Analysis and Numerics of Stochastic Gradient Flows

Von der Fakultät für Mathematik und Informatik
der Universität Leipzig
angenommene

DISSERTATION

zur Erlangung des akademischen Grades
DOCTOR RERUM NATURALIUM
(Dr.rer.nat.)

im Fachgebiet
Mathematik
vorgelegt

von Florian Johannes Kunick, M.Sc.
geboren am 03.04.1994 in Memmingen

Die Annahme der Dissertation wurde empfohlen von:

1. Professor Dr. Felix Otto (Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig)
2. Professor Dr. José Carrillo (University of Oxford, Oxford)

Die Verleihung des akademischen Grades erfolgt mit Bestehen der Verteidigung am 20.09.2022 mit dem Gesamtprädikat magna cum laude.

Erklärungen

- Ich erkenne die Promotionsordnung der Fakultät für Mathematik und Informatik der Universität Leipzig vom 2. Mai 2013 in der Fassung der Ersten Änderungssatzung vom 16. Oktober 2017 an.
- Die eingereichte Arbeit wurde nicht in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt.
- Es haben keine früheren erfolglosen Promotionsversuche stattgefunden.

....
(Florian Johannes Kunick)

Bibliographische Daten

Analysis and Numerics of Stochastic Gradient Flows
(Analysis und Numerik Stochastischer Gradientenflüsse)
Kunick, Florian Johannes
Universität Leipzig, Dissertation, 2022
141 Seiten, 10 Abbildungen, 122 Referenzen

Abstract

In this thesis we study three stochastic partial differential equations (SPDE) that arise as stochastic gradient flows via the fluctuation-dissipation principle.
For the first equation we establish a finer regularity statement based on a generalized Taylor expansion which is inspired by the theory of rough paths.
The second equation is the thin-film equation with thermal noise which is a singular SPDE. In order to circumvent the issue of dealing with possible renormalization, we discretize the gradient flow structure of the deterministic thin-film equation. Choosing a specific discretization of the metric tensor, we resdiscover a well-known discretization of the thin-film equation introduced by Grün and Rumpf that satisfies a discrete entropy estimate. By proving a stochastic entropy estimate in this discrete setting, we obtain positivity of the scheme in the case of no-slip boundary conditions. Moreover, we analyze the associated rate functional and perform numerical experiments which suggest that the scheme converges.
The third equation is the massive φ^4_2-model on the torus which is also a singular SPDE. In the spirit of Bakry and Émery, we obtain a gradient bound on the Markov semigroup. The proof relies on an L^2-estimate for the linearization of the equation. Due to the required renormalization, we use a stopping time argument in order to ensure stochastic integrability of the random constant in the estimate. A postprocessing of this estimate yields an even sharper gradient bound. As a corollary, for large enough mass, we establish a local spectral gap inequality which by ergodicity yields a spectral gap inequality for the φ^4_2-measure.

Acknowledgments

I am very grateful to my advisors, Prof. Dr. Benjamin Gess and Prof. Dr. Felix Otto, for the opportunity to do research with them. Thinking of all the frequent meetings makes me realize how much I have learned in the previous years. I also consider myself to be very fortunate to have been part of this institute where they have created such an international and stimulating environment to do Mathematics.

My special thanks go to my coauthors Rishabh Gvalani and Pavlos Tsatsoulis. Their door was always open for me to ask questions, and the countless hours spent at the blackboard have been a pleasure.

I would like to thank the other, former and current, PhD-students, Marius Neuß, Pablo Linares, Markus Tempelmayr, Francesco Mattesini, Fenna Müller and Christian Wagner, with whom I had many fruitful mathematical discussions. Moreover, I want to thank Daniel Heydecker and Lukas Koch for reading parts of this thesis, and Tobias Ried for all the encouraging conversations. The first half of my PhD also was positively impacted by Angkana Rüland and her former group.

Finally, I want to thank Antonia for her constant support, for cheering me up when I needed it the most, and, most importantly, for just being by my side.

Contents

CHAPTER 1

Introduction

A *gradient flow* is an ordinary differential equation (ODE) that (locally) minimizes a given functional, often referred to as energy. It does so by infinitesimally flowing into the direction of steepest descent in the energy landscape. The notion of steepest descent is intimately connected to the notion of a gradient, but in order for a gradient to exist, an inner product is needed or, in other words, a geometry. Hence, more precisely, a gradient flow is an ODE that minimizes an energy as fast as a given geometry allows it to. Often, the *configuration space* is infinite-dimensional and then such a gradient flow is given by a partial differential equation (PDE). For example, there are two important gradient flows associated with minimizing the Dirichlet energy. While in the Euclidean geometry such a gradient flow would correspond to the *heat equation*, in the Wasserstein geometry, known from the theory of optimal transportation, it corresponds to the *thin-film equation* and the Dirichlet energy has the interpretation of the surface tension. These examples highlight the fact that a change in geometry can have striking differences; while the heat equation is a second-order, linear PDE, the thin-film equation is a fourth-order, quasi-linear PDE.

When introducing fluctuations to the equation, and thus turning it into a *stochastic gradient flow*, the *fluctuation–dissipation theorem* suggests a noise that is compatible with the geometry. This means that the noise term does not depend on the energy, but only on the geometry. Hence, while a gradient flow, like the diffusion equation, can have multiple (even infinitely many) gradient flow structures, they give rise to completely different stochastic gradient flows. A feature that all of them have in common is that the fluctuation-dissipation theorem suggests the same invariant measure; the *Gibbs-measure* associated to the energy. This is not completely correct, though; while the Gibbs-measure indeed does not depend on the geometry, it depends on the underlying configuration space. Formally, the Gibbs measure and the energy are in a one-to-one correspondence, but in the infinite-dimensional setting it happens that the Gibbs measure does not make any sense. In that case one needs to perform a *renormalization* procedure in order to rigorously define it. While in thermodynamics an invariant measure (corresponding to equilibrium) is postulated and the interest lies mostly in the dynamics of the system, in quantum field theory a program of *stochastic quantization* has been suggested. Quantum field theorists are concerned with certain infinite-dimensional Gibbs measures that are not a priori well-defined. Stochastic quantization refers to the procedure of introducing a Langevin equation that is supposed to have as an invariant measure exactly this Gibbs measure. In that case, the time and the dynamics are completely artificial, serving the only purpose of sampling from the Gibbs measure after long times.

In the infinite-dimensional case, stochastic gradient flows are often *singular* stochastic partial differential equations: as in the case of infinite-dimensional Gibbs measures, these equations contain nonlinear terms that are ill-defined. This is due to a rough, stochastic forcing term: a white noise. Similarly to stochastic integrals, probabilistic techniques are necessary to give sense to these products. Since the emergence of the theories of *regularity structures* and *paracontrolled distributions* a lot of progress has been made to study singular SPDEs in the subcritical regime. These include many equations arising in quantum field theory via stochastic quantization, but also random growth interface models like the KPZ equation have been considered.

The approach of Bakry and Émery to prove functional inequalities which quantify ergodicity or convergence to equilibrium has been very successful since it was introduced. The data provided by a gradient flow structure fits well in the framework of this approach. Their insight essentially was that the geometry of the configuration space, as well as the convexity of the energy landscape, play a decisive factor, if the associated stochastic dynamics converge to equilibrium exponentially fast. A weaker condition for these functional inequalities concern the commutativity of the semigroup and the gradient which come in the form of heat kernel estimates and are of independent interest.

We want to stress that, in the following, we will be mostly formal. In particular, we will not specify any regularity or integrability conditions, which means that some expressions are not well-defined. While in the finite-dimensional setting gradient flow structures are mostly rigorous, in the infinite-dimensional setting they have to be taken with a grain of salt. The same warning applies to the fluctuation-dissipation theorem; we focus on giving a heuristic overview of the underlying principles that come up in this thesis. The contributions of the subsequent chapters are to make some of the heuristic arguments presented rigorous in the introduction, at least in special cases.

1.1. The diffusion equation and Brownian motion

The diffusion equation

$$\begin{cases} \partial_t u = \Delta u \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d \\ u|_{t=0} = f \end{cases} \tag{1.1}$$

is a prime example of a second-order, parabolic partial differential equation. Its solution is explicitly given by the semigroup $(P_t)_{t\geq 0}$ defined as

$$P_t f(x) := \int_{\mathbb{R}^d} p_t(x-y) f(y)\,\mathrm{d}y$$

for $f : \mathbb{R}^d \to \mathbb{R}$ where p_t is the *heat kernel*

$$p_t(x) := \frac{1}{(4\pi t)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4t}} \tag{1.2}$$

and $|\cdot|$ denotes the Euclidean norm. Now consider a Brownian particle that starts at $x \in \mathbb{R}^d$, i.e. a solution to the following stochastic differential equation (SDE)

$$\begin{cases} \mathrm{d}X_t = \sqrt{2}\,\mathrm{d}W_t \\ X_0 = x \end{cases} \tag{1.3}$$

where $(W_t)_{t\geq 0}$ is a d-dimensional Wiener process, i.e. $W_t = \left(W_t^1, \dots, W_t^d\right)$ and $(W_t^i)_{t\geq 0}$ are independent Brownian motions. Then one has the following two consequences of the Itô formula (cf. [**9**, p.5]):

$$\mathbb{E}[f(X_t)|X_0 = x] = \int_{\mathbb{R}^d} p_t(x-y) f(y)\,\mathrm{d}y$$

for all bounded and measurable $f : \mathbb{R}^d \to \mathbb{R}$ and, in particular,

$$\mathbb{P}(X_t \in B | X_0 = x) = \int_B p_t(x-y)\,\mathrm{d}y$$

for all Borel sets $B \subset \mathbb{R}^d$. This already hints at a close relationship of the PDE (1.1) and the SDE (1.3). Indeed, let $N \in \mathbb{N}$ and let $(X_t^i)_{t\geq 0}$ be independent solutions to (1.3) for $i = 1, \dots, N$. Then we define the *empirical measure*

$$\rho_N(t) := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^i} \tag{1.4}$$

and notice that by the law of large numbers, for any bounded and measurable $f : \mathbb{R}^d \to \mathbb{R}$ we have

$$\langle \rho_N(t), f \rangle := \frac{1}{N} \sum_{i=1}^N f(X_t^i) \overset{N\to\infty}{\longrightarrow} \mathbb{E}[f(X_t)|X_0 = x] \quad \text{a.s..} \tag{1.5}$$

In other words, (1.5) implies

$$\rho_N(t) \overset{w}{\rightharpoonup} P_t^* \delta_x \quad \text{a.s.} \tag{1.6}$$

where $(P_t^*)_{t\geq 0}$ is the dual semigroup, i.e., in particular, we have $P_t^* \delta_x(f) = P_t f(x)$.

1.2. Gradient flows

Let u be a solution to (1.1). We integrate by parts to compute

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\int_{\mathbb{R}^d}|\nabla u|^2\,\mathrm{d}x = \int_{\mathbb{R}^d}\nabla u\cdot\nabla\partial_t u\,\mathrm{d}x = -\int_{\mathbb{R}^d}|\Delta u|^2\,\mathrm{d}x = -\int_{\mathbb{R}^d}|\partial_t u|^2\,\mathrm{d}x \le 0. \tag{1.7}$$

This shows that under the flow of (1.1) the *Dirichlet energy*

$$E(u) := \frac{1}{2}\int_{\mathbb{R}^d}|\nabla u|^2\,\mathrm{d}x \tag{1.8}$$

decreases and the rate of the *dissipation* of the energy is measured in the L^2-geometry. This suggests that the diffusion equation (1.1) is an instance of a *gradient flow*. Indeed, (1.1) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}u = -\nabla E(u)$$

where $\nabla E(u) \in L^2(\mathbb{R}^d)$ is the L^2-gradient defined via the duality

$$\mathrm{diff}E|_u.\dot{u} := \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0}E(u+\varepsilon\dot{u}) = (\nabla E(u),\dot{u})_{L^2(\mathbb{R}^d)}.$$

for all sufficiently nice test functions $\dot{u}$.

In general, a gradient flow structure consists of a triple $(\mathcal{M}, E, g)$ where $\mathcal{M}$ is a configuration space, $E : \mathcal{M} \to \mathbb{R}$ is an energy and g is a metric tensor on $\mathcal{M}$[1]. This data gives rise to an ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}u = -\nabla E(u) \tag{1.9}$$

which is then referred to as a gradient flow, where ∇ denotes the Riemannian gradient with respect to g. The interpretation of (1.9) is that

$$\mathrm{diff}E|_u.\dot{u} + g_u(\partial_t u, \dot{u}) = 0 \tag{1.10}$$

for all $\dot{u} \in T_u\mathcal{M}$. The choice $\dot{u} = \partial_t u$ in (1.10) yields the *energy estimate*

$$\frac{\mathrm{d}}{\mathrm{d}t}E(u) = -g_u(\partial_t u, \partial_t u). \tag{1.11}$$

The discussion above shows that the diffusion equation (1.1) is a gradient flow with respect to the L^2-geometry and the Dirichlet energy (1.8). The abstract energy estimate (1.11) corresponds to (1.7). While the Hilbertian structure that arises from the Dirichlet energy is certainly pertinent to (1.3) – the *Cameron–Martin space* of Brownian motion is essentially the (homogeneous) Sobolev space $H^1(\mathbb{R}^d)$ (cf. [**103**, Theorem (2.2), p.339]) – there is another gradient flow structure of the heat equation (1.1) revealing an even deeper connection.

