1 + \frac{1}{2\eta}\right) + \frac{2\beta(f)}{\eta}\right]}. \tag{4.1.40}$$

We check that with such choice, if $\langle X, MX\rangle \leq 0$

$$\langle X, (J + \delta I)A^{\varepsilon}(u)X\rangle \geq \kappa_- \|X\|_2^2, \tag{4.1.41}$$

where

$$\kappa_- := \min\left(\frac{N}{2}\left[N - \delta\left[\left(1 + \frac{f_{\max}}{f_{\min}}\right)\left(1 + \frac{1}{2\eta}\right) + \frac{2\beta(f)}{\eta}\right]\right], \frac{\delta}{2}(\kappa_0 - \beta(f))\right) > 0. \tag{4.1.42}$$

Now in the case where $\langle X, MX\rangle > 0$

$$\begin{aligned}\langle X, (J + \delta I)A^{\varepsilon}(u)X\rangle \geq{}& N\left[1 - \beta(f)\frac{\delta}{\eta}\right]\|U\|_2^2 \\ &+ \delta\left[1 - \beta(f) - \beta(f)N^2\eta\right]\|V\|_2^2.\end{aligned} \tag{4.1.43}$$

Therefore, under Condition 4.0.1, $1 > \beta(f)$, and we can choose

$$\eta < \eta_+ := \frac{1}{2}\frac{1 - \beta(f)}{\beta(f)N^2}, \tag{4.1.44}$$

and

$$\delta < \delta_+(\eta) := \eta\beta(f). \tag{4.1.45}$$

Thus

$$\langle U + V, (J + \delta I)A^{\varepsilon}(u)(U + V)\rangle \geq \kappa_+ \|X\|_2^2, \tag{4.1.46}$$

where

$$\kappa_+ := \min\left(N\left[1-\beta(f)\frac{\delta}{\eta}\right], \frac{\delta}{2}\left[1-\beta(f)\right]\right) > 0. \tag{4.1.47}$$

To conclude, take $\eta < \min(\eta_-, \eta_+)$ and $\delta < \min(\delta_-(\eta), \delta_+(\eta))$, to obtain

$$\langle X, (J+\delta I)A^\varepsilon(u)X\rangle \geq \kappa \|X\|_2^2, \tag{4.1.48}$$

where $\kappa := \min(\kappa_-, \kappa_+) > 0$. □

We are now ready to prove uniqueness.

Proposition 4.1.4. *Under Condition 4.0.1, there exists at most one solution in* $C^{1+\alpha/4,2+\alpha/2}$ *to Problem* (4.1.3).

Proof. Let p and q be two solutions in $(C^{1+\alpha/4,2+\alpha/2})^N$. By integrating by parts, we have the estimate

$$\|\partial_x p_n\|_{L^2}^2 \leq \|\partial_{xx} p_n\|_{L^\infty} \|p_n\|_{L^1},\ 1 \leq n \leq N \tag{4.1.49}$$

we see that $\partial_x p, \partial_x q \in C(0,T,L^2)^N$.

Let $\Gamma \in S^{++}$ and $\kappa > 0$ be given by Proposition 4.1.3, such that for all $u \in (\mathbb{R}_+)^N$

$$\langle X, \Gamma A^\varepsilon(u)X\rangle \geq \kappa \|X\|_2^2. \tag{4.1.50}$$

Set $p' = \sqrt{\Gamma}p$ and $q' = \sqrt{\Gamma}q$. p' and q' are in $C^{1+\alpha/4,2+\alpha/2}$, and their spatial gradient in $C(0,T,L^2)$, and they solve respectively the system of PDEs

$$\begin{cases} \partial_t p' = \frac{1}{2}\partial_x[\sigma^2 A'(p)\partial_x p'] + +\partial_x[(\sigma\partial_x\sigma + \frac{1}{2}\sigma^2)B'(p)p'],\ (t,x) \in [0,T]\times\mathbb{R}, \\ p'_n(0,x) = \sqrt{\Gamma}P_n(x),\ x \in \mathbb{R}, 1 \leq n \leq N, \end{cases} \tag{4.1.51}$$

and

$$\begin{cases} \partial_t q' = \frac{1}{2}\partial_x[\sigma^2 A'(q)\partial_x q'] + \partial_x[(\sigma\partial_x\sigma + \frac{1}{2}\sigma^2)B'(q)q'], \ (t,x) \in [0,T]\times\mathbb{R}, \\ q'_n(0,x) = \sqrt{\Gamma}P_n(x), \ x\in\mathbb{R}, 1\le n\le N, \end{cases} \tag{4.1.52}$$

where the operators A' and B' are respectively defined by $A'(u) := \sqrt{\Gamma}A^\varepsilon(u)\sqrt{\Gamma}^{-1}$ and $B'(u) := \sqrt{\Gamma}B^\varepsilon(u)\sqrt{\Gamma}^{-1}$, $u \in (\mathbb{R}_+)^N$. A' satisfies the same coercivity property as A. Indeed, for all $X\in\mathbb{R}$ and $u\in(\mathbb{R}_+)^N$

$$\langle\sqrt{\Gamma}X, \sqrt{\Gamma}A^\varepsilon(u)\sqrt{\Gamma}^{-1}\sqrt{\Gamma}X\rangle = \langle X, \Gamma A^\varepsilon(u)X\rangle \ge \kappa\|X\|_2^2 \ge \tilde{\kappa}\|\sqrt{\Gamma}X\|_2^2, \tag{4.1.53}$$

where $\tilde{\kappa} := \frac{\kappa}{N+\delta}$, $(N+\delta)^{-1}$ being the smallest eigenvalue of Γ^{-1}.

Multiplying (4.1.51) by $p'-q'$ and integrating gives for all $t\in[0,T]$

$$\begin{aligned} \int\langle p'-q', \partial_t p'\rangle = &-\frac{1}{2}\int\langle\partial_x(p'-q'), A'(p)\partial_x p'\rangle \\ &-\int\left(\sigma\partial_x\sigma + \frac{1}{2}\sigma^2\right)\langle\partial_x(p'-q'), B'(p)p'\rangle. \end{aligned} \tag{4.1.54}$$

A similar equation holds for q'. Taking the difference with the previous equation and integrating over $[0,t]$ gives

$$\begin{aligned} \frac{1}{2}\int\|p'-q'\|_2^2 = &-\frac{1}{2}\int_0^t\int\langle\partial_x(p'-q'), A'(p)\partial_x p' - A'(q)\partial_x q'\rangle dt \\ &-\int_0^t\int\left(\sigma\partial_x\sigma + \frac{1}{2}\sigma^2\right)\langle\partial_x(p'-q'), B'(p)p' - B'(q)q'\rangle dt. \end{aligned} \tag{4.1.55}$$

Rewrite the integrand of the first integral of the right-hand side as

$$\begin{aligned}\langle \partial_x(p'-q'), A'(p)\partial_x p' - A'(q)\partial_x q'\rangle &= \langle \partial_x(p'-q'), A'(p)\partial_x(p'-q')\rangle \\ &+ \langle \partial_x(p'-q'), (A'(p)-A'(q))\partial_x q'\rangle \\ &\geq -\tilde{\kappa}\|\partial_x(p'-q')\|_2^2 \\ &- C\|q\|_{(C^{1+\alpha/2,2+\alpha})^N}\|\partial_x(p'-q')\|_2\|A'(p)-A'(q)\|_2.\end{aligned} \tag{4.1.56}$$

Likewise for the second integral

$$\begin{aligned}\sigma\partial_x\sigma\langle \partial_x(p'-q'), B'(p)p' - B'(q)q'\rangle &= \sigma\partial_x\sigma\langle \partial_x(p'-q'), B'(p)(p'-q')\rangle \\ &+ \sigma\partial_x\sigma\langle \partial_x(p'-q'), (B'(p)-B'(q))q'\rangle \\ &\geq -C\|\partial_x(p'-q')\|_2\|p'-q'\|_2 \\ &- C\|q\|_{(C^{1+\alpha/2,2+\alpha})^N}\|\partial_x(p'-q')\|_2\|B'(p)-B'(q)\|_2,\end{aligned} \tag{4.1.57}$$

where we used the boundedness of $B'(p)$, σ and $\partial_x\sigma$.

We check easily the existence of a constant C depending on Γ, $f_{\max}$, $f_{\min}$ and ε such that

$$\|A'(p)-A'(q)\|_2^2 + \|B'(p)-B'(q)\|_2^2 \leq C\|p'-q'\|_2^2. \tag{4.1.58}$$

In the view of the upper-bounds

$$\|\partial_x(p'-q')\|_2\|A'(p)-A'(q)\|_2 \leq \frac{\kappa}{4}\|\partial_x(p'-q')\|_2^2 + \frac{4}{\kappa}\|A'(p)-A'(q)\|_2^2, \tag{4.1.59}$$

$$\|\partial_x(p'-q')\|_2\|p'-q'\|_2 \leq \frac{\kappa}{4}\|\partial_x(p'-q')\|_2^2 + \frac{4}{\kappa}\|p'-q'\|_2^2, \tag{4.1.60}$$

and

$$\|\partial_x(p'-q')\|_2\|B'(p)-B'(q)\|_2 \leq \frac{\kappa}{4}\|\partial_x(p'-q')\|_2^2 + \frac{4}{\kappa}\|B'(p)-B'(q)\|_2^2, \tag{4.1.61}$$

we conclude that for all $t \in [0,T]$ and some constant $C > 0$

$$\int \|p' - q'\|_2^2 \leq C\|q\|^2_{(C^{1+\alpha/4,2+\alpha/2})^N} \int_0^t \int \|p' - q'\|_2^2 dt. \tag{4.1.62}$$

Gronwall's inequality yields $p' = q'$ a.e. on $[0,T] \times \mathbb{R}$ and by continuity of p and q, we have immediately that $p = q$.

$\square$

4.2 Propagation of chaos

In this section, we prove the propagation of chaos and introduce an intermediate mollified version of SDE (4.0.4). For each $M \geq 1$, let $(\widetilde{X}_t^{i,M})_{i\geq 1, t\geq 0}$ be a particle system such that the i-th particle starts at X_0^i and evolves according to the dynamics

$$\begin{cases} d\widetilde{X}_t^{i,M} &= -\dfrac{1}{2}\dfrac{\varepsilon + f(Y^i)\sum_{n=1}^N W_{\delta_M} * \widetilde{p}_n(t, \widetilde{X}_t^{i,M})}{\varepsilon + \sum_{n=1}^N f(n) W_{\delta_M} * \widetilde{p}_n(t, \widetilde{X}_t^{i,M})}\sigma(t, \widetilde{X}_t^{i,M})^2 dt \\ & \quad + \sqrt{\dfrac{\varepsilon + f(Y^i)\sum_{n=1}^N W_{\delta_M} * \widetilde{p}_n(t, \widetilde{X}_t^{i,M})}{\varepsilon + \sum_{n=1}^N f(n) W_{\delta_M} * \widetilde{p}_n(t, \widetilde{X}_t^{i,M})}}\sigma(t, \widetilde{X}_t^{i,M}) dB_t^i, \\ \mathbb{P}\left(\widetilde{X}_t^{i,M} \in dx \cap Y^i = n\right) = \widetilde{p}_n(t,x)dx, \end{cases}, \tag{4.2.1}$$

where we denote

$$W_{\delta_M} * \phi(x) = \int W_{\delta_M}(x-y)\phi(y)dy, \quad \phi \in L^2(\mathbb{R}), \ x \in \mathbb{R}. \tag{4.2.2}$$

We introduce the PDE system associated to SDE (4.2.1)

$$\begin{cases} \partial_t \widetilde{p} = \frac{1}{2}\partial_{xx}\left[\sigma^2 B^\varepsilon(W * \widetilde{p})\widetilde{p}\right] + \frac{1}{2}\partial_x\left[\sigma^2 B^\varepsilon(W * \widetilde{p})\widetilde{p}\right], \ (t,x) \in [0,T] \times \mathbb{R}, \\ \widetilde{p}_n(0,x) = P_n(x), \ x \in \mathbb{R}, \\ 1 \leq n \leq N. \end{cases}. \tag{4.2.3}$$

The existence of $(\widetilde{X}_t^{i,M})_{1\leq i\leq M, t\geq 0}$ is ensured by the following proposition.

Proposition 4.2.1. *If $(P_n)_{1\leq n\leq N}$ satisfies the assumption of Theorem 4.0.2, then there exists a strong solution $(\widetilde{X}_t^{i,M}, \widetilde{p}(t,\cdot))_{t\geq 0}$ to SDE (4.0.11). Moreover, $(\widetilde{p}_n)_{1\leq n\leq N}$ belongs to the space $(C^{1+\alpha/4,2+\alpha/2})^N \cap C(0,T,L^2)^N$ and there exists a constant $C>0$ independent of M and δ_M such that*

$$\sum_{n=1}^{N} \|\tilde{p}_n\|_{C^{1+\alpha/4,2+\alpha/2}} \leq C, \tag{4.2.4}$$

and

$$\sup_{0\leq t\leq T} \int \|\widetilde{p} - \hat{p}\|_2^2 \leq C\delta_M, \ \forall t\geq 0. \tag{4.2.5}$$

Proof. The existence of $(\tilde{p}_n)_{1\leq n\leq N}$ and $(\tilde{X}_t^{i,M})_{t\geq 0}, 1\leq i\leq M$, can easily be established by noticing that the mapping $\phi \in C^{1+\alpha/4,2+\alpha/2} \mapsto W_{\delta_M} * u \in C^{1+\alpha/4,2+\alpha/2}$ is non-expansive for the norm $\|\cdot\|_{C^{1+\alpha/4,2+\alpha/2}}$ and by using the techniques of the proof of Proposition 4.1.1 (see as well [65, Proposition 2.2]). The non-expansivity of the mentioned mapping makes the estimate (4.1.7) still valid and independent of δ_M, and therefore (4.2.4) is justified.