[1]more precisely on the tangent bundle $T\mathcal{M} \otimes T\mathcal{M}$

1.3. Wasserstein geometry

Again, let u be a solution to (1.1), that we think of as being positive. Integrating by parts yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d} u\ln u\,\mathrm{d}x = \int_{\mathbb{R}^d}\partial_t u(\ln u+1)\,\mathrm{d}x = \int_{\mathbb{R}^d}\Delta u(\ln u+1)\,\mathrm{d}x = -\int_{\mathbb{R}^d}\frac{|\nabla u|^2}{u}\,\mathrm{d}x = -\int_{\mathbb{R}^d} u|\nabla\ln u|^2\,\mathrm{d}x. \tag{1.12}$$

Hence, under the flow of the diffusion equation (1.1) moreover the *entropy*

$$\mathrm{Ent}(u) := \int_{\mathbb{R}^d} u\ln u\,\mathrm{d}x \tag{1.13}$$

decreases. In this case the geometry is more complicated. We first of all specify a configuration space

$$\mathcal{M} := \left\{u:\mathbb{R}^d\to\mathbb{R} : u>0, \int_{\mathbb{R}^d} u\,\mathrm{d}x = 1\right\}, \tag{1.14}$$

and for any $u\in\mathcal{M}$ the tangent space

$$T_u\mathcal{M} := \left\{\dot{u}:\mathbb{R}^d\to\mathbb{R} : \int_{\mathbb{R}^d}\dot{u}\,\mathrm{d}x = 0\right\}.$$

We next observe that $\Delta u = \nabla\cdot(u\nabla\ln u)$ which motivates the following definition in view of (1.12). The *Wasserstein metric tensor* is given by

$$g_u(\dot{u},\dot{u}) := \int_{\mathbb{R}^d} u|\nabla p|^2\,\mathrm{d}x, \quad \text{where} \quad \dot{u}+\nabla\cdot(u\nabla p) = 0 \tag{1.15}$$

for $\dot{u}\in T_u\mathcal{M}$. Then it is easy to see that one indeed has

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(u) = -g_u(\partial_t u,\partial_t u)$$

and, moreover, for any $\dot{u}\in T_u\mathcal{M}$

$$\mathrm{diff}\,\mathrm{Ent}|_u.\dot{u} + g_u(\partial_t u,\dot{u}) = 0.$$

In words, the diffusion equation (1.1) is a gradient flow with respect to the entropy and the Wasserstein geometry. This was first observed in [**69**] in the more general context of Fokker–Planck equations, and in [**95**] the metric tensor (1.15) has been first introduced. In recent years this has triggered a vast research interest in gradient flows in general, and Wasserstein gradient flows in particular. We refer to [**5**], [**109**], [**110**], to only mention a few.

The metric tensor (1.15) is referred to as the Wasserstein metric tensor due to the fact that it generates the 2-Wasserstein distance (cf. [**115**, Definition 7.1.1.]) defined as

$$W_2^2(\lambda_1, \lambda_2) := \inf\Big\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x, \mathrm{d}y) : \pi\Big(\mathbb{R}^d \times \cdot\Big) = \lambda_1, \pi\Big(\cdot \times \mathbb{R}^d\Big) = \lambda_2 \Big\}, \tag{1.16}$$

at least if restricted to measures which are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. Indeed, by the *Benamou – Brenier* formula[2] one has (cf. [**13**], [**115**, Theorem 8.1])

$$\begin{aligned} W_2^2(f_0, f_1) &= \inf_{f,v} \Big\{ \int_0^1 \int_{\mathbb{R}^d} f_t |v_t|^2 \,\mathrm{d}x \,\mathrm{d}t : \partial_t f_t + \nabla \cdot (f_t v_t) = 0, f|_{t=i} = f_i, i = 0, 1 \Big\} \\ &= \inf_f \Big\{ \int_0^1 g_{f_t}(\partial_t f_t, \partial_t f_t) \,\mathrm{d}t : f|_{t=i} = f_i, i = 0, 1 \Big\}. \end{aligned}$$

The second inequality follows from the fact that if v is a minimizer it is divergence-free and hence must be a gradient (cf. [**95**, (22)]). In other words, the Wasserstein metric tensor gives rise to the 2-Wasserstein distance in the same way that a general Riemannian metric gives rise to a distance on a Riemannian manifold.

1.3.1. More examples of gradient flows. The diffusion equation (1.1) is an example of a PDE that admits many[3] gradient flow structures. We want to introduce two more gradient flows that play a role in this thesis. These are the *thin-film equation* and the *Allen–Cahn equation*.

1.3.1.1. *The thin-film equation.* The thin-film equation on $\mathbb{R}^d$ is given by

$$\begin{cases} \partial_t h + \nabla \cdot (M(h) \nabla \Delta h) = 0 \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d \\ h|_{t=0} = h_0 \end{cases} \tag{1.17}$$

for some initial condition $h_0 > 00$. We again think of the solution h as being positive. The equation (1.17) is a fourth-order, parabolic, quasi-linear and degenerate PDE that describes the time evolution of the height function h of a very thin liquid on a solid. Here $M(h)$ is referred to as the mobility and it describes the boundary conditions of the fluid velocity at the solid-liquid interface. We are mostly interested in a power-law type non-linearity, i.e. $M(h) = h^m$. Indeed, the most relevant case is $M(h) = h^3$ corresponding to Dirichlet boundary conditions. For a more in-depth discussion we refer to Section 3.3.

A quick computation shows that (1.17) is a gradient flow on the configuration space $\mathcal{M}$ (1.14) with respect to the Dirichlet energy E (1.8) and the *generalized* Wasserstein metric tensor given by

$$g_h(\dot{h}, \dot{h}) := \int_{\mathbb{R}^d} M(h) |\nabla \varphi|^2 \,\mathrm{d}x, \quad \text{where} \quad \dot{h} + \nabla \cdot (M(h) \nabla \varphi) = 0$$

[2] modulo some technicalities

[3] in fact infinitely many

for $\dot{h} \in T_h\mathcal{M}$ (cf. [**4**], [**94**]). A similar metric tensor, for concave M, has been considered in [**35**] and [**20**]. Since (1.17) is derived by a lubrication approximation of the Navier-Stokes equations in $d = 3$ (cf. [**50**]), the natural dimension for (1.17) is $d = 2$. By assuming a shear flow, it is also not unreasonable to consider (1.17) for $d = 1$. In the case of $d = 1$ non-negative weak solutions to (1.17) have been first obtained in [**16**].

A major interest concerning equation (1.17) is that of preservation of positivity. In the simplest form this would amount to the question if

$$h_0 > 0 \implies h_t > 0 \quad \text{for} \quad t > 0. \tag{1.18}$$

We first note that (1.18) depends crucially on the degeneracy of the mobility $M(h)$. Indeed, there is no comparison principle for fourth-order equations and in the case of $M(h) = 1$ there exists a counter-example such that (1.18) does not hold (cf. [**17**]). On the other hand, for $d = 1$, it was shown in [**14**] that (1.18) holds if $M(h) = h^m$ for $m \geq \frac{7}{2}$. Hence, an important question remains if positivity is preserved in the case of $M(h) = h^3$. In the case of $d = 2$ much less is known. One should note that in general (1.17) must be seen as a so-called *free boundary problem* with free boundary given by $\partial\{h > 0\}$ where additional boundary conditions need to be imposed (cf. [**49**]).

Moreover, there has been important developments concerning numerical discretizations of (1.17). This includes, in particular, the articles [**54**] and [**121**] where the authors propose a discretization that preserves the so-called entropy estimate which is in turn related to the question of positivity (1.18). Indeed, in their discretization, the property (1.18) holds for $M(h) = h^m$ if $m \geq 2$. However, this result is not uniform in the discretization parameter and thus can not be used to infer the same result for the continuum case.

1.3.1.2. *The Allen–Cahn equation.* The Allen–Cahn equation on $\mathbb{R}^d$ is given by

$$\begin{cases} \partial_t u = \Delta u - \frac{1}{\varepsilon^2} W'(u) & \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d \\ u|_{t=0} = u_0 \end{cases} \tag{1.19}$$

for some initial condition u_0 and a potential $W : \mathbb{R} \to \mathbb{R}$. The PDE (1.19) arises as the gradient flow of the energy defined by

$$F_\varepsilon(u) := \int_{\mathbb{R}^d} \frac{\varepsilon}{2}|\nabla u|^2 + \frac{1}{\varepsilon} W(u)\,\mathrm{d}x \tag{1.20}$$

and the L^2-inner product (modulated by the factor $\frac{1}{\varepsilon}$).

A typical example for W is the *double-well potential*, i.e. $W(u) := \frac{1}{4}(1 - u^2)^2$. Hence, in this case, the minimizer of F_ε should concentrate on the set $\{-1, 1\}$ as $\varepsilon \to 0$. Indeed, in [**88**] it was shown that if F_ε is posed on a bounded domain, then it converges[4] to the

[4] in the sense of Γ-convergence to be precise

perimeter functional of that domain as $\varepsilon \to 0$. Moreover, there is a close connection to *mean curvature flow*[5] which describes the evolution of surfaces according to their mean curvature (cf. [**10**, Section 6.1.3]). It was also recently shown that if one starts (1.19) with sufficiently mixing (and random) initial data, then after some time the generated fronts also evolve according to mean curvature flow (cf. [**60**]).

1.4. Fluctuating gradient flows

For a given gradient flow we now want to introduce randomness into the system which will usually mean a random forcing term involving white noise. To this end, consider again the empirical measure ρ_N (cf. (1.4)). Following the (formal) computation in [**32**][6] one finds that ρ_N is formally an exact solution to the d-dimensional *Dean–Kawasaki* equation

$$\partial_t \rho = \Delta\rho + \sqrt{\frac{2}{N}}\,\nabla\cdot(\sqrt{\rho}\,\eta) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^d \tag{1.21}$$

where η is d-dimensional space-time white noise. Note that for $N \to \infty$, the equation (1.21) formally converges to the deterministic diffusion equation. This reflects again the fact that for N large enough, the empirical measure ρ_N can be reasonable well be approximated by the diffusion equation (1.1) (cf. (1.6)). Hence, the advantage of the Dean–Kawasaki equation is that it describes exactly the behaviour of N independent Brownian particles while the disadvantage is that it is a much more complicated and stochastic PDE. While (1.6) provides a *macroscopic* description of the particle density – it is purely deterministic – the description of (1.21) is *mesoscopic* and retains randomness. Indeed, the precise meaning of (1.21) is debatable. Due to the low regularity of the space-time white noise, the product $\sqrt{\rho}\eta$ does not a priori make sense, even for $d = 1$. In fact, (1.21) is an instance of a singular SPDE and even in this class stands out since it is a *supercritical* equation, which, roughly speaking, means that the non-linearity only becomes more pronounced on small scales. In this case there are no known techniques in order to give a pathwise sense to (1.21). Nevertheless, (1.21) has been the object of a vast amount of research. In [**75**] it was shown that even weak solutions, i.e. solutions to the martingale problem corresponding to (1.21), are sums of Dirac measures. This poses the question if (1.21) is just a complicated way of speaking about the empirical measure. The recent work [**24**] establishes that suitable discretizations of (1.21) and (1.4) are arbitrarily close in a weak norm, meaning that (1.21) can be used in order to approximate the diffusion of independent Brownian particles. Moreover, in [**41**] a solution theory for (1.21) with colored noise is established whereas in [**40**] a large deviations principle in the scaling regime $\frac{1}{N} \ll \frac{1}{K}$, where K denotes the cut-off in the noise, is proved.

[5]which coincidentally is also a gradient flow

[6]and omitting the potential V

1.4.1. The fluctuation–dissipation theorem. We now want to explain a different way to derive (1.21). Consider again the diffusion equation (1.1) and its gradient flow structure with respect to the entropy and the Wasserstein geometry, i.e. we write (1.1) as

$$\partial_t u = \nabla \cdot \left(u \nabla \frac{\delta}{\delta u} \mathrm{Ent}(u) \right)$$

where $\frac{\delta}{\delta u}$ denotes the L^2-derivative. Formally, in Riemannian geometry the Riemannian gradient is given by

$$\nabla_g = g^{-1} \nabla \tag{1.22}$$

where ∇ denotes the standard Euclidean gradient. One can make sense of (1.22) in local coordinates, but otherwise the inverse of a Riemannian metric tensor g or generally of a bilinear form does not make any sense. What we mean by g^{-1} is the inverse of the operator corresponding to g with respect to a reference metric tensor given by the L^2-inner product[7]. More precisely, integrating by parts in (1.15) yields

$$g_u(\dot{u}, \dot{u}) = \int_{\mathbb{R}^d} \dot{u} \mathcal{K}_u^{-1} \dot{u} \, \mathrm{d}x$$

where

$$\mathcal{K}_u \dot{u} := \nabla \cdot (u \nabla \dot{u}) \tag{1.23}$$

and then we identify $\mathcal{K}_u$ with g_u^{-1}. Often, the operator $\mathcal{K}$ is referred to as *Onsager operator*. With this notation in hand the diffusion equation (1.1) takes the form

$$\partial_t u = -\mathcal{K}_u \frac{\delta}{\delta u} \mathrm{Ent}(u).$$

The *fluctuation-dissipation theorem* (cf. [**122**, (1.57)]) proposes that in order to introduce fluctuations in a meaningful way – which will be explained shortly – a noise must be added to the equation whose covariance structure matches the geometry. More precisely, that means that fixing the gradient flow structure with respect to the entropy and the Wasserstein geometry, the diffusion equations involving fluctuations takes the form

$$\begin{aligned} \partial_t u &= -\mathcal{K}_u \frac{\delta}{\delta u} \mathrm{Ent}(u) + \sqrt{2\beta^{-1}} \sqrt{\mathcal{K}_u} \eta \\ &= \Delta u + \sqrt{2\beta^{-1}} \nabla \cdot (\sqrt{u}\, \eta) \end{aligned} \tag{1.24}$$

where η is a vector-valued space-time white noise and $\beta > 0$. The operator $\sqrt{\mathcal{K}_u}$ is the square root of the positive-definite and symmetric operator $\mathcal{K}_u$. Thus we have recovered the Dean–Kawasaki equation[8] (1.21). Adding fluctuations in this way suggests that the

[7] in finite-dimensions this is the matrix representation of g

[8] this has already been noted in Dean's original paper [**32**]

equation (1.24) has an invariant measure and this invariant measure is given by the *Gibbs measure*

$$\frac{1}{Z}e^{-\beta \mathrm{Ent}(u)}\,\mathrm{d}u \tag{1.25}$$

where Z is a normalization constant which makes the measure (1.25) into a probability measure and $\mathrm{d}u$ formally is the Lebesgue measure on the corresponding configuration space: in this case (1.14). In order for (1.25) to be invariant for (1.24), the stochastic integral has to be chosen accordingly. Since (1.14) is infinite-dimensional, it is known that there exists no such Lebesgue measure. Nevertheless, the construction of (1.25) was addressed in [**117**].

For a general gradient flow given by an energy E and the geometry induced by an Onsager operator $\mathcal{K}$, the fluctuation-dissipation theorem suggests a corresponding stochastic gradient flow of the form

$$\partial_t u = -\mathcal{K}_u \frac{\delta}{\delta u}E(u) + \sqrt{2\beta^{-1}}\sqrt{\mathcal{K}_u}\eta \tag{1.26}$$

with Gibbs measure given by

$$\mathrm{d}\nu_s := \frac{1}{Z}e^{-\beta E(u)}\,\mathrm{d}u \tag{1.27}$$

where Z is a normalization constant and $\mathrm{d}u$ is formally the Lebesgue measure on the associated configuration space. Again, in general, the stochastic integral in (1.26) has to be chosen in such a way that (1.27) is the invariant measure.

In order to illustrate the relevance of different gradient flow structures, let us again consider the diffusion equation as a gradient flow with respect to the L^2-metric and the Dirichlet energy E (1.8). Applying the fluctuation–dissipation theorem yields the *stochastic heat equation*

$$\partial_t v = \Delta v + \sqrt{2\beta^{-1}}\xi \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^d \tag{1.28}$$

where ξ is (real-valued) space-time white noise and the postulated invariant measure is given by

$$\mathrm{d}\mu(v) := \frac{1}{Z}e^{-\beta E(v)}\,\mathrm{d}v. \tag{1.29}$$

Here, μ can be interpreted as a Gaussian measure. Indeed, $E(v) = \frac{1}{2}(-\Delta v, v)_{L^2(\mathbb{R}^d)}$ is a quadratic functional, and hence μ is Gaussian with covariance operator $(-\beta\Delta)^{-1}$. This covariance operator corresponds to the so-called *Gaussian free field*[9] (cf. [**107**]). The equation (4.10) is also known as the *Edward–Wilkinson process* (cf. [**85**]) and plays an important role in the study of subcritical singular SPDE.

[9]most often, this measure is considered on a bounded domain for $d = 2$

The fluctuation-dissipation theorem is closely related to the notion of *detailed balance* (cf. [**100**, (4.97)]). Assuming that (1.26) has a unique invariant measure given by (1.27), then formally one has $\mathcal{L}^*\rho_s = 0$ where $\rho_s(u) := \frac{1}{Z}e^{-\beta E(u)}$ and $\mathcal{L}$ is the *Kolmogorov operator*, also referred to as the generator of (1.26). Then

$$\partial_t \rho = \mathcal{L}^* \rho \tag{1.30}$$

is the *Fokker–Planck equation.* A stochastic process satisfies the detailed balance condition (cf. [**100**, (4.97)]) if its associated Fokker–Planck equation (1.30) can be written as a continuity equation[10]

$$\partial_t \rho = \nabla \cdot J(\rho)$$

with *probability flux* J and if, moreover,

$$J(\rho_s) = 0.$$

Under this condition the operator $\mathcal{L}$ is symmetric on $L^2(\mathrm{d}\nu_s)$ and it is well-known that this implies that the resulting stochastic process is time-reversible (cf. [**100**, Section 4.6], [**37**, (49)]). In thermodynamics the interpretation is that in equilibrium it should not be possible to tell if time is going forward or backward. We want to stress that the preceding discussion is very formal on the level of infinite-dimensional stochastic gradient flows. In that setting, the Kolmogorov operator $\mathcal{L}$ is a priori only defined on cylindrical functions and if the equation in question requires renormalization it is in general not possible to specify $\mathcal{L}$. In the finite-dimensional setting, on the other hand, this is classical (cf. [**100**]).

1.4.1.1. *The stochastic thin-film equation.* We have already seen that the thin-film equation (1.17) is a gradient flow with respect to the Dirichlet energy E (1.8) and the Wasserstein metric tensor g (1.15) on the configuration space (1.14). Hence it can be written as

$$\partial_t h = -\mathcal{K}_h \frac{\delta}{\delta h} E(h)$$

where $\mathcal{K}_h \dot{h} := \nabla \cdot \left(M(h) \nabla \dot{h} \right)$ (cf. (1.23)). According to the fluctuation-dissipation theorem, the corresponding *stochastic thin-film equation* (cf. [**30**]) takes the form

$$\begin{aligned} \partial_t h &= -\mathcal{K}_h \frac{\delta}{\delta h} E(h) + \sqrt{2\beta^{-1}} \sqrt{\mathcal{K}_h} \eta \\ &= -\nabla \cdot (M(h) \nabla \Delta h) + \sqrt{2\beta^{-1}} \nabla \cdot \left(\sqrt{M(h)} \eta \right) \end{aligned} \tag{1.31}$$

[10] here $\nabla \cdot$ might refer to a infinite-dimensional divergence

where η is vector-valued space-time white noise. The invariant measure is given by

$$\mathrm{d}\nu(h) = \frac{1}{Z} e^{-\beta E(h)} \,\mathrm{d}h. \tag{1.32}$$

On first sight, the interpretation of (1.32) appears to be the same as for the stochastic heat equation (4.10), i.e. that $\mathrm{d}\nu$ looks like the Gaussian free field (1.29). This is where the configuration space makes a difference: while $\mathrm{d}u$ formally denotes the Lebesgue measure on $L^2(\mathbb{R}^d)$, $\mathrm{d}h$ denotes the Lebesgue measure on $\mathcal{M}$.Hence, the interpretation of (1.32) is that of a Gaussian free field conditioned to be positive with spatial average being one. For more details we refer to Section 3.4. Also, (1.31) is a singular SPDE and the renormalization should be chosen in such a way that (1.32) is the invariant measure of (3.13). See also Section 3.4.2.

1.4.1.2. *The φ^4_d-model.* Euclidean quantum field theory is concerned with the construction of Gibbs measures of the form

$$\mathrm{d}\nu(u) = \frac{1}{Z} e^{-F(u)} \,\mathrm{d}u \tag{1.33}$$

where F is given by (1.20) (for $\varepsilon = 1$) and $\mathrm{d}u$ is formally the Lebesgue measure on $L^2(\mathbb{R}^d)$ [11]. If $W(u) = \frac{1}{4}|u|^4$ [12], the measure ν is called the *φ^4_d-measure.* Since, again, the Lebesgue measure on $L^2(\mathbb{R}^d)$ does not exist, the interpretation of ν is[13]

$$\mathrm{d}\nu(u) = \frac{1}{Z} e^{-\int_{\mathbb{R}^d} \frac{1}{4}|u|^4 \,\mathrm{d}x} \,\mathrm{d}\mu(u)$$

where μ is the Gaussian free field (1.29). It is well-known that the support of the Gaussian free field is the Hölder space $C^{\frac{2-d}{2}-\varepsilon}$ for any $\varepsilon > 0$ [14]. Since for $d \geq 2$ this is a space of distributions the term $|u|^4$ is a priori ill-defined and hence it is expected that renormalization is necessary. In [**99**] it was suggested to sample from (1.33) by writing down a stochastic gradient flow via the fluctuation-dissipation theorem that has (1.33) as an invariant measure. This procedure is now known as *stochastic quantization* and it gives rise to the *φ^4_d-model*

$$\partial_t u = \Delta u - u^3 + \sqrt{2}\,\xi \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d \tag{1.34}$$

where ξ is space-time white noise. The geometry is chosen to be $L^2(\mathbb{R}^d)$ which means that the Onsager operator is nothing but the identity. We want to note that in the case of the stochastic thin-film equation, at least for $d = 1$, the postulated invariant measure

[11] of particular interest is the case $d = 4$

[12] somewhat of a simplification of the double-well potential

[13] in $d = 3$ the measure has been constructed but it is not absolutely continuous with respect to the Gaussian free field. This ansatz still gives the correct heuristic.

[14] This can be seen by a scaling argument and a variant of Kolmogorov's continuity theorem. In [**107**] it is proven that the Gaussian free field for $d \geq 2$ can only be realized as a distribution which is enough for our argument.

is not in need of a renormalization whereas in this case the invariant measure has to be renormalized. The φ^4_3-model has been considered in [**59**]. For more details on the φ^4_d-model, in particular for $d = 2$, we refer to Chapter 4.

1.4.2. Singular SPDEs. We have already seen SPDEs, i.e. (1.21), (1.31), (1.34), that have been referred to as being *singular*; let us now explain this notion. The following theorem is concerned with the multiplication of distributions (cf. [**6**, Theorem 2.52], Lemma 4.17).

Theorem 1.1. *Let $\varphi \in C^\alpha$ and $\psi \in C^\beta$. Then it holds that $\varphi \cdot \psi \in C^{\min\{\alpha,\beta\}}$ if and only if $\alpha + \beta > 0$.*

Here, the function space C^α for $\alpha \in \mathbb{R}$ denotes a Hölder spaces of possibly negative regularity, which usually is identified with the Besov space $B^\alpha_{\infty,\infty}$ (cf. [**6**, Section 2]). To illustrate Theorem 1.1 we consider again a one-dimensional Brownian motion $\{B_t\}_{t\geq 0}$. It is well-known[15] that $B \in C^{\frac{1}{2}-\varepsilon}$ for all $\varepsilon > 0$[16]. The theory of stochastic integration is essentially concerned with the product $B_t \frac{\mathrm{d}}{\mathrm{d}t} B_t$ where the derivative is in the sense of distributions. Since $\frac{\mathrm{d}}{\mathrm{d}t} B \in C^{-\frac{1}{2}-\varepsilon}$, the product is not-well defined according to Theorem 1.1 since $\frac{1}{2} - \varepsilon + \left(-\frac{1}{2} - \varepsilon\right) = -2\varepsilon < 0$. In order to give sense to this product probabilistic techniques are necessary.

Motivated by this observation, a SPDE is called singular if there are products in the equation which are not well-defined since their regularity does not sum up to a positive number. This is due to the low regularity of the space-time white noise. A space-time white noise on $\mathbb{R}_+ \times \mathbb{R}^d$ is a Gaussian centered random field $\{\xi(f)\}_{f \in L^2(\mathbb{R}_+ \times \mathbb{R}^d)}$ with covariance given by

$$\mathbb{E}\left[\xi(f)\xi(f')\right] = \left(f, f'\right)_{L^2(\mathbb{R}_+ \times \mathbb{R}^d)}$$

which is often informally written as

$$\mathbb{E}[\xi(t,x)\xi(s,y)] = \delta(t-s)\delta(x-y). \tag{1.35}$$

Writing the covariance as in (1.35) is misleading, since ξ cannot be realized as a (random) function. The equation (1.1) is invariant under the parabolic scaling $(t,x) \mapsto (\lambda^2 t, \lambda x)$ ($\lambda > 0$) and hence this is the natural scaling that we want to assume when treating second-order SPDEs like (1.21) or (1.34). Then the rescaled space-time white noise $\xi(t,x) \mapsto \xi_\lambda(t,x) := \xi(\lambda^2 t, \lambda x)$[17] has the covariance

$$\mathbb{E}\left[\xi_\lambda(f)\xi_\lambda(f')\right] = \lambda^{-(d+2)}\left(f, f'\right)_{L^2(\mathbb{R}_+ \times \mathbb{R}^d)}$$

[15] by Kolmogorov's continuity theorem
[16] on a finite time-interval $[0, T]$
[17] this is meant in the sense of distributions

and hence $\hat{\xi}(t,x) := \lambda^{\frac{d+2}{2}}\xi(\lambda^2 t, \lambda x)$ and $\xi(t,x)$ have the same covariance. Existence of ξ is established by Kolmogorov's existence theorem (cf. [**103**, Theorem 3.2]) and then the distribution of ξ is uniquely determined by its covariance, implying that $\hat{\xi}$ and ξ have the same distribution. This scale invariance suggests that by Kolmogorov's continuity theorem there is a version of ξ, which, by abuse of notation, we will also refer to as ξ that satisfies (locally) $\xi \in C^{-\frac{d+2}{2}-\varepsilon}$ for all $\varepsilon > 0$[18] (cf. [**102**, Lemma 18])[19].

Let us again consider the φ^4_d-model. Rescaling (1.34) according to

$$\lambda x = \hat{x}, \quad \lambda^2 t = \hat{t}, \quad \lambda^{\frac{d-2}{2}} u = \hat{u}, \quad \lambda^{\frac{d+2}{2}} \xi = \hat{\xi}$$

yields

$$\partial_{\hat{t}} \hat{u} = \hat{\Delta}\hat{u} - \lambda^{4-d}\, \hat{u}^3 + \hat{\xi} \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d.$$

Trivially, we have the following behavior

$$\lim_{\lambda \to 0} \lambda^{4-d} = \begin{cases} 0, & d < 4 \\ 1, & d = 4 \\ \infty, & d > 4, \end{cases}$$

which, in words, means that zooming in on small scales for $d < 4$ has the effect that the non-linearity vanishes, for $d = 4$ it has no effect and for $d > 4$ it has the effect that the non-linearity blows up. If the non-linearity vanishes on small scales we call the singular SPDE *subcritical*, if the rescaling leaves the equation invariant we call it *critical* and otherwise the equation is called *supercritical*. If the equation is subcritical it locally *looks like* the stochastic heat equation (4.10). For (4.10) standard Schauder estimates yield that the solution v satisfies $v \in C^{\frac{2-d}{2}-\varepsilon}$ for all $\varepsilon > 0$ (cf. [**108**], [**58**, Lemma A.9]). Since regularity is a local property after all, this suggests that the solution u of the φ^4_d-model, for $d < 4$, i.e. if it is subcritical, satisfies $u \in C^{\frac{2-d}{2}-\varepsilon}$ for all $\varepsilon > 0$. In the case of $d = 1$, u is actually a function and there is no problem defining the non-linearity u^3. However, for $d = 2$ it is expected that u is not a function anymore and hence by Theorem 1.1 multiplication with itself is not possible and the φ^4_d-model (1.34) is a singular SPDE for $d > 1$. Other examples of subcritical singular SPDEs include the KPZ equation and the stochastic thin-film equation in $1 + 1$ dimensions. In $2 + 1$ dimensions the KPZ equation and the stochastic thin-film equation are critical. The Dean–Kawasaki equation is supercritical in any dimension.