In order to establish (4.2.5), we shall show that the proof of [65, Lemma 2.6] carries to parabolic systems. Mimicking the computation of [65, Equation (2.8)], we can write that the difference $\hat{p} - \widetilde{p}$ solves the system

$$\partial_t(\hat{p} - \widetilde{p}) = \frac{1}{2}\partial_x[A^{\varepsilon}(\hat{p})\partial_x(\hat{p} - \widetilde{p})] + a_1\partial_x(\hat{p} - \widetilde{p}) + a_2(\hat{p} - \widetilde{p}) + \phi. \tag{4.2.6}$$

where the matrix valued functions $a_1(t,x)$ and $a_2(t,x)$ are bounded, depend on $\hat{p}$ and $\tilde{p}$, and the vector valued function $\phi(t,x)$ depends on $\widetilde{p} - W * \widetilde{p}$ and its first and second derivatives. Moreover, ϕ satisfies

$$\sum_{n=1}^{N} \sup_{x\in\mathbb{R}} |\phi_n(x)| \leq C\delta_M, \tag{4.2.7}$$

for some constant $C>0$ independent of δ_M.

Next, we observe that under Condition 4.0.1, there exist $\Gamma \in S^{++}$ and $\kappa > 0$ such that for all $X \in \mathbb{R}^N$

$$\langle X, \Gamma A^\varepsilon(\hat{p})X\rangle \geq \kappa \|X\|_2^2. \tag{4.2.8}$$

Like in the proof of Proposition 4.1.4, set $\hat{p}' = \sqrt{\Gamma}\hat{p}$, $\widetilde{p}' = \sqrt{\Gamma}\widetilde{p}$, $A'(\hat{p}) = \sqrt{\Gamma}A^\varepsilon(\hat{p})\sqrt{\Gamma}^{-1}$, $a_1' = \sqrt{\Gamma}a_1\sqrt{\Gamma}^{-1}$, $a_2' = \sqrt{\Gamma}a_2\sqrt{\Gamma}^{-1}$ and $\phi' = \sqrt{\Gamma}\phi$. With (4.1.53), there exists $\kappa' > 0$ such that for all $X \in \mathbb{R}^N$

$$\langle X, A'X\rangle \geq \kappa' \|X\|_2^2. \tag{4.2.9}$$

Moreover

$$\begin{aligned}\frac{1}{2}\int \|\hat{p}' - \widetilde{p}'\|_2^2 = &-\frac{1}{2}\int_0^t \int \langle \partial_x(\hat{p}' - \widetilde{p}'), A'(\hat{p})\partial_x(\hat{p}' - \widetilde{p}')\rangle dt \\ &+ \int_0^t \int \langle \hat{p}' - \widetilde{p}', a_1'\partial_x(\hat{p}' - \widetilde{p}')\rangle dt + \int_0^t \int \langle \hat{p}' - \widetilde{p}', a_2(\hat{p}' - \widetilde{p}')\rangle dt \\ &+ \int_0^t \int \langle \hat{p}' - \widetilde{p}', \phi'\rangle dt.\end{aligned} \tag{4.2.10}$$

Using the boundedness of a_1', a_2' and standard techniques, it is easy to see that for some appropriate constant $C > 0$ and for all $t \in [0, T]$

$$\begin{aligned}&\int \langle \hat{p}' - \widetilde{p}', a_1'\partial_x(\hat{p}' - \widetilde{p}')\rangle + \int \langle \hat{p}' - \widetilde{p}', a_2(\hat{p}' - \widetilde{p}')\rangle \\ &\leq \frac{\kappa'}{2}\int \|\partial_x(\hat{p}' - \widetilde{p}')\|_2^2 + C\int \|\hat{p}' - \widetilde{p}'\|_2^2.\end{aligned} \tag{4.2.11}$$

In the view of the uniform coercivity of A', and ϕ', and the estimate

$$\int \langle \hat{p}' - \widetilde{p}', \phi'\rangle \leq C\delta_M \sum_{n=1}^N \int (\hat{p}_n + \widetilde{p}_n) = C\delta_M, \tag{4.2.12}$$

we conclude that there exists some constant $C > 0$ such that for all $t \in [0, T]$

$$\frac{1}{2}\int \|\hat{p}' - \widetilde{p}'\|_2^2 \leq C\delta_M + C\int_0^t \frac{1}{2}\int \|\hat{p}' - \widetilde{p}'\|_2^2 dt. \tag{4.2.13}$$

The result is proved by applying Gronwall's inequality and using the obvious inequality

$\|\hat{p}-\widetilde{p}\|_2 \leq \|\hat{p}'-\widetilde{p}'\|_2$, for some constant $C>0$ depending only on Γ and independent of δ_M. $\square$

Let $1 \leq i \leq M$. To simplify the notation, we will drop the subscripts and superscripts in i and M, and we denote $\delta = \delta_M$, $W = W_{\delta_M}$, $X = X^i$, $\widetilde{X} = \widetilde{X}^i$, $\hat{X} = \hat{X}^i$, and for $j \neq i$, $X^j = X^{j,M}$ and $\widetilde{X}^j = \widetilde{X}^{j,M}$. Denote $W_{\delta_M} * X_t(x) = \frac{1}{M}\sum_{j=1}^M W_{\delta_M}(x - X_t^j)$ and $W^f_{\delta_M} * X_t(x) = \frac{1}{M}\sum_{j=1}^M f(Y^j) W_{\delta_M}(x - X_t^j)$. Under this notation, the volatility coefficients of X, $\widetilde{X}$ and $\hat{X}$ are respectively given by

$$\Sigma_t := f(Y)\sqrt{\frac{\varepsilon + W * X_t(X_t)}{\varepsilon + W^f * X_t(X_t)}}\sigma(t, X_t), \tag{4.2.14}$$

$$\widetilde{\Sigma}_t := \sqrt{\frac{\varepsilon + f(Y)\sum_{n=1}^N W * \widetilde{p}_n(t, \widetilde{X}_t)}{\varepsilon + \sum_{n=1}^N f(n) W * \widetilde{p}_n(t, \widetilde{X}_t)}}\sigma(t, \widetilde{X}_t), \tag{4.2.15}$$

and

$$\hat{\Sigma}_t := \sqrt{\frac{\varepsilon + f(Y)\sum_{n=1}^N \hat{p}_n(t, \hat{X}_t)}{\varepsilon + \sum_{n=1}^N f(n)\hat{p}_n(t, \widetilde{X}_t)}}\sigma(t, \hat{X}_t). \tag{4.2.16}$$

Notice immediately that

$$\sigma_0^2 \frac{\inf f}{\sup f} \leq \Sigma^2, \widetilde{\Sigma}^2, \hat{\Sigma}^2 \leq \frac{\sup f}{\inf f}\sigma_1^2. \tag{4.2.17}$$

In the rest of the section, we denote, for notational simplicity, by C any constant depending on N, ε, f, σ_0, σ_1, and $|W_1|_\infty$.

The proof of Theorem 4.0.3 requires the Propositions 4.2.2 and 4.2.3 below.

Proposition 4.2.2. *Let $1 \leq i \leq M$. Then for all $T > 0$.*

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} (X_t^{i,M} - \widetilde{X}_t^{i,M})^2\right] \leq \frac{C}{M}\frac{\exp\left(CT\delta_M^{-4}\right)}{\delta_M^2}. \tag{4.2.18}$$

Proof. According to the Burkholder–Davis–Gundy inequality, for any $s \in [0,T]$

$$\begin{aligned} \mathbb{E}\left[\sup_{0\le t\le s}(X_t - \widetilde{X}_t)^2\right] &\le C\mathbb{E}\left[\langle X - \widetilde{X}\rangle_s\right] \\ &= C\int_0^s \mathbb{E}\left[(\Sigma_t - \widetilde{\Sigma}_t)^2\right] dt. \end{aligned} \tag{4.2.19}$$

For $x \in \mathbb{R}$ and $(U^j)_{1\le j\le M} \in \mathbb{R}^M$, we abbreviate $\frac{1}{M}\sum_{j=1}^M W(x - U^j)$ by $W * U(x)$ and $\frac{1}{M}\sum_{j=1}^M f(Y^j)W(x - U^j)$ by $W^f * U(x)$.

The boundedness Σ_t, $\widetilde{\Sigma}_t$, σ and $\partial_x\sigma$ gives for all $t \in [0,T]$

$$\begin{aligned} |\Sigma_t - \widetilde{\Sigma}_t| &\le C|\Sigma_t^2 - \widetilde{\Sigma}_t^2| \\ &\le C\left|\frac{\varepsilon + W^f * X_t(X_t)}{\varepsilon + W * X_t(X_t)} - \frac{\varepsilon + \sum_{n=1}^N f(n)W * \widetilde{p}_n(t,\widetilde{X}_t)}{\varepsilon + \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)}\right| \\ &\quad + C|\sigma(t,X_t) - \sigma(t,\widetilde{X}_t)| \\ &\le C\left|\frac{\varepsilon + W^f * X(X_t)}{\varepsilon + W * X_t(X_t)} - \frac{\varepsilon + W^f * X_t(X_t)}{\varepsilon + \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)}\right| \\ &\quad + C\left|\frac{\varepsilon + W^f * X(X_t)}{\varepsilon + \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)} - \frac{\varepsilon + \sum_{n=1}^N f(n)W * \widetilde{p}_n(t,\widetilde{X}_t)}{\varepsilon + \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)}\right| \\ &\quad + C|X_t - \widetilde{X}_t| \\ &\le C\left|W * X_t(X_t) - \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)\right| \\ &\quad + C\left|W^f * X_t(X_t) - \sum_{n=1}^N f(n)W * \widetilde{p}_n(t,\widetilde{X}_t)\right| + C|X_t - \widetilde{X}_t|. \end{aligned} \tag{4.2.20}$$

Let $t \in [0,T]$. After estimating the term $\mathbb{E}\left[\left|W^f * X_t(X_t) - \sum_{n=1}^N f(n)W * \widetilde{p}_n(t,\widetilde{X}_t)\right|^2\right]$, we will apply the estimate with $f = 1$ to treat the term $\mathbb{E}\left[\left|W * X_t(X_t) - \sum_{n=1}^N W * \widetilde{p}_n(t,\widetilde{X}_t)\right|^2\right]$.

Upper-bound

$$\left|W^f * X_t(X_t) - \sum_{n=1}^N f(n)W * p_n(t,\widetilde{X}_t)\right| \le \text{(I)} + \text{(II)} + \text{(III)}, \tag{4.2.21}$$

where each term (I), (II) and (III) is respectively defined by

$$(\mathrm{I}) := \left| W^f * X_t(X_t) - W^f * X_t(\widetilde{X}_t) \right|, \tag{4.2.22}$$

$$(\mathrm{II}) := \left| W^f * X_t(\widetilde{X}_t) - W^f * \widetilde{X}_t(\widetilde{X}_t) \right|, \tag{4.2.23}$$

and

$$(\mathrm{III}) := \left| W^f * \widetilde{X}_t(\widetilde{X}_t) - \sum_{n=1}^{N} f(n) W * \widetilde{p}_n(t, \widetilde{X}_t) \right|. \tag{4.2.24}$$

By using $\sup_{x\in\mathbb{R}} |W'(x)| \leq \frac{C}{\delta^2}$, we find

$$\begin{aligned} \mathbb{E}\left[(\mathrm{I})^2\right] &= \mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M} \mathbb{E}\left[f(Y^j)(W(X_t - X_t^j) - W(\widetilde{X}_t - X_t^j))\right]\right)^2\right] \\ &\leq \frac{C}{\delta^4}\mathbb{E}\left[|X_t - \widetilde{X}_t|^2\right]. \end{aligned} \tag{4.2.25}$$

Likewise

$$\begin{aligned} \mathbb{E}\left[(\mathrm{II})^2\right] &= \mathbb{E}\left[\left(\frac{1}{M}\sum_{j=1}^{M} f(Y^j)(W(\widetilde{X}_t - X_t^j) - W(\widetilde{X}_t - \widetilde{X}_t^j))\right)^2\right] \\ &\leq \frac{C}{\delta^4 M}\mathbb{E}\left[\left(\sum_{j=1}^{M} |X_t^j - \widetilde{X}_t^j|^2\right)\right] \\ &\leq \frac{C}{\delta^4}\mathbb{E}\left[|X_t - \widetilde{X}_t|^2\right], \end{aligned} \tag{4.2.26}$$

where we used that by symmetry $\mathbb{E}\left[|X_t^j - \widetilde{X}_t^j|^2\right] = \mathbb{E}\left[|X_t - \widetilde{X}_t|^2\right]$ for all $j = 1, \ldots, M$.

By independence of the $(\widetilde{X}_t^j)_{1\leq j\leq M}$ and using that

$$\sum_{n=1}^{N} f(n) W * \widetilde{p}_n(t, x) = \mathbb{E}\left[f(Y^j) W(x - \widetilde{X}_t^j)\right], \; 1 \leq j \neq i \leq N, \tag{4.2.27}$$

we get

$$\mathbb{E}\left[(\mathrm{III})^2\right] = \int \mathbb{E}\left[\left(\frac{1}{M}\sum_{j\neq i} f(Y^j)W(x-\widetilde{X}_t^j) - \mathbb{E}\left[f(Y^j)W(x-\widetilde{X}_t^j)\right]\right)^2\right] \times\mathbb{P}\left(\widetilde{X}_t \in dx\right). \tag{4.2.28}$$

Therefore

$$\mathbb{E}\left[(\mathrm{III})^2\right] \leq C\int \mathbb{E}\left[\left(\frac{1}{M}\sum_{j\neq i}\left(f(Y^j)W(x-\widetilde{X}_t^j) - \mathbb{E}\left[f(Y^j)W(x-\widetilde{X}_t^j)\right]\right)\right)^2\right]\mathbb{P}\left(\widetilde{X}_t \in dx\right) + \frac{C}{M^2}\int \mathbb{E}\left[f(Y)W(x-\widetilde{X}_t)\right]^2\mathbb{P}\left(\widetilde{X}_t \in dx\right). \tag{4.2.29}$$

Making use of the bound $\sup_{x\in\mathbb{R}}|W(x)| \leq \frac{C}{\delta}$, we see that the second integral in the right-hand side of (4.2.29) is dominated by

$$\int \mathbb{E}\left[f(Y)W(x-\widetilde{X}_t)\right]^2\mathbb{P}\left(\widetilde{X}_t \in dx\right) \leq C\int W(x-y)^2\mathbb{P}\left(\widetilde{X}_t \in dx\right)\mathbb{P}\left(\widetilde{X}_t \in dy\right) \leq \frac{C}{\delta^2}. \tag{4.2.30}$$

In estimating the first integral of (4.2.29), we utilize the independence of the $(\tilde{X}_t^j)_{1\leq j\leq N}$, for all $x\in\mathbb{R}$, which gives

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{M}\sum_{j\neq i}\left(f(Y^j)W(x-\widetilde{X}_t^j) - \mathbb{E}\left[f(Y^j)W(x-\widetilde{X}_t^j)\right]\right)\right)^2\right] \\ &= \mathrm{Var}\left(\frac{1}{M}\sum_{j\neq i} f(Y^j)W(x-\widetilde{X}_t^j)\right) \\ &= \frac{M-1}{M^2}\mathrm{Var}\left(f(Y)W(x-\widetilde{X}_t)\right) \\ &\leq \frac{C}{M\delta^2}. \end{aligned} \tag{4.2.31}$$

Putting together (4.2.25), (4.2.26) and (4.2.29)-(4.2.31), we infer

$$\mathbb{E}\left[\left|W^f * X_t(X_t) - \sum_{n=1}^{N} f(n) W * \widetilde{p}_n(t, \widetilde{X}_t)\right|^2\right] \leq \frac{C}{M\delta^2} + \frac{C}{\delta^4}\mathbb{E}\left[|X_t - \widetilde{X}_t|^2\right]. \tag{4.2.32}$$

Taking $f = 1$ leads to the same estimate for $\mathbb{E}\left[\left|W * X_t(X_t) - \sum_{n=1}^{N} W * \widetilde{p}_n(t, \widetilde{X}_t)\right|^2\right]$.