In recent years there were many developments in the field of singular SPDEs. One of the earliest examples of a successful solution theory of a singular SPDE has been [**25**] where

[18] this refers to a parabolic (negative) Hölder-space defined via a scale-invariant metric
[19] Note the reminiscence to Sobolev-embeddings

the φ_2^4-model has been considered. The main insight in this article was to decompose the solution into a irregular, but linear part, namely the solution to the stochastic heat equation, and a regular, but non-linear part. This is known as the *Da Prato–Debussche trick*. In order to proceed with the latter, a renormalization procedure is necessary to define powers of the stochastic heat equation. At the same time, the theory of rough paths (cf. [**83**], [**55**]) was being developed which, in a similar way, was concerned with a pathwise treatment of stochastic integrals like $B_t \frac{\mathrm{d}}{\mathrm{d}t} B_t$. One of the main insights was to define the stochastic integral *off-line*, for example by the usual Itô-calculus, which together with a suitable topology made the solution map of a standard SDE continuous. A vast generalization of these ideas has been the development of *regularity structures* (cf. [**59**]). Like in the theory of rough paths, the idea is to define certain *trees* off-line which also possibly involves renormalization; in the φ_2^4-model this would correspond to the square and the cubic power of the stochastic heat equation. Then the solution can be considered as a functional of the trees. At the same time, another approach using *paracontrolled distributions* (cf. [**58**]) was developed relying on Fourier theoretic concepts. In particular, in paracontrolled distributions a singular product is considered as a paraproduct and then uses an ansatz for the solution known from *controlled rough paths* (cf. [**55**]). Also, a variant of regularity structures (cf. [**81**]) for quasi-linear equations has been introduced which could be used in order to study the stochastic thin-film equation.

A unifying principle that underlies both approaches is that there is a clear distinction between a probabilistic step which involves the off-line construction of the auxiliary elements, and the analytic step which consists of the solution theory taking the off-line products into account. The probabilistic step usually also consists of a renormalization procedure which in the analytic step is reflected in the form of a counter term. This counter term is in general not unique and corresponds to choosing a stochastic integral. Indeed, in [**19**] the renormalization group has dimension 54 and the authors identify various subspaces corresponding to Itô-integration, Stratonovich-integration or both! For the stochastic thin-film equation we would be interested in a counter term that ensures that the postulated invariant measure (1.32) is indeed invariant for the equation (1.31). In the case of the φ_2^4-model on the torus this has been addressed in [**111**].

1.4.3. The theory of large deviations and gradient flows. The theory of *large deviations* (cf. [**45**], [**114**]) delivers yet another justification to consider the diffusion equation (1.1) as a gradient flow with respect to the Wasserstein geometry and the entropy. Informally, a sequence of real valued stochastic process $\left\{Y_t^N\right\}_{t\geq 0, N\in\mathbb{N}}$ is said to satisfy a (conditional) large deviations principle with rate functional I if (cf. [**1**, (2)])

$$\mathbb{P}\Big(Y_t^N \simeq \lambda | Y_0^N \simeq \lambda_0\Big) \sim e^{-N I_t(\lambda;\lambda_0)} \quad \text{as} \quad n \to \infty \tag{1.36}$$

for some probability measures λ, λ_0 on $\mathbb{R}$. For the precise definition of a large deviations principle we refer to [**45**] and [**114**]. In [**1**] the authors were concerned with a large deviations principle for the empirical measure ρ_N (cf. (1.4)). Indeed, they state a large deviations principle for ρ_N (cf. [**1**, Theorem 1, Lemma 2]) in the spirit of (1.36) with rate functional given by

$$I_t(\lambda, \lambda_0) := \inf_{\Lambda} H(\Lambda|\Lambda_0)$$

where the infimum runs over all probability measures Λ on $\mathbb{R}^2$ that have λ and λ_0 as marginals, $\Lambda_0(\mathrm{d}x\,\mathrm{d}y) := \lambda_0(\mathrm{d}x)p_t(x-y)\,\mathrm{d}y$ and p_t is given by (1.2) for $d = 1$. Moreover, H denotes the *relative entropy*, i.e.

$$H(\Lambda|\Lambda_0) := \begin{cases} \int_{\mathbb{R}^2} f(x,y)\log f(x,y)\Lambda_0(\mathrm{d}x, y), & \text{if} \Lambda \ll \Lambda_0, f := \frac{\mathrm{d}\Lambda}{\mathrm{d}\Lambda_0} \\ +\infty, & \text{otherwise.} \end{cases}$$

Then their main observation is (cf. [**1**, Theorem 3]) that

$$I_t(\cdot\,;\lambda_0) \simeq \frac{1}{2}K_t(\cdot\,;\lambda_0) \quad \text{as} \quad t \to 0 \tag{1.37}$$

where

$$K_t(\lambda;\lambda_0) := \frac{1}{2t}W(\lambda,\lambda_0)^2 + \mathrm{Ent}(\lambda) - \mathrm{Ent}(\lambda_0),$$

W is the 2-Wasserstein distance (1.16) and Ent is the entropy (1.13) (for $d = 1$). Thinking of t now as a time-step reveals the deep insight of (1.37) connecting the theory of large deviations and gradient flows. In [**69**] it was shown that under mild conditions on the initial condition λ_0 the iterative minimization scheme (cf. [**1**])

$$\lambda_n \in \operatorname*{argmin}_{\lambda} K_t(\lambda, \lambda_{n-1})$$

has a unique solution and the piecewise (in time) constant interpolation converges weakly to the solution of (1.1) with initial condition λ_0. As pointed out in [**2**, Section 6], (1.37) also implies the variant of (1.36)

$$\mathbb{P}(\rho_N(t) \simeq \rho | \rho_N(0) \simeq \rho_0) \sim e^{-N\frac{W(\rho,\rho_0)^2}{4t}} \quad \text{as} \quad N \to \infty$$

which explicitly connects Wasserstein geometry and Brownian particles[20] (cf. [**101**, Section 4]). The authors of [**2**, Conjecture 6.1] conjecture that any gradient flow structure is connected to a stochastic process by a large deviations principle. For example, in [**34**]

[20]Note that $W(\delta_x, \delta_y) = |x-y|$

the case of the diffusion equation is extended to a porous medium type equation and the zero range process.

1.5. The Bakry–Émery approach to functional inequalities

A gradient flow structure consisted of the triple $(\mathcal{M}, E, g)$, and the fluctuation–dissipation theorem provided us with a stochastic gradient flow that ensured that the Gibbs measure $\mathrm{d}\mu(u) := \frac{1}{Z}e^{-E(u)}\,\mathrm{d}u$ is invariant. Evidently, there is a one-to-one correspondence between the energy E and the invariant measure μ, but as we have seen this formal definition can lead to problems, in the sense that some renormalization procedure is necessary such that μ even makes sense. Nonetheless, a *Markov triple* consists of the data (X, μ, Γ) where X is a state space, μ a (invariant) measure on X and Γ is a *carré du champ* operator, i.e. a bilinear operator on functions $X \to \mathbb{R}$, satisfying a certain condition (cf. [**9**, Definition 3.1.1]). This notion of carré du champ is very general and one should think of the choice

$$\Gamma(\zeta, \zeta) := g(\nabla\zeta, \nabla\zeta).$$

This data already gives rise to a bilinear form which we will refer to as a *Dirichlet form*[21](cf. [**84**])

$$\mathcal{E}(\zeta, \zeta) := \int_X \Gamma(\zeta, \zeta)\,\mathrm{d}\mu$$

and under very general conditions (cf. [**84**, Definition 3.1]) it is known that $\mathcal{E}$ gives rise to a stochastic process $\{M_t\}_{t\geq 0}$ on X (cf. [**84**, Theorem 3.5]). For example, this condition is met if there is a operator $\mathcal{L}$ such that

$$\mathcal{E}(\zeta, \zeta) = \int_X \zeta(-\mathcal{L}\zeta)\,\mathrm{d}\mu$$

and then $\mathcal{L}$ should be thought of as the generator of the process $\{M_t\}_{t\geq 0}$. At this point, it should be mentioned that the stochastic process $\{M_t\}_{t\geq 0}$ does not have to be unique.

For our purposes, we think of the data (X, μ, Γ) being supplemented with a semigroup $(P_t)_{t\geq 0}$, which then gives rise to a generator $\mathcal{L}$, such that μ is invariant for $(P_t)_{t\geq 0}$, and the carré du champ satisfies

$$\Gamma(\zeta, \zeta') = \frac{1}{2}(\mathcal{L}(\zeta\zeta') - \zeta\mathcal{L}\zeta' - \zeta'\mathcal{L}\zeta).$$

The insight of Bakry–Émery (cf. [**8**]) was that comparing the derivative of the semigroup and the semigroup of a derivative can unveil regularization properties of the semigroup which in turn gives rise to regularity properties of the invariant measure μ. To be more

[21] to be precise, a Dirichlet form is a bilinear form satisfying certain conditions [**84**, Definition 4.5]

precise, they considered the *gradient bound*[22] (cf. [**9**, Theorem 3.2.3])

$$\Gamma(P_t\zeta) \leq e^{-2Ct}P_t(\Gamma(\zeta)) \tag{1.38}$$

and the *strong gradient bound* (cf. [**9**, Theorem 3.2.4])

$$\sqrt{\Gamma(P_t\zeta)} \leq e^{-Ct}P_t\Big(\sqrt{\Gamma(\zeta)}\Big) \tag{1.39}$$

where $0 < C < \infty$. Clearly, the strong gradient bound implies the gradient bound by Jensen's inequality. These gradient bounds are key in the Bakry–Émery approach to functional inequalities since a strong gradient bound will imply a log-Sobolev inequality whereas a gradient bound will imply a spectral gap inequality. Note that (1.38) and (1.39) are in general not true and great effort has been put into determining sufficient conditions. Certainly the most famous sufficient condition is the Bakry–Émery curvature condition $CD(C,\infty)$ which holds if and only if

$$\Gamma_2(\zeta) \geq C\Gamma(\zeta) \tag{1.40}$$

for all ζ in a sufficiently large class of functions on X (cf. [**9**, (1.16.6)]). Here, Γ_2 is the iterated carré du champ defined by (cf. [**9**, (1.16.2)])

$$\Gamma_2(\zeta,\zeta') := \frac{1}{2}(\mathcal{L}\Gamma(\zeta,\zeta') - \Gamma(\zeta,\mathcal{L}\zeta') - \Gamma(\mathcal{L}\zeta,\zeta')).$$

The Γ_2-operator is closely related to the celebrated Bochner formula (cf. [**9**, Theorem C.3.3]) known from Riemannian geometry. Indeed, if X is a (finite) Riemannian manifold, $\Gamma(\zeta,\zeta) := |\nabla\zeta|^2$ with $\mathcal{L} := \Delta - \nabla E \cdot \nabla$, then

$$\Gamma_2(\zeta) = |\nabla\nabla\zeta|^2 + \mathrm{Ric}(\nabla\zeta,\nabla\zeta) + \nabla\zeta\cdot\nabla\nabla E\nabla\zeta$$

where $\nabla\nabla$ denotes the Hessian and Ric denotes the Ricci-curvature tensor on X [23] (cf. [**9**, 1.16.4]). In the same setting, if P_t is given by the heat kernel on X, the gradient bound (1.39) is equivalent to $\mathrm{Ric}(X) \geq C$ (cf. [**116**, Theorem 1.3]). In turn, this relationship gave rise to a *synthetic* definition of Ricci-curvature on non-smooth spaces (cf. [**109**], [**110**], [**82**]).

As already mentioned, a consequence of the strong gradient bound (1.39) is that the measure μ satisfies a log-Sobolev inequality whereas the gradient bound (1.38) implies a spectral gap inequality. We quickly recall the notion of these inequalities. A measure μ is said to satisfy a logarithmic Sobolev inequality if for all sufficiently nice functions ζ on X it holds that

$$\int_X \zeta^2\log\zeta^2\,\mathrm{d}\mu - \int_X \zeta^2\,\mathrm{d}\mu\log\Big(\int_X \zeta^2\,\mathrm{d}\mu\Big) \leq 2C\mathcal{E}(\zeta,\zeta)$$

[22] from now on we put $\Gamma(\zeta) := \Gamma(\zeta,\zeta)$

[23] all of the notions, e.g. $|\cdot|^2$ and ∇, are defined with respect to the Riemannian metric on X

whereas it satisfies a spectral gap inequality if

$$\int_X \zeta^2 \,\mathrm{d}\mu - \left(\int_x \zeta \,\mathrm{d}\mu\right)^2 \leq 2C\mathcal{E}(\zeta,\zeta). \tag{1.41}$$

If μ satisfies a log-Sobolev inequality then it also satisfies a spectral gap inequality; in fact, a spectral gap inequality can be regarded as a linearized log-Sobolev inequality (cf. [**9**, Proposition 5.1.3]). A class of measures that satisfy a log-Sobolev inequality are Gaussian measures (cf. [**9**, Proposition 5.5.1], [**36**, Proposition 4.1]). The importance of these two inequalities stems from the fact that they quantify ergodicity: the spectral gap inequality is equivalent to exponential decay in variance (cf. [**9**, Theorem 4.2.5]), while the log-Sobolev inequality is equivalent to exponential decay in entropy (cf. [**9**, Theorem 5.2.1]) which in turn by the Pinsker–Csizsár–Kullback inequality (cf. [**9**, 5.2.2]) implies exponential convergence in the total variation norm. Moreover, the log-Sobolev implies a hypercontractivity estimate for the associated semigroup (cf. [**9**, 5.2.3]). On the other hand, the appeal of these inequalities stem from the fact they are stable under tensorization (cf. [**9**, Proposition 4.3.1, Proposition 5.2.7]) which makes them useful in infinite dimensional settings. Recall that the standard Sobolev inequality on $\mathbb{R}^d$ (cf. [**39**, Theorem 1, p.263]) reads

$$\|u\|_{L^{p^*}(\mathbb{R}^d)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^d)}, \quad \text{where} \quad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{d} \tag{1.42}$$

where u is compactly supported. Clearly, the inequality (1.42) gets worse as $d \to \infty$ whereas the log-Sobolev inequality retains an improvement in terms of a logarithm. Similarly, the spectral gap inequality (1.41) is just an infinite-dimensional version of the famous Poincaré inequality (cf. [**39**, Theorem 1, p.275]). A sufficient criterion for a log-Sobolev inequality is the strong gradient bound (1.39) whereas the gradient bound (1.41) only implies the weaker spectral gap inequality (cf. the discussion in [**22**, p.237]). Note that if X is a Riemannian manifold as above, the curvature dimension condition (1.40) is satisfied if

$$\mathrm{Ric} + \nabla\nabla E \geq C\mathrm{Id}$$

and in this setting a log-Sobolev inequality holds for the measure $\frac{1}{Z}e^{-E}\,\mathrm{d}x$ where $\mathrm{d}x$ denotes the Riemannian volume measure (cf. [**9**, Proposition 5.7.1]). This suggests that Gibbs measures on infinite-dimensional state spaces, as discussed in Section 1.4, satisfy a log-Sobolev inequality provided they are convex. Since a renormalization procedure, which destroys the convexity, is often necessary in infinite dimensions, this is nevertheless unclear. Recently, progress has been made on generalizing the approach of Bakry and Émery in a certain sense (cf. [**11**]). This approach does not use the Markov semigroup, but the semigroup associated with the Polchinski equation and has been successful in proving a log-Sobolev inequality for the φ^4_d-measure for $d = 2, 3$ (cf. [**12**]). We refer to Chapter 4 for more details on gradient bounds for the φ^4_2-model.

1.6. Structure of the thesis and main contributions

The thesis consists of the following three articles: [**77**], [**48**] which is joint work with Benjamin Gess, Rishabh Gvalani and Felix Otto, and [**78**] which is joint work with Pavlos Tsatsoulis.

Chapter 2 is based on [**77**] and is concerned with a fully non-linear SPDE in divergence form with rough forcing. The forcing should be thought of as a colored Gaussian noise, although no probabilistic tools are used in the chapter. In [**97**] a solution theory for this equation has already been established, showing that the solution has a certain regularity. In general, this regularity is sharp as can be seen by considering Brownian motion. In this chapter we show a better regularity for the solution if one subtracts the solution to an anistropic stochastic heat equation which takes the non-linearity into account (cf. Theorem 2.2). This type of regularity is a sort of modelledness assumption à la [**55**] and is also at the core of [**98**].

Chapter 3 is based on [**48**] and is concerned with the derivation of a thin-film equation with noise or stochastic thin-film equation. By using the gradient flow structure, we can postulate an invariant measure and make an ansatz for the stochastic dynamics using the fluctuation–dissipation principle in the form of a variational Fokker–Planck equation. In order to circumvent the issue of solving this infinite-dimensional equation, we discretize in the space variable which yields by standard arguments a (high-dimensional) system of SDEs. The main observation is that, in general, an additional drift term appears which can be interpreted as the stochastic integral that ensures invariance of the (discrete) Gibbs measure. Discretizing the metric tensor in a specific way, as well as choosing specific coordinates, leads us to recover a well-known discretization of the mobility in the thin-film operator. This discretization was discovered in order to preserve the entropy estimate in the discrete setting which yields positivity of the deterministic scheme. Similarly, the main theorem (cf. Theorem 3.8) of the chapter is a stochastic entropy estimate which in turn also yields positivity in the case of no-slip boundary conditions, whereas the discretization which was used in the literature before does not preserve positivity. Furthermore, the gradient flow structure gives rise to a rate functional from the theory of large deviations. It turns out that finiteness of the rate functional has implications on the path being able to become zero at some point, as well as yielding Hölder regularity in time and space. Finally, numerical simulations suggest that both considered schemes do converge and they seem to converge to the same object.

Chapter 4 is based on [**78**] and is concerned with the φ^4_2-model on the torus. It is known that this equation has an invariant measure which is in Gibbs form and the energy is formally convex. Due to the necessary renormalization procedure this convexity is destroyed though and the standard Bakry–Émery machinery does not apply immediately. We analyze the linearized (with respect to the initial condition) equation and establish an energy estimate using a crucial stopping time argument. Then we postprocess this estimate which yields a sharper gradient bound for the Markov semigroup (cf. Theorem 4.1). As in the Bakry–Émery approach this yields a (local) spectral gap inequality (cf. Theorem 4.2) with a Dirichlet form that is (almost) optimal in terms of local regularity. Moreover, by ergodicity this implies the same spectral gap inequality for the φ^4_2-measure (cf. Corollary 4.3).

CHAPTER 2

A first order description of a nonlinear SPDE with colored noise

In this chapter we consider a nonlinear stochastic partial differential equation in divergence form. In contrast to the rest of the thesis, the forcing term is a Gaussian noise, that is white in time, but colored in space. This makes sure that the gradient of the solution is Hölder-continuous, but it will not be differentiable. This SPDE is a stochastic gradient flow if the non-linearity is a potential[1]. We prove a generalized Taylor expansion of the difference between the solution to the SPDE and the solution to its linearization around a fixed basepoint.

This chapter is based on the article [**77**].

2.1. Introduction and statement of main results

Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be given. We consider the following stochastic partial differential equation (SPDE)

$$\begin{cases} \partial_t u = \nabla \cdot A(\nabla u) + \xi \quad \text{on} \quad \mathbb{R} \times \mathbb{R}^d \\ u|_{t\leq 0} = 0. \end{cases} \tag{2.1}$$

The forcing term ξ is a space-time Gaussian noise, which is white in time, and periodic, colored and stationary in space. More precisely, we consider a Gaussian process ξ, formally defined via its covariance

$$\mathbb{E}\big[\xi(t,x)\xi(t',x')\big] = \delta(t-t')K(x-x'),$$

where $K : \mathbb{R}^d \to \mathbb{R}$ is periodic (cf. Section 2.3), and that for convenience is localized in the time interval $[0,1]$. The spatial covariance function K is chosen in such a way that the solution v to the linearized equation, i.e. the stochastic heat equation (SHE)

$$\begin{cases} \partial_t v = \Delta v + \xi \quad \text{on} \quad \mathbb{R} \times \mathbb{R}^d \\ v|_{t\leq 0} = 0, \end{cases} \tag{2.2}$$

[1]and if the noise is white in space as well

is not only a (continuous) function, but differentiable, and we have

$$[\nabla v]_\alpha < \infty \text{ a.s.} \tag{2.3}$$

for some $\alpha \in \left(\frac{1}{2}, 1\right)$[2]. For the precise notation we refer to Section 2.2. Since v is a linear functional of the Gaussian field ξ – and thus is itself Gaussian – more is true. Indeed, in [**97**, Lemma 3] Gaussian moments for v were established, meaning that there exists a $C_0 > 0$ such that

$$\mathbb{E}\left[\exp\left(\frac{1}{C_0}[\nabla v]_\alpha^2\right)\right] < \infty. \tag{2.4}$$

Moreover, we assume that the non-linearity A is elliptic (cf. (2.11)). Under these assumptions, it is expected that we should also have

$$[\nabla u]_\alpha < \infty \text{ a.s.} \tag{2.5}$$

and in fact [**97**, Corollary 1] yields the (a priori) estimate

$$[\nabla u]_\alpha \lesssim [\nabla v]_\alpha^{\frac{\alpha}{\alpha_0}} + [\nabla v]_\alpha \tag{2.6}$$

for an $\alpha_0 \in (0, 1)$ as in [**97**, Lemma 1][3]. Here, we note that the implicit constant in $\lesssim$ only depends and will depend on the dimension, α and the ellipticity of A (cf. Section 2.2).

In particular, we want to stress that for spatially colored noise, (2.1) is *not* a singular SPDE. In fact, the SPDE (2.1) with space-time white noise is, in general, not subcritical in any space-dimension, in the sense that by zooming in on small scales, the nonlinear terms blow up. For a reference concerning singular (semi-linear) SPDEs and the notion of subcriticality we refer to [**59**].

We should mention, that there are no probabilistic arguments in this chapter and if ξ is a distribution of suitable regularity such that (2.3) holds, the same result can be obtained. Nevertheless, we have chosen to consider the case where ξ is a (Gaussian) noise, on the one hand because we are heavily building on the work [**97**][4] and on the other hand because a typical example of a distribution with a negative Hölder-regularity is a realization of some Gaussian field with a short correlation length. Thus, everything in this chapter can be seen as a pathwise analysis, which is in the spirit of the theory of regularity structures (cf. [**59**]), where there is a clear distinction between the deterministic and probabilistic steps. Usually, the probabilistic arguments involve the construction of singular products along with the corresponding renormalization. In our setting no

[2]This is in contrast to the SHE with space-time white noise forcing, where already in dimension 2 the solution is not a function anymore
[3]coming from the DeGiorgi–Nash theorem
[4]where the analysis is also purely deterministic except for (2.4)

renormalization is necessary and hence it appears to be natural that the arguments are purely deterministic.

In [**97**, p.70, Theorem 1], it is shown that, under suitable assumptions on A – that for convenience we recall in Section 2.2 – a (unique) solution to (2.1) exists such that (2.6) and thus (2.5) hold. Due to the rough setting, the authors introduce a spatial increment operator δ_y (cf. (2.15)) with which they linearize equation (2.1)[5]. Then it is convenient (and a guiding principle) to subtract the increment $\delta_y v$ (where v solves (2.17)) in order to get rid of the noise ξ. Hence $\delta_y(u-v)$ satisfies a variable, but linear coefficient equation and the celebrated DeGiorgi–Nash theorem[6] yields an a priori estimate for its Hölder-norm for some $\alpha_0 \in (0,1)$. Postprocessing this estimate in turn establishes (2.5) for $\alpha = \alpha_0$. In the next step, this estimate is upgraded using standard $C^{1+\alpha}$-Schauder theory as well as the stochastic estimate (2.4), thus yielding (2.5) for any $\alpha \in (0,1)$. In general, (2.5) is sharp; in case of Brownian motion it is well-known that its paths have Hölder–regularity of at most $\frac{1}{2}$ but not better. We want to address the following question: Is there a way to give a finer description of the regularity of u?

In this chapter we intend to give such a regularity statement in the following way. It is often the case that the difference of the solution to a non-linear equation with a rough driving signal and the solution to the linearized equation is more regular. A prominent example that has been treated in the recent years is the φ^4_2-model (cf. [**25**]). Indeed, its solution φ and the solution f to the stochastic heat equation in dimension 2 are distributions but their difference $\varphi - f$ is smoother, in particular a function. This has been exploited in [**25**] in order to construct a solution which is local in time. In fact, this is also a guiding principle for the theory of (controlled) rough paths (cf. [**55**]) and ultimately in the theory of regularity structures (cf. [**59**]), where this concept has been vastly generalized.

The main result of this chapter is that ∇u is *modelled* after ∇v which essentially means that a certain *a priori*-estimate holds (cf. (2.10)). In the context of singular SPDEs such a modelling assumption is used in the following way. First, one constructs a singular product on the level of the linear, but irregular model, i.e. ∇v in this case. In a second step, one constructs the singular product on the level of the solution ∇u using the singular product on the level of the model and the assumption that ∇u is modelled after ∇v. Due to the nonlinear nature of the problem it is not sufficient to just consider the solution v to the SHE, but for all space-time points $z = (t', x')$ we consider the solution to an anisotropic SHE.

[5] as opposed to taking the derivative of the equation due to the low regularity of the noise
[6] a localized version thereof, to be precise

Let $a(t',x') := DA(\nabla u(t',x')) \in \mathbb{R}^{d\times d}$. Then we write $v_{a(t',x')}$ for the solution to the anisotropic stochastic heat equation, i.e. $v_{a(t',x')}$ solves

$$\begin{cases} \partial_t v_{a(t',x')} = \nabla \cdot a(t',x')\nabla v_{a(t',x')} + \xi \quad \text{on} \quad \mathbb{R}\times\mathbb{R}^d \\ v_{a(t',x')}|_{t\le 0} = 0 \end{cases} \tag{2.7}$$

with ξ as in (2.2). By the $C^{1+\alpha}$-Schauder theory developed in [**97**] we get the following uniform[7] estimate.

Lemma 2.1. *There exists a solution $v_{a(t',x')}$ to* (2.7) *where $a(t',x') = DA(\nabla u(t',x'))$ and A satisfies* (2.11)*,* (2.12) *as well as* (2.13)*, and we have*

$$\sup_{(t',x')\in\mathbb{R}\times\mathbb{R}^d} \Big[\nabla v_{a(t',x')}\Big]_\alpha \lesssim [\nabla v]_\alpha^{\frac{\alpha}{\alpha_0}} + [\nabla v]_\alpha.$$

Then we can state our main result.

Theorem 2.2. *Let u be a solution to the equation*

$$\begin{cases} \partial_t u = \nabla \cdot A(\nabla u) + \xi \quad \text{on} \quad \mathbb{R}\times\mathbb{R}^d \\ u|_{t\le 0} = 0 \end{cases}$$

where A satisfies (2.11)*,* (2.12) *and* (2.13)*, and where ξ is a Gaussian noise that is white in time, periodic, stationary and colored in space such that the corresponding solution v to the stochastic heat equation* (2.2) *satisfies $[\nabla v]_\alpha < \infty$ for some $\alpha \in (\frac{1}{2},1)$. Moreover, let $a(t',x') = DA(\nabla u(t',x')) \in \mathbb{R}^{d\times d}$ and let $v_{a(t',x')}$ be a solution to* (2.7)*. Then there exists a family of symmetric matrices $(B(t',x'))_{(t',x')}$ such that for all $x,x' \in \mathbb{R}^d$ and $t,t' \in \mathbb{R}$ it holds a.s. that*

$$\Big|\nabla u(t,x) - \nabla u(t',x') - \Big(\nabla v_{a(t',x')}(t,x) - \nabla v_{a(t',x')}(t',x')\Big) - B(t',x')(x-x')\Big| \lesssim d^{2\alpha}((t,x),(t',x')) \tag{2.8}$$

and $\lesssim$ denotes $\le C$ where C depends only on $[\nabla v]_\alpha$, λ as well as Λ, α and d.