If we insert this in (4.2.20) and use (4.2.19), we obtain

$$\mathbb{E}\left[\sup_{0\leq s\leq t}(X_s - \widetilde{X}_s)^2\right] \leq \frac{1}{M}\frac{C}{\delta^2} + \frac{C}{\delta^4}\int_0^t \mathbb{E}\left[\sup_{0\leq s\leq r}(X_s - \widetilde{X}_s)^2\right] dr. \tag{4.2.33}$$

Applying Gronwall's inequality proves (4.2.18) and concludes the proof of the proposition. □

Proposition 4.2.3. *Let $1 \leq i \leq M$. Then for some constant $C > 0$ independent of δ*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\widetilde{X}_t^{i,M} - \hat{X}_t^i|^2\right] \leq C\delta_M. \tag{4.2.34}$$

Proof. In the same manner of (4.2.20), for some constant $C > 0$

$$|\widetilde{\Sigma}_t - \hat{\Sigma}_t| \leq \frac{C}{\delta}\sum_{n=1}^{N}\left|W * \widetilde{p}_n(t, \widetilde{X}_t) - \hat{p}_n(t, \hat{X}_t)\right| + C|\widetilde{X}_t - \hat{X}_t|. \tag{4.2.35}$$

For all $t \in [0, T]$

$$\begin{aligned}\left|W * \widetilde{p}_n(t, \widetilde{X}_t) - \hat{p}_n(t, \hat{X}_t)\right| \leq &\left|W * \hat{p}_n(t, \hat{X}_t) - \hat{p}_n(t, \hat{X}_t)\right| \\ &+ \left|W * \hat{p}_n(t, \widetilde{X}_t) - W * \hat{p}_n(t, \hat{X}_t)\right| \\ &+ \left|W * \widetilde{p}_n(t, \widetilde{X}_t) - W * \hat{p}_n(t, \widetilde{X}_t)\right|.\end{aligned} \tag{4.2.36}$$

Next, using the boundedness of the gradient of $\hat{p}_n$, we estimate the first two terms of the left-hand side of (4.2.36) as follows

$$\begin{aligned} \left|W * \hat{p}_n(t, \hat{X}_t) - \hat{p}_n(t, \hat{X}_t)\right| &\leq \sup_{x\in\mathbb{R}} \int W_1(y)|\hat{p}_n(t, x-\delta y) - \hat{p}_n(t,x)|dy \\ &\leq \delta \|\hat{p}_n\|_{C^{1+\alpha/2,2+\alpha}} \int |y| W_1(y) dy, \end{aligned} \tag{4.2.37}$$

and

$$\left|W * \hat{p}_n(t, \widetilde{X}_t) - W * \hat{p}_n(t, \hat{X}_t)\right| \leq \|\hat{p}_n\|_{C^{1+\alpha/2,2+\alpha}} |\widetilde{X}_t - \hat{X}_t|. \tag{4.2.38}$$

Therefore, there exists a constant $C > 0$ independent of δ such that

$$\left|W * \hat{p}_n(t, \hat{X}_t) - \hat{p}_n(t, \hat{X}_t)\right| + \left|W * \hat{p}_n(t, \widetilde{X}_t) - W * \hat{p}_n(t, \hat{X}_t)\right| \leq C(\delta + |\widetilde{X}_t - \hat{X}_t|). \tag{4.2.39}$$

For the last term of (4.2.36), write

$$\begin{aligned} \mathbb{E}\left[\left|W * \widetilde{p}_n(t, \widetilde{X}_t) - W * \hat{p}_n(t, \widetilde{X}_t)\right|^2\right] &= \sum_{m=1}^{N} \int |W * (\widetilde{p}_n - \hat{p}_n)|^2 p_m \\ &\leq \|p\|_{(C^{1+\alpha/4,2+\alpha/2})^N} \int |W * (\widetilde{p}_n - \hat{p}_n)|^2 \\ &\leq \|p\|_{(C^{1+\alpha/4,2+\alpha/2})^N} \int |\widetilde{p}_n - \hat{p}_n|^2 \\ &\leq \|p\|_{(C^{1+\alpha/4,2+\alpha/2})^N} \delta, \end{aligned} \tag{4.2.40}$$

where we used Proposition 4.2.1.

Combining the previous estimate and (4.2.36) we deduce the existence of $C > 0$ independent of δ such that

$$\mathbb{E}\left[|\widetilde{\sigma}_t - \hat{\sigma}_t|^2\right] \leq C\left(\delta^2 + \mathbb{E}\left[|\widetilde{X}_t - \hat{X}_t|^2\right]\right). \tag{4.2.41}$$

Burkholder–Davis–Gundy inequality and Gronwall's inequality yield

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\widetilde{X}_t - \hat{X}_t|^2\right] \leq C\delta. \tag{4.2.42}$$

□

We are now ready for the proof of Theorem 4.0.3.

Proof of Theorem 4.0.3. According to Propositions 4.2.2 and 4.2.3 above, there exists a constant $C > 0$ such that for all $1 \leq i \leq N$

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left(X_t^{i,M} - \hat{X}_t^{i,M}\right)^2\right] \leq \frac{C}{M}\frac{\exp\left(CT\delta_M^{-4}\right)}{\delta_M^2} + C\delta_M. \tag{4.2.43}$$

It is enough to choose $(\delta_M)_{M\geq 1}$ such that $\lim_{M\to\infty}\delta_M = 0$ and

$$\lim_{M\to\infty}\frac{1}{M}\frac{\exp\left(CT\delta_M^{-4}\right)}{\delta_M^2} = 0. \tag{4.2.44}$$

□

Chapter 5

Neural joint S&P 500/VIX smile calibration

In this article we tackle the problem of jointly calibrating an arbitrage-free model on the Standard & Poor's 500 Index (SPX) to SPX options, Chicago Board Options Exchange's Volatility Index (VIX) futures, and VIX options. This is known to be a difficult problem, especially for short maturities, which had eluded quants for many years (see [50]). While parametric models have produced approximate fits (see, e.g., [3, 8, 9, 21, 36, 73, 96, 97, 39, 99]), the first exact solution[1] came with the nonparametric discrete-time model of [50], whose minimum-entropy technique can be seen as a nonlinear optimal transport approach. This approach was later extended to continuous time to produce jointly calibrating nonparametric diffusive models [43, 48]. In this article, we also consider overparametrized diffusive models to solve the joint calibration problem, but rather than casting it as a nonlinear optimal transport problem, we solve it using neural stochastic differential equations (SDEs).

Neural SDEs, which have been introduced or used in [108, 64, 60, 40, 23, 71], are of the nonparametric (or overparametrized) type: they are SDEs whose drift and diffusion coefficients are chosen to be neural networks, depending on potentially many more

[1] in the sense that, as opposed to a best parametric fit, the numerical method converges to an exact fit

parameters than there are options and futures to calibrate to. By the universal approximation theorem for neural networks, neural SDEs have the potential of approximating any SDE. By minimizing a loss function over the many parameters, we build a diffusive model that solves the joint calibration problem accurately, often within bid-ask spreads in our numerical tests. In this article we only consider two-dimensional Markovian neural SDEs. The first component X is the log-spot, while the second component Y drives the instantaneous volatility together with X. The main benefit of overparametrization is that it offers a lot of flexibility; the main drawback of our approach is the lack of interpretability of the parameters and of process Y. Note, however, that even in popular stochastic volatility models, such as the Bergomi models, the instantaneous volatility is modeled as a function of (typically, Ornstein-Uhlenbeck) factors that do not have a clear financial interpretation either.

Our main contribution is that we show that a general one-factor stochastic local volatility (SLV) model can solve the joint calibration problem accurately, often within bid-ask spreads in our numerical tests, provided we allow for enough flexibility on the drift and diffusion coefficients, all functions of (t, X_t, Y_t). Note that [43] uses an $(n+1)$-dimensional SDE, where n is the number of calibrated VIX expiries, while the Schrödinger bridge approach of [48] requires fine-tuning the volatility-of-volatility coefficient, which is not optimized upon. Our model shows that the joint calibration problem can be accurately solved with a two-dimensional Markovian SDE, regardless of the number of calibrated VIX expiries, and with no need of choosing a volatility-of-volatility. Interestingly, the joint calibration actually forces the SLV model to be a pure path-dependent volatility (PDV) model, confirming the findings in [49]: the "spot-vol" correlation is pushed to its lower bound -1. PDV models have recently been shown to be good candidates for approximately solving the joint calibration problem [39, 55].

The natural practical application of our model is the pricing and hedging of structured products by exotics desks. With this model, the pricing and hedging of structured products on the SPX indeed takes into account the whole information given by SPX smiles (the risk-neutral distributions of future SPX values) as well as the whole information brought by VIX futures and VIX smiles (the risk-neutral distributions of some future SPX implied volatilities). Once the model is calibrated overnight (3 hours in our tests), it is extremely simple to implement and use, as it is a Markov model in very low dimension (two).

The rest of this chapter is structured as follows. We describe our neural SDE model in Section 5.1. The neural joint calibration procedure is explained in Section 5.2. Finally, implementation details and numerical results are reported in Sections 5.3 and 5.4 respectively.

This chapter is based on [56].

5.1 The model

Let $(T^s_j)_{1\le j\le N^s_T}$ (resp. $(T^v_j)_{1\le j\le N^v_T}$) be a collection of ordered SPX (resp. VIX) options expiries. Define

$$T = \max(T^s_{N^s_T}, T^v_{N^v_T} + \tau), \tag{5.1.1}$$

where $\tau = \frac{30}{365}$ (30 days). We aim at building a model that jointly calibrates the SPX smiles at $(T^s_j)_{1\le j\le N^s_T}$ and the VIX futures and the VIX smiles at $(T^v_j)_{1\le j\le N^v_T}$. Let $(S_t)_{t\ge 0}$ denote the price of the SPX. We assume deterministic interest rates r_t and repo q_t, inclusive of dividend yield, and for $u \ge t$ we denote by

$$f_{t,u} := S_t \exp\left(\int_t^u (r_s - q_s)ds\right), \tag{5.1.2}$$

the SPX u-forward at time t. We denote by $X = (\log S_t/f_{0,t})_{t\ge 0}$ the log-price.

We consider a general SLV model. Denote Y a stochastic process driving the volatility together with X. The SLV model is initialized at $(0,0)$ at time 0 and follows the following risk-neutral dynamics for $t \in [0,T]$:

$$\begin{cases} dX_t = -\frac{1}{2}\sigma_X(t,X_t,Y_t)^2 dt + \sigma_X(t,X_t,Y_t)dB_t^1, \\ dY_t = \mu_Y(t,X_t,Y_t)dt + \sigma_Y(t,X_t,Y_t)\big(\rho(t,X_t,Y_t)dB_t^1, \\ \qquad\qquad\qquad\qquad + \sqrt{1-\rho(t,X_t,Y_t)^2}dB_t^2\big) \end{cases} \tag{M}$$

where $(B_t^1)_{t\geq 0}$ and $(B_t^2)_{t\geq 0}$ are two independent Brownian motions, σ_X is the volatility of the log-price, μ_Y is the drift of Y and σ_Y the volatility of Y, and ρ is the correlation between the two Brownian motions driving the dynamics of X and Y.

In this Markov model, with classical notations, the VIX^2 at t is given by

$$\begin{aligned} \text{VIX}_t^2 := -\frac{2}{\tau}\mathbb{E}\left[\log \frac{S_{t+\tau}}{f_{t,t+\tau}}\middle| \mathcal{F}_t\right] &= -\frac{2}{\tau}\mathbb{E}\left[X_{t+\tau} - X_t\middle| \mathcal{F}_t\right] \\ &= \mathbb{E}\left[R_t|X_t,Y_t\right] =: v(t,X_t,Y_t), \end{aligned}$$

where R_t

$$R_t := \frac{1}{\tau}\int_t^{t+\tau} \sigma_X(s,X_s,Y_s)^2 ds. \tag{5.1.3}$$

is the realized variance over the 30-day period. The VIX is defined by $\text{VIX}_t = \sqrt{\text{VIX}_t^2}$.

5.2 Neural calibration

Let $D_j^{\text{SPX}} = \{(k, I^{\text{SPX}}(T_j^s,k)) : k \in K_j^{\text{SPX}}\}$ be a collection of strikes and out-the-money (OTM) implied volatilities[2] for the SPX at T_j^s. Similarly, let $D_j^{\text{VIX}} = \{(k, C^{\text{VIX}}(T_j^v,k), P^{\text{VIX}}(T_j^v,k), I_j^{\text{VIX}}(T_j^v,k)) : k \in K_j^{\text{VIX}}\}$ be a collection of target strikes, call prices, put prices, and OTM implied volatilities for the VIX at T_j^v; the market VIX future at T_j^v is denoted $f\text{VIX}(T_j^v)$; $P(t) = \exp\big(-\int_0^t r_s ds\big)$ denotes the price of the zero coupon of maturity $t \geq 0$.