One way to think of (4.27) is as a generalized Taylor expansion. Indeed, rewriting (4.27) slightly as

$$\Big|(\nabla u(t,x) - \nabla v_{a(t',x')}(t,x)) - (\nabla u(t',x') - \nabla v_{a(t',x')}(t',x')) - B(t',x')(x-x')\Big| \lesssim d^{2\alpha}((t,x),(t',x')). \tag{2.9}$$

the symmetric matrix $B(t',x')$ plays the role of the Hessian of the function $u(t',\cdot) - v_{a(t',x')}(t',\cdot)$ at the basepoint (t',x'). Since $2\alpha > 1$ it is natural, that an affine correction appears. Moreover, since $2\alpha > 1$, the matrix B can be seen to be unique and

[7]in the basepoint

does not depend on (t,x), or more precisely it does not depend on $d((t,x),(t',x'))$. Note that (4.28) is essentially the modelledness condition[8] of [**98**, Definition 3.1, p.880], which is an extension of controlled rough paths (cf. [**55**]). In [**98**], they construct singular products on the level of the stochastic heat equation using probabilistic arguments. They then use the modelledness condition [**98**, Definition 3.1, p.880] to lift these singular products to the nonlinear setting. Since (2.1) is not singular, the regularity theory in [**97**] does not rely on such a modelledness. Nevertheless, Theorem 2.2 states that such a modelledness holds true. Of course, (4.27) also implies that $\nabla u(t',\cdot) - \nabla v_{a(t',x')}(t',\cdot)$ is differentiable in x' as well as the improved Hölder-regularity in time $\left|\nabla u(t,x') - \nabla v_{a(t',x')}(t,x') - \left(\nabla u(t',x') - \nabla v_{a(t',x')}(t',x')\right)\right| \lesssim |t-t'|^\alpha$ at $z = (t',x')$.

From now on, we set $w_{a(t',x')} := u - v_{a(t',x')}$. Then we define the *modelling constant*

$$M := \sup_{z=(t',x')} \inf_B \sup_{r>0} r^{-2\alpha} \left\| \nabla w_{a(t',x')} - B_{x'} \right\|_{P_r(z)},$$

where the infimum ranges over all affine functions $B_{x'}(x) := B(x-x')+b$ with $B \in \mathbb{R}^{d\times d}$ being symmetric and $b \in \mathbb{R}^d$, and $P_r(z)$ is the parabolic cylinder of radius r defined in Section 2.2. In order to prove Theorem 2.2 we show that we have

$$M \le C(d,\lambda,\Lambda,\alpha,[\nabla v]_\alpha). \tag{2.10}$$

Then it is straightforward to see that the optimal choice is $b = \nabla w_{a(t',x')}(t',x')$ and hence (2.10) yields Theorem 2.2. Note that the definition of M is essentially the same as the definition of the modelling constant in [**98**, Definition 3.1]. Moreover, we want to make the following connection. Apart from the dependence of $w_{a(t',x')}$ on the basepoint $z = (t',x')$, the semi-norm M bears close resemblance to the Hölder-norm defined in [**76**, Section 3.3] where Hölder-regularity is measured in terms of how well a function is approximated by polynomials.

2.2. Notation and assumptions

Our notation and assumptions are the same as in [**97**, Section 2]. For the convenience of the reader we will recall them here. The non-linearity $A : \mathbb{R}^d \to \mathbb{R}^d$ is assumed to be continuously differentiable and uniformly elliptic in the sense that there exists $\lambda > 0$ such that

$$\eta \cdot DA(x)\eta \ge \lambda|\eta|^2 \text{ for all } x,\eta \in \mathbb{R}^d \tag{2.11}$$

and we have

$$|DA(x)\eta| \le |\eta| \text{ for all } x,\eta \in \mathbb{R}^d \tag{2.12}$$

[8]on the level of the solution and not the gradient for $\sigma \equiv 1$

where by DA we denote the Jacobian of A. Moreover, we assume that there exists $\Lambda > 0$ such that

$$|DA(x) - DA(y)| \leq \Lambda |x-y| \text{ for all } x, y \in \mathbb{R}^d. \tag{2.13}$$

For $\alpha \in (0,1)$ the semi-norm $[\cdot]_\alpha$ is defined as

$$[f]_\alpha := \sup_{z \neq z' \in \mathbb{R}\times\mathbb{R}^d} \frac{|f(z) - f(z')|}{d^\alpha(z,z')} < \infty \tag{2.14}$$

for all space-time functions $f : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}$ or vector fields $f : \mathbb{R}\times\mathbb{R}^d \to \mathbb{R}^d$ where

$$d((t,x),(t',x')) := |t-t'|^{\frac{1}{2}} + |x-x'|$$

denotes the *Carnot–Caratheodory* metric and by abuse of notation $|\cdot|$ refers either to the absolute value or the Euclidean norm on $\mathbb{R}^d$. Naturally, the space C^α denotes all functions f such that $[f]_\alpha < \infty$. For $r > 0$ and $z = (t', x')$, by

$$P_r(z) := (t' - r^2, t') \times B_r(x')$$

we denote the parabolic cylinder centered around z with radius r. Then $\|\cdot\|_{P_r(z)}$ is the supremum norm on $P_r(z)$. We will also frequently write $[f]_{\alpha, P_r(z)}$ which is the same semi-norm as in (2.14) restricted to $P_r(z)$.

For $y \in \mathbb{R}^d$ we define the spatial increment operator δ_y as

$$\delta_y f(t,x) := f(t, x+y) - f(t,x) \tag{2.15}$$

where f is either a scalar or a vector field. Then, by the mean value theorem, we can linearize our non-linearity according to

$$\delta_y A(\nabla u) = a_y \nabla \delta_y u$$

where

$$a_y(t,x) = \int_0^1 DA(\theta \nabla u(t, x+y) + (1-\theta)\nabla u(t,x))\, \mathrm{d}\theta \in \mathbb{R}^{d\times d}.$$

The assumptions on the Jacobian DA translate to estimates on a_y as follows. We have

$$\eta \cdot a_y(t,x)\eta \geq \lambda |\eta|^2 \text{ for all } x, \eta \in \mathbb{R}^d, t \in \mathbb{R}$$

as well as

$$|a_y(t,x)\eta| \leq |\eta| \text{ for all } x, \eta \in \mathbb{R}^d, t \in \mathbb{R}$$

and

$$[a_y]_\alpha \leq [\nabla u]_\alpha. \tag{2.16}$$

Let ψ be a smooth, positive and radially symmetric mollifier that satisfies $\operatorname{supp}\psi \subset B_1(0)$ as well as $\int_{\mathbb{R}^d} \psi \, \mathrm{d}x = 1$. The radial symmetry assumption also ensures that first moments vanish, i.e. for all $i = 1, \ldots, d$ it holds $\int_{\mathbb{R}^d} \psi(x) x_i \, \mathrm{d}x = 0$. Then we write $\psi_r(x) := \frac{1}{r^d}\psi(\frac{x}{r})$ and we define for any function f

$$f_r := f_{r,0} := f * \psi_r$$

as well as for $i = 1, \ldots, d$

$$f_{r,i} := f * (\partial_i \psi)_r.$$

For vector fields[9] this notation is to be understood entrywise. Moreover, we will always write $B_{x'}$ for the affine function

$$B_{x'}(x) := B(x - x') + b$$

where $B \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$.

2.3. The stochastic heat equation

In the work [**97**], they consider a stationary, spatially periodic Gaussian noise ξ on $\mathbb{R}\times\mathbb{R}^d$, localized in the time interval $[0,1]$, that is white in time and colored in space, and such that the corresponding stochastic heat equation

$$\begin{cases} \partial_t v = \Delta v + \xi \quad \text{on} \quad \mathbb{R}\times\mathbb{R}^d \\ v|_{t\le 0} = 0 \end{cases} \tag{2.17}$$

satisfies

$$[\nabla v]_\alpha < \infty \text{ a.s.} \tag{2.18}$$

for some $\alpha \in (\frac{1}{2}, 1)$. More specifically, they consider a Gaussian field $\xi = \{\xi(f)\}_f$ with $1-$periodic, positive definite covariance function $K : \mathbb{R}^d \to \mathbb{R}$ such that for smooth test-functions f, g

$$\mathbb{E}[\xi(f)\xi(g)] = \int_0^1 \int_{[0,1]^d} \int_{[0,1]^d} f(t,x) K(x-y) g(t,y) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}t$$

and $\xi(f)$ is normally distributed with mean zero. Such a Gaussian field is easily seen to exist by Kolmogorov's consistency theorem. The fact that K only depends on one variable yields stationarity of ξ. Periodicity of K translates to periodicity of ξ. Positive definiteness of K implies that its Fourier transform $\hat{K}$ is real-valued and non-negative and requiring that K is symmetric, i.e. $K(-x) = K(x)$, yields $\hat{K}(-k) = \hat{K}(k)$. Most

[9] such as a gradient

importantly, requiring that the Fourier transform of K satisfies

$$\hat{K}(k) \lesssim (1+|k|^2)^{-\frac{s}{2}}$$

for all $k \in (2\pi\mathbb{Z})^d$, where $s = 2\alpha + d \in (d, d+2)$ finally implies (2.18) (cf. [**97**, Lemma 3]).

Now we give the proof for Lemma 2.1.

Proof of Lemma 2.1. For any $(t', x') \in \mathbb{R} \times \mathbb{R}^d$ we apply Theorem 1 of [**97**] to A given by

$$A(\nabla u) = a(t', x')\nabla u$$

where $u = v_{a(t',x')}$[10]. Thus we get a unique solution $v_{a(t',x')}$ that satisfies $\left[\nabla v_{a(t',x')}\right]_\alpha < \infty$ a.s.. Then we can apply Corollary 1 of [**97**] which yields uniformly in (t', x') the estimate

$$\left[\nabla v_{a(t',x')}\right]_\alpha \leq C(d, \lambda, \Lambda, \alpha)([\nabla v]_\alpha + [\nabla v]_\alpha^{\frac{\alpha}{\alpha_0}})$$

and hence the conclusion. □

2.4. Deterministic estimates

From now on we fix a space-time point $z = (t', x')$. We set $w_{a(t',x')} := u - v_{a(t',x')}$, where we recall that $a(t', x') = DA(\nabla u(t', x'))$. For notational convenience we will drop the subscripts referring to $a(t', x')$.

In the first step, for simplicity, we focus on the spatial part of the modelling constant. The proof is elementary and essentially an extension of the proof of Lemma 1 in [**97**].

Proposition 2.3. *Let $x' \in \mathbb{R}^d$. For $f : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla f \in C^{\alpha}$*[11]*, we have*

$$\sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla f - B_{x'}\|_{B_r(x')} \lesssim \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\leq l} \inf_{k\in\mathbb{R}^d} \|\nabla \delta_y f - k\|_{B_l(x')} =: N$$

where $B_{x'}(x) = B(x - x') + b$, $B \in \mathbb{R}^{d\times d}$ symmetric, $b \in \mathbb{R}^d$.

Proof. First of all, we assume that f is smooth. We denote by $\{e_i\}_{i=1,\dots,d}$ the standard orthonormal basis of $\mathbb{R}^d$. Let $k = k(y, l)$ be the (near) optimal constant for N. Fix $i, j = 1, \dots, d$ and $l > 0$. Note that for all y_1, y_2 we have the identity $\delta_{y_1+y_2}\partial_i f = \delta_{y_2}\partial_i f(\cdot + y_1) + \delta_{y_1}\partial_i f$.
Then we estimate

$$|k_i(2le_j, 2l) - 2k_i(le_j, 2l)| \leq \left\|k_i(2le_j, 2l) - \delta_{2le_j}\partial_i f\right\|_{B_{2l}(x')} + \left\|k_i(le_j, 2l) - \delta_{le_j}\partial_i f\right\|_{B_{2l}(x')}$$

[10] In this case A is non-deterministic, but since the analysis is pathwise and all the constants depend on A only through the ellipticity constants λ, Λ, which are deterministic, the proofs are unaffected.
[11] only in space in this case

$$+ \left\| k_i(le_j, 2l) - \delta_{le_j}\partial_i f \right\|_{B_l(x')}$$
$$\le 3(2l)^{2\alpha} N$$

and similarly

$$\begin{aligned} |k_i(le_j, 2l) - k_i(le_j, l)| &\le \left\| k_i(le_j, 2l) - \delta_{le_j}\partial_i f \right\|_{B_{2l}(x')} + \left\| k_i(le_j, l) - \delta_{le_j}\partial_i f \right\|_{B_l(x')} \\ &\le 2(2l)^{2\alpha} N. \end{aligned}$$

Combining these two estimates yields via the triangle inequality

$$|k_i(2le_j, 2l) - 2k_i(le_j, l)| \lesssim Nl^{2\alpha}.$$

Hence for any $l > 0$ and $n \in \mathbb{N}$ we get

$$\left| \frac{k_i(\frac{l}{2^n}e_j, \frac{l}{2^n})}{\frac{l}{2^n}} - \frac{k_i(\frac{l}{2^{n+1}}e_j, \frac{l}{2^{n+1}})}{\frac{l}{2^{n+1}}} \right| \lesssim Nl^{2\alpha-1}(2^{-n})^{2\alpha-1}.$$

Using our assumption that $2\alpha - 1 > 0$ we see that the corresponding sequence is Cauchy and thus there exists $a_{ij}(l) \in \mathbb{R}$ such that

$$\frac{k_i(\frac{l}{2^n}e_j, \frac{l}{2^n})}{\frac{l}{2^n}} \to a_{ij}(l) \text{ for } n \to \infty$$

and, by dyadic summation[12], this yields

$$\left| \frac{k_i(le_j, l)}{l} - a_{ij}(l) \right| \lesssim Nl^{2\alpha-1}. \tag{2.19}$$