For each strike k and maturity t, we define the price of the call and put of strike k

[2]By OTM implied volatility we mean the implied volatility of the OTM call or put option. The forward value is computed via call-put parity so that implied volatilities behave smoothly around the money.

and maturity t, respectively, under model (M) as

$$\begin{aligned} C_m^{\mathrm{SPX}}(t,k) &= P(0,t)\mathbb{E}\left[(S_t-k)_+\right], \\ P_m^{\mathrm{SPX}}(t,k) &= P(0,t)\mathbb{E}\left[(k-S_t)_+\right] \end{aligned} \tag{5.2.1}$$

and we denote by $I_m^{\mathrm{SPX}}(t,k)$ the associated OTM implied volatility. Similarly, we denote the prices of call and put options on the VIX for strike k and maturity t by

$$\begin{aligned} C_m^{\mathrm{VIX}}(t,k) &= P(0,t)\mathbb{E}\left[(\mathrm{VIX}_t-k)_+\right], \\ P_m^{\mathrm{VIX}}(t,k) &= P(0,t)\mathbb{E}\left[(k-\mathrm{VIX}_t)_+\right]. \end{aligned} \tag{5.2.2}$$

The model price of the VIX future expiring at t is defined as $f\mathrm{VIX}_m(t) = \mathbb{E}\left[\mathrm{VIX}_t\right]$. In order to jointly calibrate Model (M) to the smiles $(D_j^{\mathrm{SPX}})_{1\leq j\leq N_T^s}$ and $(D_j^{\mathrm{VIX}})_{1\leq j\leq N_T^v}$ and to VIX futures, we look for

$$\sigma_X, \mu_Y, \sigma_Y, \rho \in \operatorname{argmin} L(\sigma_X, \mu_Y, \sigma_Y, \rho), \tag{5.2.3}$$

where the loss L is defined by

$$
\begin{aligned}
L(\sigma_X, \mu_Y, \sigma_Y, \rho) = {} & w_{f\mathrm{VIX}} \frac{1}{N_T^v} \sum_{j=1}^{N_T^v} \left(\frac{f\mathrm{VIX}_m(T_j^v)}{f\mathrm{VIX}(T_j^v)} - 1 \right)^2 \\
& + w_{\mathrm{SPX}} \frac{1}{N_T^s} \sum_{j=1}^{N_T^s} \frac{1}{|D_j^{\mathrm{SPX}}|} \\
& \times \sum_{k \in K_j^{\mathrm{SPX}}} \Delta^{\mathrm{SPX}}(T_j^s, k) \left(\frac{I_m^{\mathrm{SPX}}(T_j^s, k)}{I^{\mathrm{SPX}}(T_j^s, k)} - 1 \right)^2 \\
& + w_{\mathrm{VIX}} \frac{1}{N_T^v} \sum_{j=1}^{N_T^v} \frac{1}{|D_j^{\mathrm{VIX}}|} \\
& \times \sum_{\substack{k \in K_j^{\mathrm{VIX}} \\ k > f\mathrm{VIX}_m(T_j^v)}} \Delta^{\mathrm{VIX}}(T_j^v, k) \left(\frac{C_m^{\mathrm{VIX}}(T_j^v, k)}{C^{\mathrm{VIX}}(T_j^v, k)} - 1 \right)^2 \\
& + w_{\mathrm{VIX}} \frac{1}{N_T^v} \sum_{j=1}^{N_T^v} \frac{1}{|D_j^{\mathrm{VIX}}|} \\
& \times \sum_{\substack{k \in K_j^{\mathrm{VIX}} \\ k \leq f\mathrm{VIX}_m(T_j^v)}} \Delta^{\mathrm{VIX}}(T_j^v, k) \left(\frac{P_m^{\mathrm{VIX}}(T_j^v, k)}{P^{\mathrm{VIX}}(T_j^v, k)} - 1 \right)^2,
\end{aligned}
\tag{5.2.4}
$$

for some positive weights $w = (w_{f\mathrm{VIX}}, w_{\mathrm{SPX}}, w_{\mathrm{VIX}})$. Small bid-ask spreads are given more importance through the weights

$$
\begin{aligned}
\Delta^{\mathrm{SPX}}(T_j^s, k) &= \frac{\delta^{\mathrm{SPX}}(t, k)}{\sum_{l \in K_j^{\mathrm{SPX}}} \delta^{\mathrm{SPX}}(t, l)}, \\
\Delta^{\mathrm{VIX}}(T_j^v, k) &= \frac{\delta^{\mathrm{VIX}}(t, k)}{\sum_{l \in K_j^{\mathrm{VIX}}} \delta^{\mathrm{VIX}}(t, l)}
\end{aligned}
\tag{5.2.5}
$$

where $\delta^{\mathrm{SPX}}(t, k)$ (resp. $\delta^{\mathrm{VIX}}(t, k)$) denotes the inverse of the bid-ask spread of the OTM implied volatility $I^{\mathrm{SPX}}(t, k)$ (resp. $I^{\mathrm{VIX}}(t, k)$). Notice that we calibrate the VIX OTM call and put prices instead of the VIX implied volatilities. Indeed, model VIX futures are needed to compute model VIX implied volatilities, and it makes no sense to minimize the relative error of the VIX implied volatilities when the model and market VIX futures do not closely match.

5.2.1 Loss approximation

We approximate the loss L by discretizing in time SDE (M) via the Euler-Maruyama method. Let $N \geq 1$ be the number of Monte Carlo paths and $\Delta t > 0$ be the time step. Denote $t_n = n\Delta t$, $n \in \mathbb{N}$. We choose $\Delta t = \frac{1}{q\times 365}$ where q is an integer so that the maturities $T_j^s, T_j^v \in \Delta t\mathbb{N}$. Let $\big(\Delta B_{t_n}^{1,i}\big)_{1\leq i\leq N, 0\leq t_n\leq T}$ and $\big(\Delta B_{t_n}^{2,i}\big)_{1\leq i\leq N, 0\leq t_n\leq T}$ be independent samples of the distribution $\mathcal{N}(0,\Delta t)$.

Model (M) is approximated by the following scheme for $t_n < T$ and $1 \leq i \leq N$

$$\begin{cases} X_{t_{n+1}}^i = X_{t_n}^i - \frac{1}{2}\sigma_X(t_n, X_{t_n}^i, Y_{t_n}^i)^2\Delta t + \sigma_X(t_n, X_{t_n}^i, Y_{t_n}^i)\Delta B_{t_n}^{1,i}, \\ Y_{t_{n+1}}^i = Y_{t_n}^i + \mu_Y(t_n, X_{t_n}^i, Y_{t_n}^i)\Delta t \\ \qquad + \sigma_Y(t_n, X_{t_n}^i, Y_{t_n}^i)\big[\rho(t_n, X_{t_n}^i, Y_{t_n}^i)\Delta B_{t_n}^{1,i} \\ \qquad\qquad + \sqrt{1-\rho(t_n, X_{t_n}^i, Y_{t_n}^i)^2}\Delta B_{t_n}^{2,i}\big], \\ (X_0^i, Y_0^i) = (0,0). \end{cases} \tag{5.2.6}$$

Setting the stock price $S_{t_n}^i := f_{0,t_n} e^{X_{t_n}^i}$, the call and put prices are approximated by the Monte Carlo estimators[3]

$$\begin{aligned} \widehat{C}_m^{\text{SPX}}(t_n, k) &= \frac{1}{N}\sum_{i=1}^{N}(S_{t_n}^i - k)_+, \\ \widehat{P}_m^{\text{SPX}}(t_n, k) &= \frac{1}{N}\sum_{i=1}^{N}(k - S_{t_n}^i)_+. \end{aligned} \tag{5.2.7}$$

We denote by $\widehat{I}_m^{\text{SPX}}(t_n, k)$ the associated model OTM implied volatility.

In this discrete-time approximation, the realized variance is given by

$$R_{t_n}^i = \frac{\Delta t}{\tau}\sum_{t_n \leq t_m < t_n+\tau} \sigma_X(t_m, X_{t_m}^i, Y_{t_m}^i)^2, \tag{5.2.8}$$

[3]Other Monte Carlo schemes, reducing the variance of the estimator, could of course be used here. For the sake of simplicity, we use this basic MC estimator.

and the VIX^2 is defined by

$$\mathrm{VIX}^2_{t_n,i} = \mathbb{E}\left[R^i_{t_n}|X^i_{t_n}, Y^i_{t_n}\right], 1 \leq i \leq N. \tag{5.2.9}$$

Let $\widehat{\mathrm{VIX}^2}_{t_n,i}$ be an estimator of $\mathrm{VIX}^2_{t_n,i}$ and define

$$\widehat{\mathrm{VIX}}_{t_n,i} := \sqrt{\widehat{\mathrm{VIX}^2}_{t_n,i}}. \tag{5.2.10}$$

The VIX call and put prices are estimated by

$$\begin{aligned}\widehat{C}^{\mathrm{VIX}}_m(t_n, k) &= \frac{1}{N}\sum_{i=1}^{N}\left(\widehat{\mathrm{VIX}}_{t_n,i} - k\right)_+, \\ \widehat{P}^{\mathrm{VIX}}_m(t_n, k) &= \frac{1}{N}\sum_{i=1}^{N}\left(k - \widehat{\mathrm{VIX}}_{t_n,i}\right)_+\end{aligned} \tag{5.2.11}$$

and the VIX future price is estimated by

$$\widehat{f\mathrm{VIX}}_m(t_n) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\mathrm{VIX}}_{t_n,i}. \tag{5.2.12}$$

The loss function L in (5.2.4) is thus approximated by $\widehat{L}(\sigma_X, \mu_Y, \sigma_Y, \rho)$, which is defined as in (5.2.4) with $f\mathrm{VIX}_m$, I^{SPX}_m, C^{VIX}_m and P^{VIX}_m replaced by $\widehat{f\mathrm{VIX}}_m$, $\widehat{I}^{\mathrm{SPX}}_m$, $\widehat{C}^{\mathrm{VIX}}_m$, $\widehat{P}^{\mathrm{VIX}}_m$, respectively and the discretized minimization problem reads

$$\sigma_X, \mu_Y, \sigma_Y, \rho \in \operatorname{argmin} \widehat{L}(\sigma_X, \mu_Y, \sigma_Y, \rho). \tag{P}$$

5.2.2 Neural parameterization and minimization of $\widehat{L}$

In order to tackle the minimization problem (P), we apply a gradient descent algorithm and parameterize $(\sigma_X, \mu_Y, \sigma_Y, \rho)$ by neural networks. More precisely, let θ be a collection of weights. Let $\Phi_\theta : \mathbb{R}^3 \to \mathbb{R}^4$

$$\Phi_\theta : \begin{bmatrix} t \\ x \\ y \end{bmatrix} \in \mathbb{R}^3 \longmapsto \begin{bmatrix} \Phi_\theta^1(t,x,y) \\ \Phi_\theta^2(t,x,y) \\ \Phi_\theta^3(t,x,y) \\ \Phi_\theta^4(t,x,y) \end{bmatrix} \in \mathbb{R}^4. \tag{5.2.13}$$

be a neural network with r hidden layers of size l and weights θ. The volatility of X and Y being positive, and the correlation ρ lying in $[-1, 1]$, we choose

$$\begin{cases} \sigma_X = 1 + \tanh(\Phi_\theta^1), \\ \sigma_Y = 1 + \tanh(\Phi_\theta^2), \\ \mu_Y = \Phi_\theta^3, \\ \quad \rho = \tanh(\Phi_\theta^4). \end{cases} \tag{5.2.14}$$

Note that as a consequence we enforce that σ_X and σ_Y do not exceed 2.[4]

In order to apply a gradient descent algorithm, we need to compute the gradients $\partial_\theta \widehat{L}$. To this end, it is enough to compute the gradients $\partial_c \widehat{I}_m^{\mathrm{SPX}}(t,k)$, $\partial_\theta X_t^i$, and $\partial_\theta \widehat{\mathrm{VIX}}_{t,i}$. The gradient $\partial_c I_m^{\mathrm{SPX}}(t,k)$ is computed by using the inverse function rule. We use backpropagation through iterations (5.2.6) to compute $\partial_\theta X_t^i$.[5] The computation of the gradients $\partial_\theta \widehat{\mathrm{VIX}}_{t,i}$ is examined in the next section.

5.2.3 Differentiable VIX^2 estimator.

In order to use a gradient descent algorithm for the minimization of $\widehat{L}$, we need to estimate the VIX in an efficient and differentiable way; in particular, we need to be able to compute

[4] We could have considered $\sigma_Y = \alpha(1 + \tanh(\Phi_\theta^2))$, where α is a trainable weight. This would allow arbitrary large values for the volatility of Y. However, we notice numerically that taking $\alpha = 1$ is enough to jointly calibrate the SPX and VIX smiles. Using $\sigma_Y > 2$ is not desirable as it may produce MC estimators with a very large variance.

[5] Backpropagation requires $\mathcal{O}(l)$ memory space and takes $\mathcal{O}(l)$ time, where $l = \frac{T}{\Delta t}$ is the number of time steps in the Euler-Maruyama scheme (5.2.6). A more memory-efficient approach in $\mathcal{O}(1)$ memory space and $\mathcal{O}(l \log l)$ in time can be applied by following an adjoint method [71, 84]. However, this method requires smaller time steps.

$\partial_\theta \widehat{\mathrm{VIX}}_{t,i}$. Various methods are available to estimate

$$\mathrm{VIX}^2_{t_n,i} := \mathbb{E}\left[R^i_{t_n} \mid X^i_{t_n}, Y^i_{t_n}\right]. \tag{5.2.15}$$

- **Kernel regression.** For $W : \mathbb{R} \to \mathbb{R}_+$ a kernel and $h_X, h_Y > 0$ two bandwidths, a VIX^2 estimator is given by

$$\widehat{\mathrm{VIX}^2}_{t,i} = \frac{\sum_{j=1}^N R^j_t W_{ij}}{\sum_{j=1}^N W_{ij}}, \qquad W_{ij} := W((X^i_t - X^j_t)^2/h_X + (Y^i_t - Y^j_t)^2/h_Y)). \tag{5.2.16}$$

Optimal bandwidths are typically chosen by cross-validation to ensure the best tradeoff between bias and variance. Kernel methods implemented using the classical acceleration techniques [54] are extremely efficient in one dimension, but the curse of dimensionality decreases their efficiency in higher dimensions and may lead to high-variance estimators.

- **Linear least squares.** Let $d \in \mathbb{N}$ and $\mathbb{R}_d[X,Y] = \{[X^k Y^l]_{0 \le k+l \le d} \cdot \alpha : \alpha \in \mathbb{R}^m\}$ be the space of bivariate polynomials of degree at most d, m being the dimension of this real vector space and $[X^k Y^l]_{0 \le k+l \le d} \cdot \alpha$ denoting the polynomial with coefficients α, namely $\sum_{k+l \le d} \alpha_{k,l} X^k Y^l$. The VIX^2 estimator minimizing the quadratic error is given by

$$\widehat{\mathrm{VIX}^2}_{t,i} = [(X^i_t)^k (Y^i_t)^l]_{0 \le k+l \le d} \cdot \alpha^*, \tag{5.2.17}$$

where

$$\alpha^* \in \mathrm{argmin}_{\alpha \in \mathbb{R}^m} \frac{1}{N} \sum_{i=1}^N \left(R^i_t - [(X^i_t)^k (Y^i_t)^l] \cdot \alpha^*\right)^2. \tag{5.2.18}$$

This method is memory efficient and the estimator has a low variance, by appropriately solving the normal equation $A^T A \alpha = A^T R$ where $R = (R^i_t)_{1 \le i \le N}$ and

$$A = [(X^i_t)^k (Y^i_t)^l]_{1 \le i \le N, 0 \le k+l \le d} \in \mathbb{R}^{N \times m}. \tag{5.2.19}$$

Indeed, let $A = QS$ be the QR decomposition of A, Q being orthogonal and S upper triangular. A solution α^* is given by solving the triangular equation

$$S\alpha^* = Q^T R. \tag{5.2.20}$$

Computing the QR decomposition and solving triangular equations are numerically stable operations and can be made differentiable. Notice that solving the normal equation by inverting the matrix $A^T A + \varepsilon I_N$ requires choosing the regularization parameter $\varepsilon > 0$ (again by cross-validation for example). The choice of a basis of monomials is arbitrary; other orthogonal polynomials or transformations of X and Y can be used.

- **Nested Monte Carlo.** Let $M \in \mathbb{N}^*$. The nested Monte Carlo estimator is

$$\widehat{\mathrm{VIX}^2}_{t,i} = \frac{1}{M} \sum_{j=1}^{M} R_t^j(X_t^i, Y_t^i) \tag{5.2.21}$$

where $(R_t^j(x,y))_{1 \leq j \leq M}$ are M independent samples of the realized variance given that $(X_t^i, Y_t^i) = (x, y)$. This estimator is numerically stable and for large enough M has very low variance. However, it requires a lot of computational resources. The nested Monte Carlo methods can be made differentiable by backpropagating the gradients through the scheme (5.2.6).

- **Partial differential equation.** Notice that

$$\mathrm{VIX}_t^2(x, y) = -\frac{2}{\tau} \left(\mathbb{E}\left[X_{t+\tau} | X_t = x, Y_t = y \right] - x \right). \tag{5.2.22}$$

The conditional expectation $\mathbb{E}[X_{t+\tau}|X_t, Y_t]$ can be computed by solving the bidimensional backward parabolic equation

$$\begin{cases} \partial_t u + \frac{1}{2}\sigma_X^2 \partial_{xx}^2 u + \frac{1}{2}\sigma_Y^2 \partial_{yy}^2 u + 2\rho\sigma_X\sigma_Y \partial_{xy}^2 u - \frac{1}{2}\sigma_X^2 \partial_x u + \mu_Y \partial_y u = 0, \\ u(t+\tau, x, y) = x, \end{cases} \tag{5.2.23}$$

This equation can be solved numerically by alternating direction implicit (ADI) method for example (see [35]). This method is the most accurate and can be made differentiable by computing the gradients through the solver. However, the algorithm is computationally intensive and requires the crucial selection of the discretizing grid and boundary conditions.

We find that a good compromise between accuracy and memory usage is to choose the least squares method. The advantages of this method over nested Monte Carlo have already been highlighted in Guo and Loeper [41]. Their work gives confidence bounds for VIX futures and VIX options by computing simultaneously estimators of $\text{VIX}_t^2 = \mathbb{E}[R_t|X_t, Y_t]$ and of the stochastic integral $R - \mathbb{E}[R_t|X_t, Y_t]$, which represents the model hedging strategy. While we were able to reproduce their numerical results for the Heston model, we could not obtain tight bounds for our neural SDE model. Therefore, once our model has been calibrated, in order to check the accuracy of the least squares method, we compare the VIX and VIX smile obtained with the least squares method with those estimated with the (slow) nested Monte Carlo algorithm, which is known to be very accurate as soon as N and M are large enough.

Remark 5.2.1. *Contrary to nested Monte Carlo, least squares can lead to negative VIX^2 estimates. For the calibrated model, this happens in less than 0.01% of the simulated paths. To deal with this issue, one can clip $\widehat{VIX}^2_{t,i}$ by a small positive value h and define the estimator*

$$\widehat{VIX^2}_{t,i} = \max\left(h, \alpha^* \cdot [(X_t^i)^k (Y_t^i)^l]_{0 \le k+l \le d}\right). \tag{5.2.24}$$

This solution implies the arbitrary choice of h. We rather use filtering. An expectation

$\mathbb{E}\left[G(VIX_t)\right]$ is estimated by

$$\widehat{\mathbb{E}}\left[G(VIX_t)\right] = \frac{1}{|I|}\sum_{i\in I} G\left(\sqrt{\widehat{VIX^2}_{t,i}}\right), \tag{5.2.25}$$

where $I := \{1 \leq i \leq N : \widehat{VIX^2}_{t,i} > 0\}$. For our joint calibration problem, both methods induce a very small bias for the estimator (see Figure 5.5.2).

5.2.4 Algorithms

In this section, we provide the main algorithms used to calibrate Model (M): the training loop and the differentiable VIX computation, respectively described in Algorithms 2 and 1. For brevity, we assume the existence of the following procedures.

1. $(X_{t_n}, \partial_\theta X_{t_n}, Y_{t_n}, \partial_\theta Y_{t_n}, R_{t_n}, \partial_\theta R_{t_n})_{0\leq t_n\leq T} \leftarrow$ `Euler`$(\Delta t, T, X_0, Y_0, \theta, \Delta B_1, \Delta B_2)$.

   The inputs are a step size Δt, a final time T, N initial data $(X_0, Y_0) = (X_0^i, Y_0^i)_{1\leq i\leq N}$ at time 0, the weights θ of the neural network Φ_θ and Brownian increments ΔB_1 and ΔB_2. The procedure returns the value of X, Y at times $t_n = n\Delta t$ and the realized variance R (5.2.8), along with the gradients of these quantities with respect to θ, after $\frac{T}{\Delta t}$ iterations of Euler-Maruyama scheme (5.2.6) with the previous input parameters. The gradients are computed efficiently by backpropagation through the solver.

2. $\theta \leftarrow$ `GradientDescent`$(\theta, \partial_\theta L, lr)$.

   This procedure updates the weights θ along the gradients $\partial_\theta L$, with learning rate lr and according to a gradient descent stepper (stochastic gradient descent, Adam, Adagrad, etc.).

3. $Q, \partial_A Q, S, \partial_A S \leftarrow$ `DecompositionQR`(A).

   It computes the QR decomposition of a matrix A and the gradients $\partial_A Q$ and $\partial_A S$.

4. $x, \partial_A x, \partial_b x \leftarrow$ `SolveTriangular`(A, b).

It solves the triangular equation $Ax = b$ with unknown x and the gradients $\partial_A x$ and $\partial_b x$ of the solution x with respect to A and b.

5. $L, \partial_\theta L \leftarrow \texttt{ComputeL}(T, (X_{t_n}, \partial_\theta X_{t_n})_{0 \le t_n \le T}, (\widehat{\mathrm{VIX}}_{T_j^v}, \partial_\theta \widehat{\mathrm{VIX}}_{T_j^v})_{1 \le j \le N_T^v}, w)$.

   Finally, this function computes the loss and its gradient with respect to θ. Its implementation is only a matter of differentiating directly $\widehat{L}$.

Algorithm 1 describes the differentiable VIX computation by polynomial regression. The input d is the degree of the polynomials in the linear regression. The output are the estimated VIX and estimated gradients ∂_θVIX.

Algorithm 1 Differentiable VIX least squares estimator

---

function $\texttt{computeVIX}(X_t, \partial_\theta X_t, Y_t, \partial_\theta Y_t, R_t, \partial_\theta R_t, d)$

$A \leftarrow [X_t^{ik} Y_t^{il}]_{1 \le i \le N, 0 \le k+l \le d}$

$Q, \partial_A Q, S, \partial_A S \leftarrow \texttt{DecompositionQR}(A)$

$\alpha^*, \partial_S \alpha^*, \partial_{Q^T R} \alpha^* \leftarrow \texttt{SolveTriangular}(S, Q^T R_t)$

$\partial_\theta \alpha^* \leftarrow \partial_\theta A \left[\partial_A S \partial_S \alpha^* + \partial_A Q^T R_t \partial_{Q^T R} \alpha^*\right]$

$\mathrm{VIX}^2 \leftarrow \alpha^* \cdot A$

$\partial_\theta \mathrm{VIX}^2 \leftarrow \alpha^* \partial_\theta A + \partial_\theta \alpha^* A$

$\mathrm{VIX}, \partial_\theta \mathrm{VIX} \leftarrow \sqrt{\mathrm{VIX}^2}, \frac{\partial_\theta \mathrm{VIX}^2}{2\sqrt{\mathrm{VIX}^2}}$

return $\mathrm{VIX}, \partial_\theta \mathrm{VIX}$

end function

Algorithm 2 describes the training of the model. Its inputs are the weight of Φ_θ, the time step Δt, the weight $w = (w_{f\mathrm{VIX}}, w_{\mathrm{SPX}}, w_{\mathrm{VIX}})$ for the loss $\widehat{L}$, the degree d of the polynomials in linear regression (5.2.17), the learning rate lr of the gradient descent and Brownian increments $\Delta B^1 = (\Delta B_{t_n}^{1,i})_{1 \le i \le N, 0 \le t_n \le T}$ and $\Delta B^2 = (\Delta B_{t_n}^{2,i})_{1 \le i \le N, 0 \le t_n \le T}$. The output is the updated weight θ.

Algorithm 2 Training step

function TRAIN($\theta, N, \Delta t, w, d, lr, \Delta B^1, \Delta B^2$)

$(X_{t_n}, \partial_\theta X_{t_n}, Y_{t_n}, \partial_\theta Y_{t_n}, R_{t_n}, \partial_\theta R_{t_n})_{0\le t_n \le T} \leftarrow$ EULER($\Delta t, T, 0, 0, \theta, \Delta B_1, \Delta B_2$).

for $j \leftarrow 1$ to N_T^v **do**

$\widehat{\text{VIX}}_{T_j^v}, \partial_\theta \widehat{\text{VIX}}_{T_j^v} \leftarrow$ COMPUTEVIX($X_{T_j^v}, \partial_\theta X_{T_j^v}, Y_{T_j^v}, \partial_\theta Y_{T_j^v}, R_{T_j^v}, \partial_\theta R_{T_j^v}, d$)

end for

$L, \partial_\theta L \leftarrow$ COMPUTEL($T, (X_{t_n}, \partial_\theta X_{t_n})_{0\le t_n\le T}, (\widehat{\text{VIX}}_{T_j^v}, \partial_\theta \widehat{\text{VIX}}_{T_j^v})_{1\le j\le N_T^v}, w$).

$\theta \leftarrow$ GRADIENTDESCENT($\theta, \partial_\theta L, lr$)

return θ

end function

5.3 Numerical implementation and data

In this section we provide details of the implementations of Algorithms 1 and 2. Market data was obtained from the IvyDB Optionmetrics database, available through Wharton Research Data Service.[6]

We used `Pytorch` and Nvidia Tesla V100 GPUs for auto-differentiation and backpropagation. The `Euler` procedure is implemented using the `torchsde` library [71, 84]. The number of Monte Carlo paths is $N = 150,000$ and the time step is $\Delta t = 0.5/365$ (half a day). Φ_θ is a feedforward neural network with $r = 1$ hidden layer of width $l = 16$. We recall that the activation functions for the hidden layers are hyperbolic tangents and that the output layer is taken without activation. We use the `Adam` algorithm to perform the gradient descent. The learning rate lr is taken equal to 0.001. Implied volatilities are computed by solving a root finding problem using Brent's method. The choice of the weights $w = (w_{f\text{VIX}}, w_{\text{SPX}}, w_{\text{VIX}})$ is crucial. The VIX future should be very precisely calibrated since it is the underlying of VIX options, and also for the comparison of model and market VIX implied volatilities to be meaningful. Therefore, $w_{f\text{VIX}}$ should be rela-

[6] `https://wrds-www.wharton.upenn.edu/login/?next=/pages/get-data/optionmetrics/ivy-db-us/options/option-prices/`

tively large. The weights are taken as $(w_{f\text{VIX}}, w_{\text{SPX}}, w_{\text{VIX}}) = (30, 2, 3)$. The degree of the polynomial for the VIX^2 regression was taken to be $d = 8$; both too low a degree (large bias, underfitting) and too high a degree (large variance, overfitting) may lead to a poor VIX^2 estimation.