Note that $a_{ij}(l)$ is constant on dyadics, i.e.

$$a_{ij}(l) = a_{ij}(2^{-m}l) \tag{2.20}$$

for all $m \in \mathbb{N}$. Feeding the estimate (2.19) into N we consequently have

$$\left\| \frac{1}{l}\delta_{le_j}\partial_i f - a_{ij}(l) \right\|_{B_l(x')} \lesssim Nl^{2\alpha-1}$$

and, since f is smooth, we infer that

$$a_{ij}(l) \to \partial_{ij} f(x') \text{ for } l \to 0.$$

Hence we conclude by (2.20) that a_{ij} is constant and we have $a_{ij} = \partial_{ij} f(x')$. Moreover, the estimate

$$\left\| \frac{1}{l}\delta_{le_j}\partial_i f - \partial_{ij} f(x') \right\|_{B_l(x')} \lesssim Nl^{2\alpha-1} \tag{2.21}$$

[12] again using the fact that $2\alpha - 1 > 0$

holds. Now let $x \in B_r(x')$ and we set $y := x - x'$. Then we estimate componentwise using (2.21)

$$
\begin{aligned}
(2.22)\qquad & |\partial_i f(x) - \partial_i f(x') - \nabla \partial_i f(x') \cdot (x - x')| \\
&\leq \sum_{k=1}^{d} |\delta_{y_k e_k} \partial_i f(x'_1, \dots, x'_{k+1} + y_{k+1}, \dots, x'_d + y_d) - \partial_{ki} f(x') y_k| \\
&\lesssim N \sum_{k=1}^{d} |y_k|^{2\alpha} \lesssim N r^{2\alpha}.
\end{aligned}
$$

Now we drop the assumption that f is smooth. To this end, let $\varepsilon > 0$. Let k be the (near) optimal constant for f in N. Then we have for all $x \in B_l(x')$ and $|y| \leq l$ by the triangle inequality

$$|\delta_y f_\varepsilon(x) - k| \leq \|\delta_y f - k\|_{B_{2l}(x')}$$

and hence

$$
\begin{aligned}
(2.23)\qquad \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y| \leq l} \inf_{k^\varepsilon} \|\delta_y f_\varepsilon - k^\varepsilon\|_{B_l(x')} &\leq \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y| \leq l} \|\delta_y f_\varepsilon - k\|_{B_l(x')} \\
&\leq \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y| \leq l} \|\delta_y f - k\|_{B_l(x')}.
\end{aligned}
$$

Note that, since $\partial_i f$ is Hölder-continuous, $\partial_i f_\varepsilon$ converges uniformly to $\partial_i f$ and thus, for fixed r, it holds that $\inf_{B^\varepsilon} \|\nabla f_\varepsilon - B^\varepsilon_{x'}\|_{B_r(x')}$ converges to $\inf_B \|\nabla f - B_{x'}\|_{B_r(x')}$. Dividing by $r^{2\alpha}$ and taking the supremum over r yields lower semicontinuity of the seminorm

$$
(2.24)\qquad \sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla f - B_{x'}\|_{B_r(x')} \leq \liminf_{\varepsilon \to 0} \sup_{r>0} \frac{1}{r^{2\alpha}} \inf_{B^\varepsilon} \|\nabla f_\varepsilon - B^\varepsilon_{x'}\|_{B_r(x')}.
$$

We conclude

$$
\begin{aligned}
\sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla f - B_{x'}\|_{B_r(x')} &\overset{(2.24)}{\leq} \liminf_{\varepsilon \to 0} \sup_{r>0} \frac{1}{r^{2\alpha}} \inf_{B^\varepsilon} \|\nabla f_\varepsilon - B^\varepsilon_{x'}\|_{B_r(x')} \\
&\overset{(2.22)}{\lesssim} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y| \leq l} \inf_{k^\varepsilon} \|\delta_y f_\varepsilon - k^\varepsilon\|_{B_l(x')} \\
&\overset{(2.23)}{\lesssim} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y| \leq l} \inf_{k} \|\delta_y f - k\|_{B_l(x')}.
\end{aligned}
$$

□

In order to include time we extend Proposition 2.3 to space-time functions via the following interpolation inequality.

Corollary 2.4. *For $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ such that $\nabla f \in C^\alpha$ we have*

$$\sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla f - B_{x'}\|_{P_r(z)} \lesssim \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\le l} \inf_{k\in\mathbb{R}^d} \|\nabla \delta_y f - k\|_{P_l(z)} + \sum_{i=0}^{d} \sup_{r>0} r^{1-2\alpha} \sup_{|y|\le r} \left\| \partial_t (\delta_y f)_{r,i} \right\|_{P_r(z)}$$

Proof. Let $(t,x) \in P_r(z)$. By Proposition 2.3 there exist a symmetric matrix $B(t',x')$ and a vector $b(t',x')$ such that

$$\begin{aligned} \left|\nabla f(t',x) - \left(B(t',x')(x-x') + b(t',x')\right)\right| &\lesssim r^{2\alpha} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\le l} \inf_{k=k(t)} \|\nabla \delta_y f(t',\cdot) - k\|_{B_l(x')} \\ &\lesssim r^{2\alpha} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\le l} \inf_{k\in\mathbb{R}^d} \|\nabla \delta_y f - k\|_{P_l(z)}. \end{aligned}$$

Via the triangle inequality we split the remainder into a spatial increment and a temporal increment

$$\begin{aligned} \left|\nabla f(t,x) - \nabla f(t',x)\right| \le& |\nabla f(t,x) - (\nabla f)_r(t,x)| + \left|(\nabla f)_r(t,x) - (\nabla f)_r(t',x)\right| \\ &+ \left|\nabla f(t',x) - (\nabla f)_r(t',x)\right|. \end{aligned}$$

Again by Proposition 2.3 there exist a symmetric matrix $B(t,x')$ and a vector $b(t,x')$ such that

$$\left|\nabla f(t,x) - \left(B(t,x')(x-x') + b(t,x')\right)\right| \lesssim r^{2\alpha} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\le l} \inf_{k\in\mathbb{R}^d} \|\nabla \delta_y f - k\|_{P_l(z)}.$$

and, in order to estimate the spatial increment, we appeal to radial symmetry of our mollifier, which implies

$$B(t,x')(x-x') = \int_{\mathbb{R}^d} \psi_r(x-\zeta) B(t,x')(\zeta - x')\, \mathrm{d}\zeta$$

and, hence again by Proposition 2.3 and the triangle inequality

$$\begin{aligned} &|\nabla f(t,x) - (\nabla f)_r(t,x)| \\ &= \Big|\nabla f(t,x) - b(t,x') - B(t,x')(x-x') \\ &- \int_{\mathbb{R}^d} \psi_r(x-\zeta)\left(\nabla f(t,\zeta) - b(t,x') - B(t,x')(\zeta-x')\right) \mathrm{d}\zeta\Big| \\ &\lesssim r^{2\alpha} \sup_{l>0} \frac{1}{l^{2\alpha}} \sup_{|y|\le l} \inf_{k\in\mathbb{R}^d} \|\nabla \delta_y f - k\|_{P_l(z)}. \end{aligned}$$

In order to treat the time difference, we show

$$\|\partial_t (\nabla f)_r\|_{P_r(z)} \lesssim \sum_{i=0}^{d} \frac{1}{r} \sup_{|y|\le r} \left\| \partial_t (\delta_y f)_{r,i} \right\|_{P_r(z)}. \tag{2.25}$$

Let $i = 1, \dots, d$. For any $\hat{t} \in \mathbb{R}$ we compute, appealing to the mean value theorem in space and then Fubini's theorem

$$\begin{aligned}(\partial_i f)_r(\hat{t}, x) - \Big(\frac{1}{r}\delta_{re_i} f\Big)_r(\hat{t}, x) &= \int_{\mathbb{R}^d} \psi_r(x-\zeta)\Big(\partial_i f(\hat{t}, \zeta) - \int_0^1 \partial_i f(\hat{t}, \zeta + \theta r e_i)\,\mathrm{d}\theta\Big)\,\mathrm{d}\zeta \\ &= -\int_0^1 \int_{\mathbb{R}^d} \psi_r(x-\zeta)\delta_{\theta r e_i}\partial_i f(\hat{t}, \zeta)\,\mathrm{d}\zeta\,\mathrm{d}\theta.\end{aligned}$$

Integration by parts then yields

$$(\partial_i f)_r(\hat{t}, x) - \Big(\frac{1}{r}\delta_{re_i} f\Big)_r(\hat{t}, x) = -\int_0^1 \Big(\frac{1}{r}\delta_{\theta r e_i} f\Big)_{r,i}(\hat{t}, x)\,\mathrm{d}\theta.$$

Taking the time derivative proves (2.25).

By the mean value theorem in time, for any $x \in \mathbb{R}^d$ we have

$$(\partial_i f)_r(t, x) - (\partial_i f)_r(t', x) = (t - t')\int_0^1 \partial_t(\partial_i f)_r(\delta t + (1-\delta)t', x)\,\mathrm{d}\delta. \tag{2.26}$$

and combining (2.26) with (2.25) finally yields

$$\big|(\nabla f)_r(t, x) - (\nabla f)_r(t', x)\big| \lesssim r^2\|\partial_t(\nabla f)_r\|_{P_r(z)} \lesssim \sum_{i=0}^d r \sup_{|y|\le r}\Big\|\partial_t(\delta_y f)_{r,i}\Big\|_{P_r(z)}.$$

Hence we have proven that for $(t, x) \in P_r(z)$

$$\begin{aligned}\big|\nabla f(t,x) - \big(B(t',x')\cdot(x-x') + b(t',x')\big)\big| \lesssim & r^{2\alpha}\sup_{l>0}\frac{1}{l^{2\alpha}}\sup_{|y|\le l}\inf_{k\in\mathbb{R}^d}\|\nabla\delta_y f - k\|_{P_l(z)} \\ &+ r^{2\alpha}\sum_{i=0}^d \sup_{l>0} l^{1-2\alpha}\sup_{|y|\le l}\Big\|\partial_t(\delta_y f)_{l,i}\Big\|_{P_l(z)}\end{aligned}$$

□

Let $y \in \mathbb{R}^d$. Since $\delta_y u$ satisfies

$$\partial_t\delta_y u - \nabla\cdot a_y\nabla\delta_y u = \partial_t\delta_y v - \nabla\cdot a(t',x')\nabla\delta_y v$$

we see that $\delta_y w$ satisfies the equation

$$\partial_t\delta_y w - \nabla\cdot a_y\nabla\delta_y w = \nabla\cdot(a_y - a(t',x'))\nabla\delta_y v.$$

Then, we can write this as a constant coefficient equation

$$\partial_t\delta_y w - \nabla\cdot a(t',x')\nabla\delta_y w = \nabla\cdot g \tag{2.27}$$

by introducing

$$g := (a_y - a(t',x'))\nabla\delta_y u.$$

Note that (2.27) is now invariant under affine spatial translations, i.e. of functions of the form $\mathrm{aff}(x) := b\cdot x + a$ for some $b \in \mathbb{R}^d, a \in \mathbb{R}$. Hence, we can apply the $C^{1+\alpha}$

interior Schauder estimate [**79**, Theorem 4.8], using this invariance, as well as parabolic rescaling, to obtain that, for all $l > 0$ and all space-time points $z = (t', x')$, it holds

$$l^\alpha \inf_{\text{aff}} [\nabla(\delta_y w - \text{aff})]_{\alpha, P_l(z)} + \inf_{\text{aff}} \|\nabla(\delta_y w - \text{aff})\|_{P_l(z)} \lesssim l^\alpha [g]_{\alpha, P_{2l}(z)} + l^{-1} \inf_{\text{aff}} \|\delta_y w - \text{aff}\|_{P_{2l}(z)}$$

which is equivalent to

$$l^\alpha [\nabla \delta_y w]_{\alpha, P_l(z)} + \inf_{k \in \mathbb{R}^d} \|\nabla \delta_y w - k\|_{P_l(z)} \lesssim l^\alpha [g]_{\alpha, P_{2l}(z)} + l^{-1} \inf_{\text{aff}} \|\delta_y w - \text{aff}\|_{P_{2l}(z)}. \tag{2.28}$$

This suggests that we need to further estimate the right hand side which is captured in the next Proposition and which is an extension of [**97**, p.73, (34)]. The proof is similar.

Proposition 2.5. *Let $z = (t', x')$. For $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ continuously differentiable and $y \in \mathbb{R}^d$ we have the estimate*

$$\inf_{\text{aff}} \|\delta_y f - \text{aff}\|_{P_l(z)} \lesssim |y| \inf_B \|\nabla f - B_{x'}\|_{P_{2l}(z)}$$

where $B_{x'}(x) = B(x - x') + b$, $B \in \mathbb{R}^{d \times d}$ is symmetric and $b \in \mathbb{R}^d$.

Proof. For $B \in \mathbb{R}^{d \times d}$ symmetric we define

$$\widetilde{f}(t, x) := f(t, x) - \Big(\frac{1}{2}(x - x') \cdot B(x - x') + b \cdot (x - x')\Big).$$

Then we compute

$$\nabla \widetilde{f}(t, x) = \nabla f(t, x) - (B(x - x') + b)$$

as well as

$$\delta_y \widetilde{f}(t, x) = \delta_y f(t, x) - \Big(y \cdot B(x - x') + \frac{1}{2} y \cdot By + b \cdot y\Big).$$

Notice that for any y the map $x \mapsto y^T B(x - x') + \frac{1}{2} y^T By + b \cdot y$ is again affine. Thus, by the mean value theorem, we have $\big\|\delta_y \widetilde{f}\big\|_{P_l(z)} \le |y| \big\|\nabla \widetilde{f}\big\|_{P_{2l}(z)}$ which yields the statement. □

2.4.1. Proof of the main theorem.

We are now able to prove our main theorem.