Finally, the nested Monte Carlo VIX^2 estimator used $M = 15,000$ nested paths and the Monte Carlo estimators of VIX payoffs are computed with $N' = 20,000$ trajectories. The code is available online. [7]

5.4 Numerical results

We calibrate to market data as of October 1, 2021. We consider respectively $N_T^s = 8$ and $N_T^v = 8$ weekly and monthly SPX and VIX maturities listed below in Figures 5.5.1 and 5.5.2, spanning 9 months of SPX options and 6 months of VIX options.[8] On each figure, the market bid-ask is plotted. The SPX smiles and the VIX futures and VIX smiles are well calibrated, often within bid-ask spreads. In Figure 5.5.2, the calibrated VIX smiles are computed using the nested Monte Carlo estimator (5.2.21). The figure also allows us to verify the accuracy the least squares estimator (5.2.17).

The surfaces of the calibrated $\sigma_X(t,\cdot,\cdot)$, $\sigma_Y(t,\cdot,\cdot)$, $\mu_Y(t,\cdot,\cdot)$ and $\rho(t,\cdot,\cdot)$ are plotted in Figures 5.5.3, 5.5.4 and 5.5.5 for t = October 13, 2021, November 13, 2021 and January 13, 2022, respectively. The red scatter plot on the surface shows the values taken at MC samples $(X_t^i, Y_t^i)_{1\le i\le N}$ defined by (5.2.6). The SPX volatility σ_X is mostly an increasing function of Y, so Y plays a role similar to the instantaneous volatility σ_X, but it also depends on X— it mostly decreases with X. The volatility of Y, σ_Y is saturated at 2. We deliberately capped σ_Y at 2 to control the variance of MC estimators and see if market data could be calibrated without pushing the "vol-of-vol" σ_Y too high. The correlation ρ is thus -1 almost everywhere, so as to match the large negative SPX market skews; the

[7] `https://archive.softwareheritage.org/browse/origin/directory/?origin_url=https://github.com/intermet/neural-spx-vix-calibration-sh`

[8] Our data set does not contain option data with longer maturities.

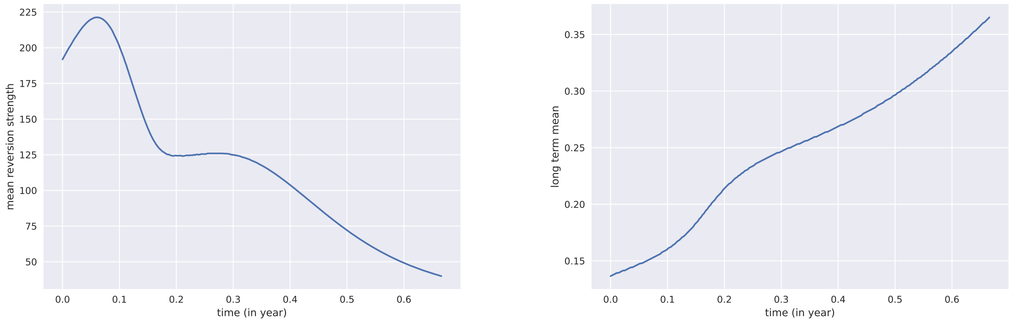


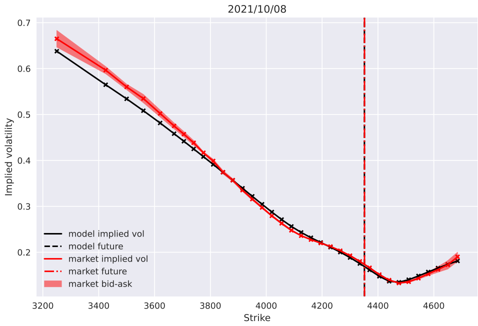
Figure 5.4.1: Left: time-dependent mean reversion, defined as the slope of $y \mapsto \mu_Y(t, 0, y)$ at the value y_t such that $\mu_Y(t, 0, y_t) = 0$. Right: long-term mean of volatility $t \mapsto \sigma_X(t, 0, y_t)$

fact that σ_X decreases with X contributes as well. Therefore only one Brownian motion drives (most of) the model dynamics and the model is (almost) purely path-dependent. This confirms the findings in [49], and that PDV models are natural candidates for solving the joint calibration problem [39, 55]. The drift μ_Y is essentially a fast decreasing linear function of Y: our neural SDE procedure learns (from scratch) that the volatility factor Y is fast mean-reverting, with a time-dependent characteristic scale of mean reversion varying from around 2 days (for small t) to around 7 days (for t around 6 months); see Figure 5.4.1. Since our neural net takes directly (t, x, y) as input and only uses smooth activation functions, the surfaces are smooth and vary smoothly with time.

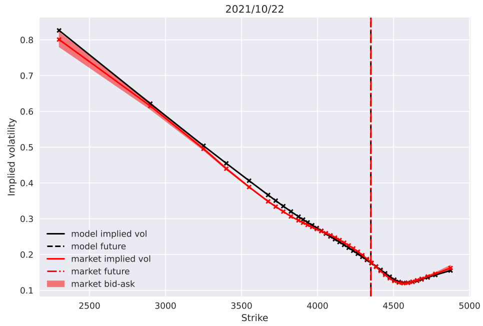
When we initialize the neural network with random weights, the calibration takes a long time (36 hours). However, initializing the neural network with the parameters calibrated on the day before greatly speeds up the calibration process, which then takes only around 3 hours to calibrate on the next business day; the calibration can thus run overnight. The computation time is also reduced by considering less maturities. To calibrate to one monthly VIX maturity (October 20, 2021) and two monthly SPX maturities (October 15 and November 19, 2021), only 4 hours are needed with random initialization (compared to 11 hours reported in [43]).

5.5 Conclusion

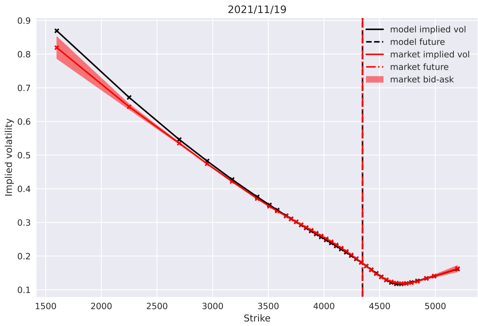
In this chapter, we have shown that a one-factor SLV model can jointly calibrate to SPX and VIX smiles and VIX futures for many maturities, provided enough flexibility is allowed on the SDE coefficients, which we model as neural networks. The calibrated model is actually a one-factor PDV model with a fast mean-reverting path-dependent factor Y which depends only on past SPX returns. Our work thus illustrates the expressivity of neural SDEs by solving the joint calibration problem with good accuracy for multiple maturities, and provides yet extra reasons to use PDV models for pricing, hedging, and risk-managing derivatives.



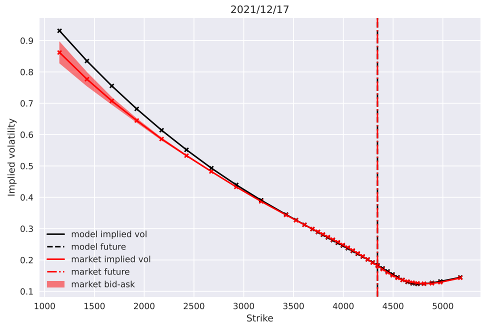
2021/10/08



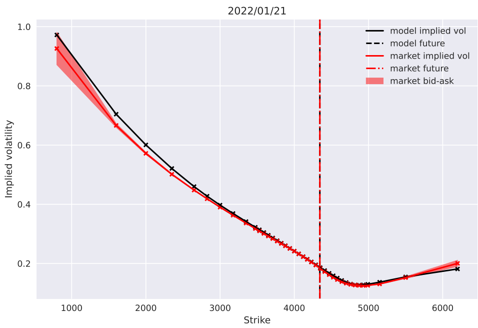
2021/10/22



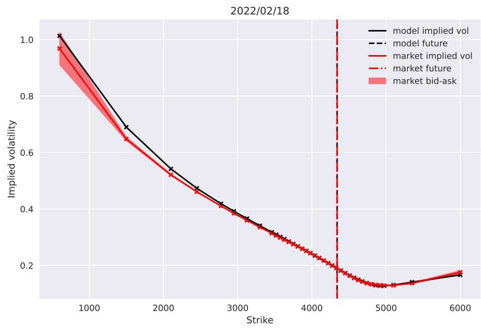
2021/11/19



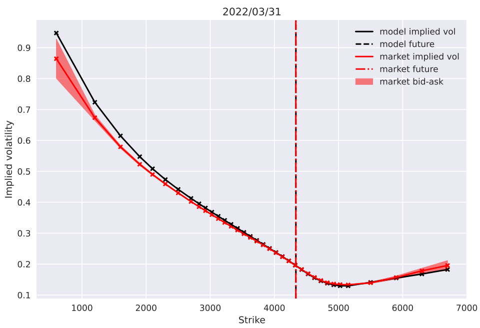
2021/12/17



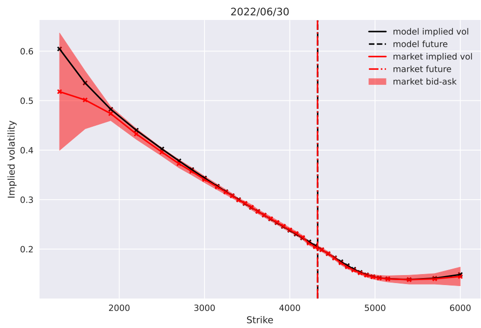
2022/01/21



2022/02/18



2022/03/31



2022/06/30

Figure 5.5.1: Calibration of the SPX smiles as of October 1, 2021.

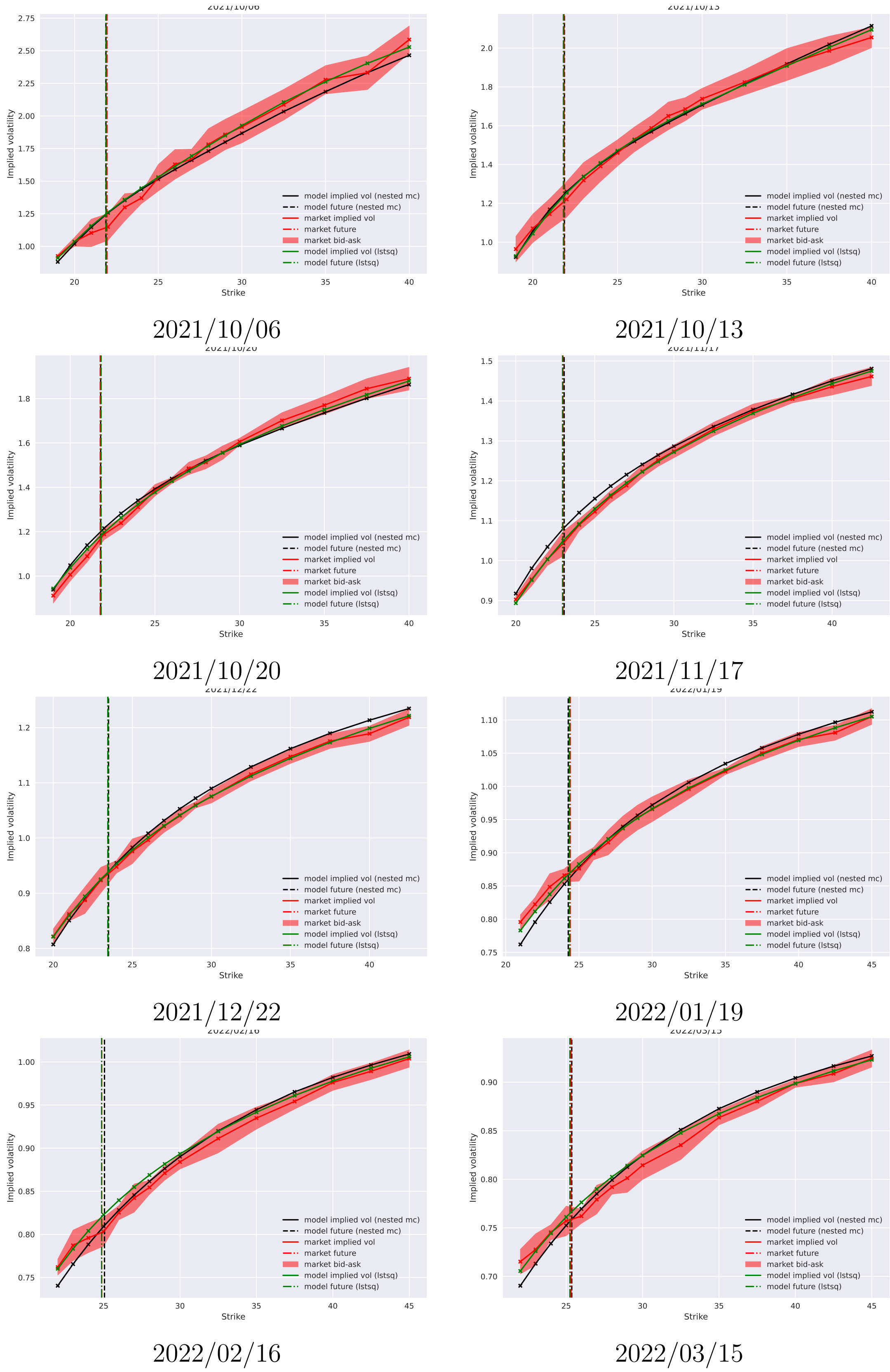
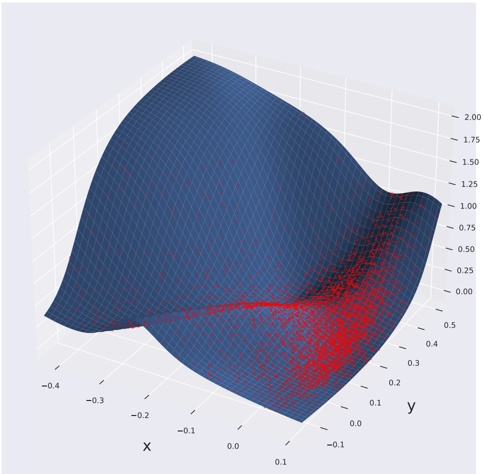
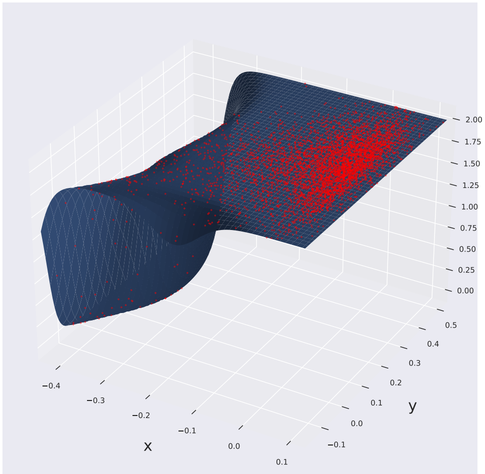


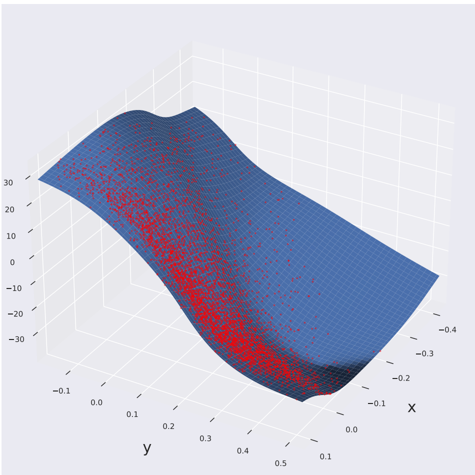
Figure 5.5.2: Calibration of the VIX futures and VIX smiles as of October 1, 2021.



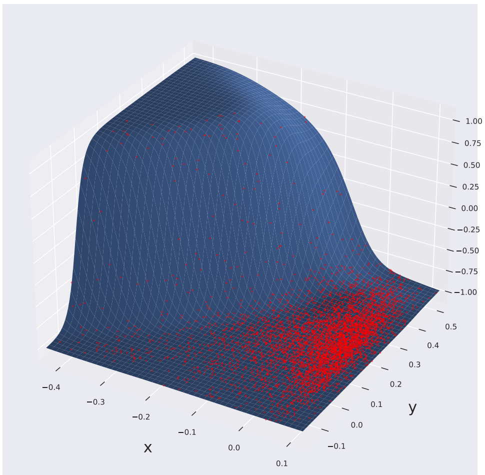
(a) $\sigma_X\,(t, x, y)$



(b) $\sigma_Y\,(t, x, y)$

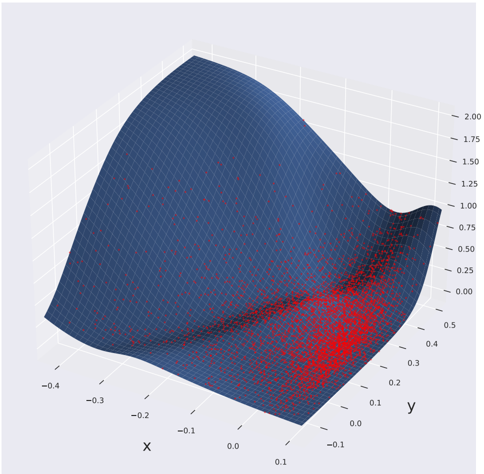


(c) $\mu_Y\,(t, x, y)$

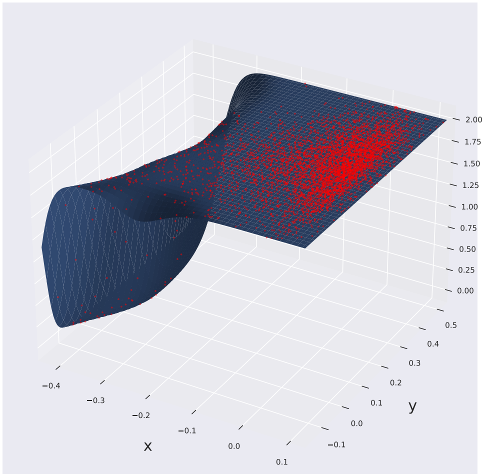


(d) $\rho\,(t, x, y)$

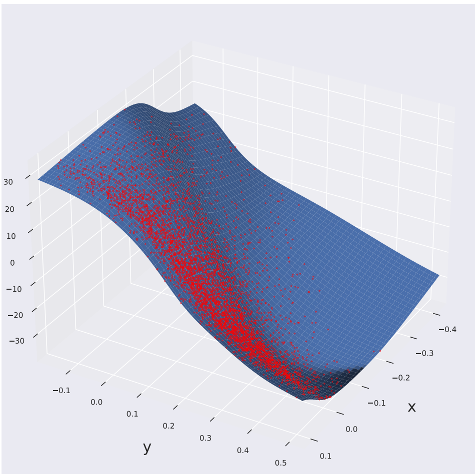
Figure 5.5.3: Plots of the optimal σ_X, σ_Y, μ_Y, ρ at time t = October 13, 2021, as of October 1, 2021.



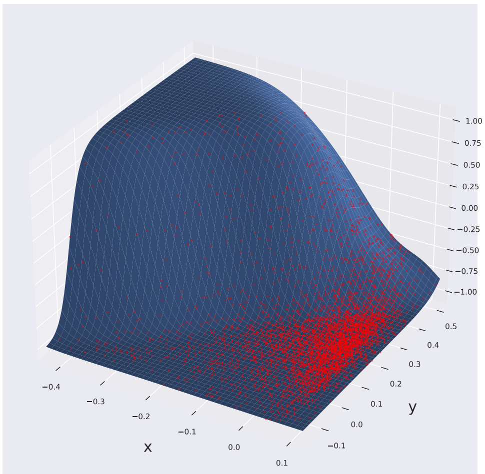
(a) $\sigma_X\,(t,x,y)$



(b) $\sigma_Y\,(t,x,y)$

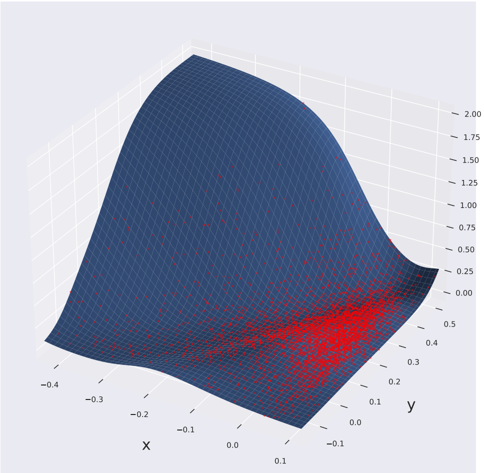


(c) $\mu_Y\,(t,x,y)$

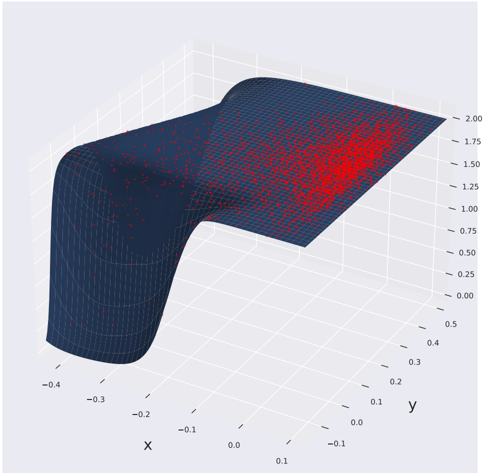


(d) $\rho\,(t,x,y)$

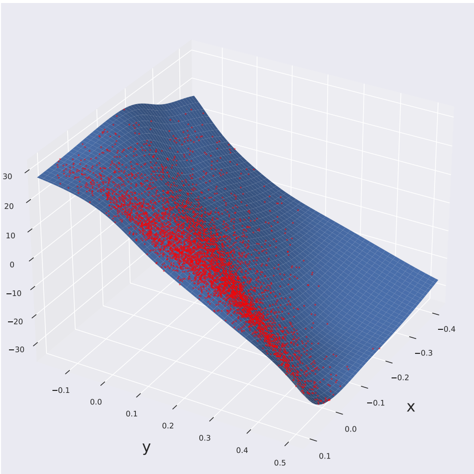
Figure 5.5.4: Plots of the optimal σ_X, σ_Y, μ_Y, ρ at time $t =$ November 13, 2021, as of October 1, 2021.



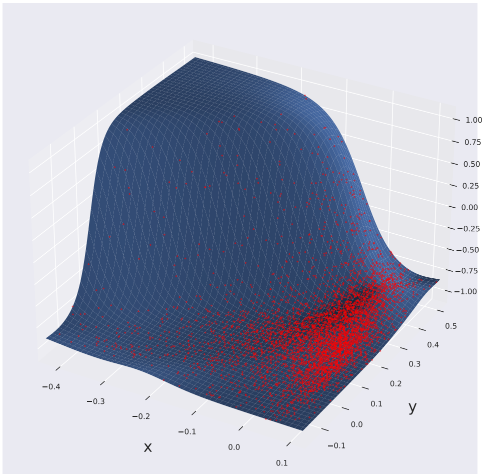
(a) $\sigma_X\,(t, x, y)$



(b) $\sigma_Y\,(t, x, y)$



(c) $\mu_Y\,(t, x, y)$



(d) $\rho\,(t, x, y)$

Figure 5.5.5: Plots of the optimal σ_X, σ_Y, μ_Y, ρ at time t = January 13, 2022, as of October 1, 2021.

Bibliography

[1] Frédéric Abergel, Rémy Tachet Des Combes, Riadh Zaatour, et al. Nonparametric model calibration for derivatives. Journal of Mathematical Finance, 7(03):571, 2017.

[2] Frédéric Abergel and Rémi Tachet. A nonlinear partial integro-differential equation from mathematical finance. Discrete and Continuous Dynamical Systems-Series A, 27(3):907–917, 2010.

[3] Eduardo Abi Jaber, Camille Illand, and Shaun Xiaoyuan Li. Joint SPX–VIX calibration with gaussian polynomial volatility models: deep pricing with quantization hints. Available at SSRN 4292544, 2022.

[4] Romain Allez and Laure Dumaz. Random matrices in non-confining potentials. Journal of Statistical Physics, 160(3):681–714, 2015.

[5] Graeme Baker and Mykhaylo Shkolnikov. A singular two-phase stefan problem and particles interacting through their hitting times. arXiv preprint arXiv:2203.06003, 2022.

[6] D. Bakry and M. Émery. Diffusions hypercontractives. In Jacques Azéma and Marc Yor, editors, Séminaire de Probabilités XIX 1983/84, pages 177–206, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg.

[7] Dominique Bakry and Michel Émery. Diffusions hypercontractives. In Séminaire de Probabilités XIX 1983/84: Proceedings, pages 177–206. Springer, 2006.

[8] Jan Baldeaux and Alexander Badran. Consistent modelling of VIX and equity derivatives using a 3/2 plus jumps model. Applied Mathematical Finance, 21(4):299–312, 2014.

[9] Chris Bardgett, Elise Gourier, and Markus Leippold. Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. Journal of Financial Economics, 131(3):593–618, 2019.

[10] Philippe Biane. Free brownian motion, free stochastic calculus and random matrice, in free probability theory. Fields Inst. Commun., 12:1–19, 1997.

[11] Philippe Biane. Free probability for probabilists. arXiv preprint math/9809193, 1998.

[12] Philippe Biane and Roland Speicher. Free diffusions, free entropy and free Fisher information. In Annales de l'Institut Henri Poincare (B) Probability and Statistics, volume 37, pages 581–606. Elsevier, 2001.

[13] François Bolley, Ivan Gentil, and Arnaud Guillin. Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations. Journal of Functional Analysis, 263(8):2430–2457, 2012.

[14] François Bolley, Ivan Gentil, and Arnaud Guillin. Uniform convergence to equilibrium for granular media. Archive for Rational Mechanics and Analysis, 208(2):429–445, 2013.

[15] Mireille Bossy and Jean-François Jabir. On the wellposedness of some mckean models with moderated or singular diffusion coefficient. In Frontiers in Stochastic Analysis–BSDEs, SPDEs and their Applications: Edinburgh, July 2017 Selected, Revised and Extended Contributions 8, pages 43–87. Springer, 2019.

[16] Alexander Vadimovich Bulinski. A new variant of the functional law of the iterated logarithm. Teoriya Veroyatnostei i ee Primeneniya, 25(3):502–512, 1980.

[17] René Carmona, François Delarue, and Aimé Lachapelle. Control of mckean–vlasov dynamics versus mean field games. Mathematics and Financial Economics, 7:131–166, 2013.

[18] José A Carrillo, Robert J McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. Revista Matematica Iberoamericana, 19(3):971–1018, 2003.

[19] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. Advances in neural information processing systems, 31, 2018.

[20] F. Comets and Th. Eisele. Asymptotic dynamics, noncritical and critical fluctuations for a geometric long-range interacting model. Communications in Mathematical Physics, 118(4):531–567, 1988.

[21] Rama Cont and Thomas Kokholm. A consistent pricing model for index options and volatility derivatives. Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics, 23(2):248–274, 2013.

[22] Christa Cuchiero, Guido Gazzani, Janka Möller, and Sara Svaluto-Ferro. Joint calibration to spx and vix options with signature-based models. arXiv preprint arXiv:2301.13235, 2023.

[23] Christa Cuchiero, Wahid Khosrawi, and Josef Teichmann. A generative adversarial network approach to calibration of local stochastic volatility models. Risks, 8(4):101, 2020.

[24] Christa Cuchiero, Stefan Rigger, and Sara Svaluto-Ferro. Propagation of minimality in the supercooled Stefan problem. The Annals of Applied Probability, 33(2):1388–1418, 2023.

[25] François Delarue, James Inglis, Sylvain Rubenthaler, and Etienne Tanré. Global solvability of a networked integrate-and-fire model of McKean–Vlasov type. The Annals of Applied Probability, 25(4):2096–2133, 2015.

[26] François Delarue, James Inglis, Sylvain Rubenthaler, and Etienne Tanré. Particle systems with a singular mean-field self-excitation. Application to neuronal networks. Stochastic Processes and their Applications, 125(6):2451–2492, 2015.

[27] François Delarue, Sergey Nadtochiy, and Mykhaylo Shkolnikov. Global solutions to the supercooled Stefan problem with blow-ups: regularity and uniqueness. Probability and Mathematical Physics, 3(1):171–213, 2022.

[28] François Delarue, Sergey Nadtochiy, and Mykhaylo Shkolnikov. Global solutions to the supercooled stefan problem with blow-ups: regularity and uniqueness. Probability and Mathematical Physics, 2019.

[29] Amir Dembo and Li-Cheng Tsai. Criticality of a randomly-driven front. Archive for Rational Mechanics and Analysis, 233(2):643–699, 2019.

[30] Catherine Donati-Martin, Benjamin Groux, and Mylene Maida. Convergence to equilibrium in the free fokker–planck equation with a double-well potential. In Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, volume 54, pages 1805–1818. Institut Henri Poincaré, 2018.

[31] Bruno Dupire et al. Pricing with a smile. Risk, 7(1):18–20, 1994.

[32] Freeman J Dyson. A brownian-motion model for the eigenvalues of a random matrix. Journal of Mathematical Physics, 3(6):1191–1198, 1962.

[33] Begoña Fernandez and Sylvie Méléard. A Hilbertian approach for fluctuations on the McKean-Vlasov model. Stochastic Processes and their Applications, 71(1):33–53, 1997.

[34] Franco Flandoli, Enrico Priola, and Giovanni Zanco. A mean-field model with discontinuous coefficients for neurons with spatial interaction. arXiv preprint arXiv:1708.04156, 2017.

[35] S. Foulon and K.J. in 't Hout. ADI finite difference schemes for option pricing in the heston model with correlation. International Journal of Numerical Analysis & Modeling, 7(2):303–320, 2010.

[36] Jean-Pierre Fouque and Yuri Saporito. Heston stochastic vol-of-vol model for joint calibration of VIX and S&P 500 options. Quantitative Finance, 18(6):1003–1016, 2018.

[37] Nina Gantert. An inversion of Strassen's law of the iterated logarithm for small time. The Annals of Probability, 21(2):1045–1049, 1993.

[38] Jim Gatheral. Consistent modeling of spx and vix options. In Bachelier congress, volume 37, pages 39–51, 2008.

[39] Jim Gatheral, Paul Jusselin, and Mathieu Rosenbaum. The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem. Risk, May 2020.

[40] Patryk Gierjatowicz, Marc Sabate-Vidales, David Šiška, Lukasz Szpruch, and Žan Žurič. Robust pricing and hedging via neural SDEs. arXiv preprint arXiv:2007.04154, 2020.

[41] Ivan Guo and Grégoire Loeper. Pricing bounds for volatility derivatives via duality and least squares Monte Carlo. Journal of Optimization Theory and Applications, 179(2):598–617, 2018.

[42] Ivan Guo, Grégoire Loeper, Jan Obłój, and Shiyi Wang. Joint modeling and calibration of spx and vix by optimal transport. SIAM Journal on Financial Mathematics, 13(1):1–31, 2022.

[43] Ivan Guo, Grégoire Loeper, Jan Obloj, and Shiyi Wang. Joint modeling and calibration of SPX and VIX by optimal transport. SIAM Journal on Financial Mathematics, 13(1):1–31, 2022.

[44] Yucheng Guo, Sergey Nadtochiy, and Mykhaylo Shkolnikov. Stefan problem with surface tension: uniqueness of physical solutions under radial symmetry. arXiv preprint arXiv:2306.02969, 2023.

[45] J Guyon. The vix future in bergomi models: Fast approximation formulas and joint calibration with s&p 500 skew, to appear in siam j. Finan. Math, 2022.

[46] Julien Guyon. The joint s&p 500/vix smile calibration puzzle solved. Risk, April, 2020.

[47] Julien Guyon. Dispersion-constrained martingale schrödinger bridges: Joint entropic calibration of stochastic volatility models to s&p 500 and vix smiles. Available at SSRN 4165057, 2022.

[48] Julien Guyon. Dispersion-Constrained Martingale Schrödinger Bridges: Joint Entropic Calibration of Stochastic Volatility Models to S&P 500 and VIX Smiles. Available at SSRN 4165057, 2022.

[49] Julien Guyon. The VIX Future in Bergomi Models: Fast Approximation Formulas and Joint Calibration with S&P 500 Skew. SIAM Journal on Financial Mathematics, 13(4):1418–1485, 2022.

[50] Julien Guyon. The joint S&P 500/VIX smile calibration puzzle solved. Risk, April 2020.

[51] Julien Guyon and Florian Bourgey. Fast exact joint s&p 500/vix smile calibration in discrete and continuous time. Available at SSRN 4315084, 2022.

[52] Julien Guyon and Pierre Henry-Labordère. Being particular about calibration. Risk, 25(1):88, 2012.

[53] Julien Guyon and Pierre Henry-Labordere. Nonlinear option pricing. CRC Press, 2013.

[54] Julien Guyon and Pierre Henry-Labordère. The smile calibration problem solved. Risk, May 2012.

[55] Julien Guyon and Jordan Lekeufack. Volatility is (mostly) path-dependent. Quantitative Finance, forthcoming, 2022.

[56] Julien Guyon and Scander Mustapha. Neural joint S&P 500/VIX smile calibration. SSRN preprint 4309576, 2022.

[57] István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. Probability theory and related fields, 71(4):501–516, 1986.

[58] Ben Hambly, Sean Ledger, and Andreas Søjmark. A McKean–Vlasov equation with positive feedback and blow-ups. The Annals of Applied Probability, 29(4):2338–2373, 2019.

[59] Pierre Henry-Labordère. Calibration of local stochastic volatility models to market smiles: A monte-carlo approach. Risk Magazine, September, 2009.

[60] Liam Hodgkinson, Chris van der Heide, Fred Roosta, and Michael W Mahoney. Stochastic normalizing flows: training SDEs as ODEs. In Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence (UAI 2021), PMLR 161, pages 1130–1140, 2020.

[61] Yueyun Hu, Daniel Pierre-Loti-Viaud, and Zhan Shi. Laws of the iterated logarithm for iterated Wiener processes. Journal of Theoretical Probability, 8:303–319, 1995.

[62] Eduardo Abi Jaber, Camille Illand, et al. The quintic ornstein-uhlenbeck volatility model that jointly calibrates spx & vix smiles. arXiv preprint arXiv:2212.10917, 2022.

[63] Antoine Jacquier, Claude Martini, and Aitor Muguruza. On vix futures in the rough bergomi model. Quantitative Finance, 18(1):45–61, 2018.

[64] Junteng Jia and Austin R Benson. Neural jump stochastic differential equations. Advances in Neural Information Processing Systems, 32, 2019.

[65] Benjamin Jourdain and Sylvie Méléard. Propagation of chaos and fluctuations for a moderate model with smooth initial data. In Annales de l'Institut Henri Poincare (B) Probability and Statistics, volume 34, pages 727–766. Elsevier, 1998.

[66] Benjamin Jourdain and Julien Reygner. Capital distribution and portfolio performance in the mean-field atlas model. Annals of Finance, 11(2):151–198, 2015.

[67] Benjamin Jourdain and Alexandre Zhou. Existence of a calibrated regime switching local volatility model. Mathematical Finance, 30(2):501–546, 2020.

[68] Mark Kac. Foundations of kinetic theory. In Proceedings of The third Berkeley symposium on mathematical statistics and probability, volume 3, pages 171–197, 1956.

[69] S Kamenomostskaya. On the stefan problem. Mat. Sb.(NS), 53:489–514, 1961.

[70] Patrick Kidger, James Foster, Xuechen Li, and Terry J Lyons. Neural sdes as infinite-dimensional gans. In International conference on machine learning, pages 5453–5463. PMLR, 2021.

[71] Patrick Kidger, James Foster, Xuechen Li, and Terry J Lyons. Neural SDEs as infinite-dimensional GANs. In Proceedings of the 38th International Conference on Machine Learning, PMLR 139, pages 5453–5463. PMLR, 2021.

[72] Arturo Kohatsu-Higa and Shigeyoshi Ogawa. Weak rate of convergence for an euler scheme of nonlinear sde's. 1997.

[73] Thomas Kokholm and Martin Stisen. Joint pricing of vix and spx options with stochastic volatility and jump models. The Journal of Risk Finance, 16(1):27–48, 2015.

[74] Praveen Kolli and Mykhaylo Shkolnikov. Spde limit of the global fluctuations in rank-based models. The Annals of Probability, 46(2):1042–1069, 2018.

[75] Praveen Kolli and Mykhaylo Shkolnikov. SPDE limit of the global fluctuations in rank-based models. The Annals of Probability, 46(2):1042–1069, 2018.

[76] Daniel Lacker, Mykhaylo Shkolnikov, and Jiacheng Zhang. Inverting the markovian projection, with an application to local stochastic volatility models. 2020.

[77] Olga Aleksandrovna Ladyzhenskaya. Linear and quasilinear elliptic equations. translated by Scripta Technica, 1968.

[78] Carlo Lancellotti. Time-asymptotic evolution of spatially uniform Gaussian Vlasov fluctuation fields. Journal of Statistical Physics, 163(4):868–886, 2016.

[79] Sean Ledger and Andreas Søjmark. Uniqueness for contagious McKean–Vlasov systems in the weak feedback regime. Bulletin of the London Mathematical Society, 52(3):448–463, 2020.

[80] Michel Ledoux and Ionel Popescu. Mass transportation proofs of free functional inequalities, and free poincaré inequalities. Journal of Functional Analysis, 257(4):1175–1221, 2009.

[81] Michel Ledoux and Ionel Popescu. Mass transportation proofs of free functional inequalities, and free Poincaré inequalities. Journal of Functional Analysis, 257(4):1175–1221, Aug 2009.

[82] Songzi Li, Xiang-Dong Li, and Yong-Xiao Xie. On the law of large numbers for the empirical measure process of generalized dyson brownian motion. Journal of Statistical Physics, 181(4):1277–1305, 2020.

[83] Songzi Li, Xiang-Dong Li, and Yong-Xiao Xie. On the law of large numbers for the empirical measure process of generalized dyson brownian motion. Journal of Statistical Physics, 181(4):1277–1305, 2020.

[84] Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. In Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, PMLR 108, 2020.

[85] Anatoliy Malyarenko. Invariant random fields on spaces with a group action. Springer Science & Business Media, 2012.

[86] Robert J McCann. A convexity principle for interacting gases. Advances in mathematics, 128(1):153–179, 1997.

[87] Henry P McKean. Propagation of chaos for a class of non-linear parabolic equations. Stochastic Differential Equations (Lecture Series in Differential Equations, Session 7, Catholic Univ., 1967), pages 41–57, 1967.

[88] Henry P McKean Jr. A class of markov processes associated with nonlinear parabolic equations. Proceedings of the National Academy of Sciences, 56(6):1907–1911, 1966.

[89] Scander Mustapha. Trend to equilibrium for granular media equations under non-convex potential and application to log gases. arXiv preprint arXiv:2011.08547, 2020.

[90] Scander Mustapha and Mykhaylo Shkolnikov. Well-posedness of the supercooled stefan problem with oscillatory initial conditions. arXiv preprint arXiv:2302.13097, 2023.

[91] Sergey Nadtochiy and Mykhaylo Shkolnikov. Particle systems with singular interaction through hitting times: application in systemic risk modeling. The Annals of Applied Probability, 29(1):89–129, 2019.

[92] Sergey Nadtochiy and Mykhaylo Shkolnikov. Mean field systems on networks, with singular interaction through hitting times. The Annals of Probability, 48(3):1520–1556, 2020.

[93] Karl Oelschläger. A law of large numbers for moderately interacting diffusion processes. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 69(2):279–322, 1985.

[94] Hiroshi Oodaira. On Strassen's version of the law of the iterated logarithm for Gaussian processes. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 21(4):289–299, 1972.

[95] Hiroshi Oodaira. The law of the iterated logarithm for Gaussian processes. The Annals of Probability, pages 954–967, 1973.

[96] Claudio Pacati, Gabriele Pompa, and Roberto Renò. Smiling twice: the Heston++ model. Journal of Banking & Finance, 96:185–206, 2018.

[97] Andrew Papanicolaou and Ronnie Sircar. A regime-switching heston model for vix and s&p 500 implied volatilities. Quantitative Finance, 14(10):1811–1827, 2014.

[98] Leonard CG Rogers and Zhan Shi. Interacting brownian particles and the wigner law. Probability theory and related fields, 95:555–570, 1993.

[99] Mathieu Rosenbaum and Jianfei Zhang. Deep calibration of the quadratic rough heston model. Risk, October 2022.

[100] Laura Sacerdote and Maria Teresa Giraudo. Stochastic integrate and fire models: a review on mathematical methods and their applications. Stochastic

biomathematical models: with applications to neuronal modeling, pages 99–148, 2013.

[101] Edward B Saff and Vilmos Totik. Logarithmic potentials with external fields, volume 316. Springer Science & Business Media, 2013.

[102] Etienne Sandier and Sylvia Serfaty. Vortices in the magnetic Ginzburg-Landau model, volume 70. Springer Science & Business Media, 2008.

[103] Yuri F Saporito, Xu Yang, and Jorge P Zubelli. The calibration of stochastic local-volatility models: An inverse problem perspective. Computers & Mathematics with Applications, 77(12):3054–3067, 2019.

[104] Bernard Sherman. A general one-phase stefan problem. Quarterly of Applied Mathematics, 28(3):377–382, 1970.

[105] Volker Strassen. An invariance principle for the law of the iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 3(3):211–226, 1964.

[106] Alain-Sol Sznitman. Topics in propagation of chaos. Lecture notes in mathematics, pages 165–251, 1991.

[107] Julian Tugaut. Convergence to the equilibria for self-stabilizing processes in double-well landscape. The Annals of Probability, 41(3A):1427–1460, 2013.

[108] Belinda Tzen and Maxim Raginsky. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. arXiv preprint arXiv:1905.09883, 2019.

[109] Cédric Villani. Topics in optimal transportation. Number 58. American Mathematical Soc., 2003.

[110] Dan Voiculescu. The analogues of entropy and of Fisher's information measure in free probability theory, I. Communications in mathematical physics, 155(1):71–92, 1993.

[111] Hao Wu, Jonas Köhler, and Frank Noé. Stochastic normalizing flows. Advances in Neural Information Processing Systems, 33:5933–5944, 2020.