Proof of Theorem 2.2. In the following B always denotes a symmetric matrix. We write $M = \sup_{z=(t,x)} M_z$ where $M_z = \inf_B \sup_{r>0} r^{-2\alpha} \big\|\nabla w_{a(t',x')} - B_{x'}\big\|_{P_r(z)}$. For now we assume that M_z is finite for any z. By Corollary 2.4, we have

$$\sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla w - B_{x'}\|_{P_r(z)} \lesssim \sup_{r>0} \frac{1}{r^{2\alpha}} \sup_{|y| \le r} \inf_k \|\nabla \delta_y w - k\|_{P_r(z)} + \sum_{i=0}^{d} \sup_{r>0} r^{1-2\alpha} \sup_{|y| \le r} \big\|\partial_t (\delta_y w)_{r,i}\big\|_{P_r(z)}. \tag{2.29}$$

Let $|y| \leq r \leq l$. First, we focus on the second term on the right hand side of (2.29). To this end, we recall (cf. (2.27)) that $\delta_y w$ satisfies the constant coefficient equation

$$\partial_t \delta_y w = \nabla \cdot a(t',x')\nabla \delta_y w + \nabla \cdot g$$

where

$$g = (a_y - a(t',x'))\nabla \delta_y u$$

and we estimate for any $i = 0, \ldots, d$

$$\begin{aligned} \|\partial_t (\delta_y w)_{r,i}\|_{P_r(z)} &\lesssim r^{\alpha-1} [a(t',x')\nabla \delta_y w + g]_{\alpha, P_{2r}(z)} \\ &\lesssim r^{\alpha-1} \Big([\nabla \delta_y w]_{\alpha, P_{2r}(z)} + [g]_{P_{2r}(z)} \Big). \end{aligned} \tag{2.30}$$

Hence we have

$$\begin{aligned} \sup_{r>0} \frac{1}{r^{2\alpha}} \inf_B \|\nabla w - B_{x'}\|_{P_r(z)} \lesssim & \sup_{r>0} \frac{1}{r^{2\alpha}} \sup_{|y|\leq r} \inf_k \|\nabla \delta_y w - k\|_{P_r(z)} \\ & + \sup_{r>0} r^{-\alpha} \sup_{|y|\leq r} \Big([\nabla \delta_y w]_{\alpha, P_{2r}(z)} + [g]_{P_{2r}(z)} \Big). \end{aligned}$$

For $|y| \leq r \leq l$ we estimate by (2.28)

$$\inf_k \|\nabla \delta_y w - k\|_{P_r(z)} \leq r^\alpha [\nabla \delta_y w]_{\alpha, P_l(z)} \lesssim r^\alpha [g]_{\alpha, P_{2l}(z)} + \frac{r^\alpha}{l^{1+\alpha}} \inf_{\text{aff}} \|\delta_y w - \text{aff}\|_{P_{2l}(z)} \tag{2.31}$$

as well as

$$[\nabla \delta_y w]_{\alpha, P_l(z)} \lesssim [g]_{\alpha, P_{2l}(z)} + \frac{1}{l^{1+\alpha}} \inf_{\text{aff}} \|\delta_y w - \text{aff}\|_{P_{2l}(z)}.$$

Now we turn to the estimate of $[g]_{\alpha, P_{2l}(z)}$. We appeal to the Lipschitz continuity of DA (cf. (2.13)) and recall the definition of $a(t',x')$ as well as a_y in order to estimate

$$\begin{aligned} |a_y(t,x) - a(t',x')| &\lesssim \int_0^1 |\theta \nabla u(t,x+y) + (1-\theta)\nabla u(t,x) - \nabla u(t',x')| \, \mathrm{d}\theta \\ &\lesssim \int_0^1 \theta |\nabla u(t,x+y) - \nabla u(t,x)| + |\nabla u(t,x) - \nabla u(t',x')| \, \mathrm{d}\theta \\ &\lesssim |y|^\alpha + d((t,x),(t',x'))^\alpha \end{aligned}$$

and thus, for $|y| \leq r \leq l$ we get

$$\|a_y - a(t',x')\|_{P_{2l}(z)} \lesssim l^\alpha.$$

By (2.16), we have

$$[a_y - a(t',x')]_{\alpha, P_{2l}(z)} = [a_y]_{\alpha, P_{2l}(z)} \lesssim [\nabla u]_{\alpha, P_{3l}(z)}.$$

Moreover, it holds that

$$\|\nabla \delta_y u\|_{P_{2l}(z)} \lesssim l^\alpha [\nabla u]_{\alpha, P_{3l}(z)}$$

as well as

$$[\nabla\delta_y u]_{\alpha,P_{2l}(z)} \lesssim [\nabla u]_{\alpha,P_{3l}(z)}.$$

Summing up, we arrive at

$$\begin{aligned}[g]_{\alpha,P_{2l}(z)} = [(a_y - a(t',x'))\nabla\delta_y u]_{\alpha,P_{2l}(z)} &\le [a_y - a(t',x')]_{\alpha,P_{3l}(z)}\|\nabla\delta_y u\|_{P_{3l}(z)} \\ &\quad + \|a_y - a(t',x')\|_{P_{3l}(z)}[\nabla\delta_y u]_{\alpha,P_{3l}(z)} \\ &\lesssim l^\alpha[\nabla u]_{\alpha,P_{3l}(z)}\end{aligned} \tag{2.32}$$

for $|y| \le r \le l$. Combining (2.32) with (2.31) yields for $|y| \le r \le l$

$$r^\alpha[\nabla\delta_y w]_{\alpha,P_l(z)} \lesssim l^{2\alpha}[\nabla u]_{\alpha,P_{3l}(z)} + \frac{r^\alpha}{l^{1+\alpha}}\inf_{\text{aff}}\|\delta_y w - \text{aff}\|_{P_{2l}(z)}. \tag{2.33}$$

Then using Proposition 2.5 we further estimate (2.33) to the effect that

$$r^\alpha[\nabla\delta_y w]_{\alpha,P_l(z)} \lesssim l^{2\alpha}[\nabla u]_{\alpha,P_{3l}(z)} + \left(\frac{r}{l}\right)^{1+\alpha}\inf_B\|\nabla w - B_{x'}\|_{P_{4l}(z)}. \tag{2.34}$$

Summing up, we get by (2.30) and (2.34)

$$\begin{aligned}&\frac{1}{r^{2\alpha}}\sup_{|y|\le r}\inf_k\|\nabla\delta_y w - k\|_{P_r(z)} + \sum_{i=0}^d r^{1-2\alpha}\sup_{|y|\le r}\|\partial_t(\delta_y w)_r\|_{P_r(z)} \\ &\lesssim \left(\frac{l}{r}\right)^{2\alpha}[\nabla u]_{\alpha,P_{3l}(z)} + \left(\frac{r}{l}\right)^{1-\alpha}\frac{1}{l^{2\alpha}}\inf_B\|\nabla w - B_{x'}\|_{P_{4l}(z)}.\end{aligned} \tag{2.35}$$

Let $K \ge 1$ and set $l = Kr$. By introducing $M'_z := \sup_{r>0}\frac{1}{r^{2\alpha}}\inf_B\|\nabla w - B_{x'}\|_{P_r(z)}$ we estimate combining (2.29) and (2.35)

$$M'_z \lesssim K^{2\alpha}[\nabla u]_\alpha + K^{\alpha-1}M_z$$

Thus, after using (2.6) as well as Lemma 2.1 we arrive at

$$M'_z \lesssim K^{2\alpha}C([\nabla v]_\alpha) + K^{\alpha-1}M_z$$

where the implicit constant in $\lesssim$ depends only on d, λ, Λ and α and $C([\nabla v]_\alpha)$ depends polynomially on $[\nabla v]_\alpha$. Since $\alpha > \frac{1}{2}$, using the reasoning in Step 3 in the proof of Lemma 3.6 of [**98**] the matrix B is independent of r, i.e.

$$M_z \lesssim M'_z$$

and hence

$$M_z \lesssim K^{2\alpha}C([\nabla v]_\alpha) + K^{\alpha-1}M_z.$$

Choosing K sufficiently large, and since $\alpha < 1$, as well as the assumption that M_z is finite, we can absorb $K^{\alpha-1}M_z$ into the left hand side and end up with

$$M_z \lesssim C([\nabla v]_\alpha) \tag{2.36}$$

and taking the supremum in z we get

$$M \lesssim C([\nabla v]_\alpha).$$

In order to get rid of the assumption that M_z is finite we appeal to the following approximation argument. Up to now we made the abbreviation $v = v_{a(t',x')}$ but from now on by v we mean the solution to the SHE (2.17). As in the proof of [**97**, Theorem 1], we can consider smooth approximations u_ε, $v_{a(t',x'),\varepsilon}$ and v_ε to u, $v_{a(t',x')}$ and v such that

$$\nabla\Big(u_\varepsilon - v_{a(t',x'),\varepsilon}\Big) \to \nabla\Big(u - v_{a(t',x')}\Big) = \nabla w$$

uniformly on compact sets and such that $[\nabla v_\varepsilon]_\alpha \lesssim [\nabla v]_\alpha$. Then the corresponding constant $M_{z,\varepsilon}$ is finite for all z and hence (2.36) applies and we have

$$M_\varepsilon \lesssim C([\nabla v_\varepsilon]_\alpha) \lesssim C([\nabla v]_\alpha).$$

Hence, by the same reasoning as in Proposition 2.3, we conclude by lower-semicontinuity

$$M \lesssim \liminf_{\varepsilon\to 0} M_\varepsilon \lesssim C([\nabla v]_\alpha). \tag{2.37}$$

In other words, (2.37) means that for all $r > 0$ and all space-time points $z = (t', x')$ there exists a family of symmetric matrices $(B(t', x'))_{(t',x')}$ and a family of vectors $(b(t', x'))_{(t',x')}$, such that we have

$$|\nabla u(t,x) - \nabla v_{a(t',x')}(t,x) - (B(t',x')(x-x') + b(t',x'))| \lesssim r^{2\alpha} \text{ for } (t,x) \in P_r(z).$$

Letting $r \to 0$ implies $(t,x) \to (t',x')$ and hence the optimal b is of the form

$$b(t',x') = \nabla u(t',x') - \nabla v_{a(t',x')}(t',x').$$

Then we can choose $r = d((t,x),(t',x'))$ and thus

$$\begin{aligned}|\nabla u(t,x) - \nabla v_{a(t',x')}(t,x) - (\nabla u(t',x') - \nabla v_{a(t',x')}(t',x')) - B(t',x')(x-x')|\\ \lesssim d^{2\alpha}((t,x),(t',x')),\end{aligned}$$

which we wanted to prove. □

CHAPTER 3

Structure-preserving discretization of the stochastic thin-film equation

In this chapter we consider the thin-film equation. It is known that this equation has a gradient flow structure with respect to a (generalized) Wasserstein metric and the usual Dirichlet energy. Based on that, the fluctuation-dissipation theorem gives rise to a stochastic thin-film equation. This equation is a singular stochastic partial differential equation which is out of scope of the framework of regularity structures for now. In order to circumvent this issue, we discretize the gradient flow structure and rediscover a well-known discretization of the (deterministic) thin-film equation that preserves the so-called entropy estimate. In the stochastic setting this entropy estimate then yields positivity for the solution in the case that the mobility arises from the no-slip boundary condition. Moreover, we show that the discretization of the stochastic thin-film equation, considered in the literature before, does not preserve positivity and we perform various numerical experiments to compare the discretizations in question.

This chapter is based on the article [**48**] which is joint work with Benjamin Gess, Rishabh S. Gvalani and Felix Otto.

3.1. Introduction

The thin-film equation models the evolution of the height h of a liquid film over a solid flat substrate, as driven by capillarity[1] and limited by viscosity. In the considered regime of small slope ($|\partial_x h| \ll 1$) and due to the no-slip boundary condition at the liquid-solid interface, viscous dissipation is so strong that the liquid's inertia can typically be neglected. Hence the dynamics are determined by a quasi-static balance between capillary and viscous forces. The lubrication approximation, which is based on a modulated Poiseuille Ansatz for the fluid velocity, leads to a fourth-order parabolic equation with a mobility that cubically degenerates in the film height.

In this chapter, we are interested in the thin-film equation driven by the noise that models thermal fluctuations. That noise takes the form of a conservative white noise with a multiplicative non-linearity. The specific form of the multiplicative non-linearity

[1]surface tension

– it is given by the square root of the mobility – formally arises from the fluctuation-dissipation principle (cf. [**30**, (4)]). While there exist elements of a well-posedness theory for (spatially) more regular forms of the noise in the mathematical literature (cf. [**44**], [**47**] and [**29**] and the next section for a detailed discussion), the stochastic partial differential equation (SPDE) we are interested in is expected to require a renormalization, and is theoretically uncharted. However, at least in $1+1$-space dimensions[2] as considered in this chapter, the invariant measure (on configuration space) of the SPDE does not require a renormalization. In this chapter we ignore the issue of renormalization and focus on spatial[3] discretizations of this SPDE.

The main issue is that the configuration space $\{h > 0\}$, which after discretization has the structure of an orthant, obviously has a boundary. The related preservation of positivity[4] has been at the core of the analysis of the deterministic thin-film equation, both on the continuum level (cf. [**16, 14, 27**]) and others, and on the level of spatial discretization (cf. [**54, 121**]). We refer to the end of the section for a more in-depth overview. The preservation of strict positivity is intimately related to what is called the entropy estimate, i. e. the existence of a Lyapunov functional on configuration space that blows up when h approaches zero. This Lyapunov functional depends on the mobility, and thereby arises from kinetics and dissipation, and thus is actually unrelated to the notion of entropy in thermodynamic equilibrium theory. In fact, the blowing up of the entropy as $h \downarrow 0$ is a consequence of a sufficiently strong degeneracy of the mobility. Of course, both in the discrete and the continuum case, such a touch-down can be suppressed by introducing a disjoining pressure. However, this feature comes with an additional (vertical) length scale of molecular size, and which one thus would like to avoid resolving. In this chapter, we therefore disregard this energetic mechanism preventing touch-down, and just focus on the above-mentioned kinetic mechanism.

In case of the thin-film equation with thermal noise, which in its discretized version describes a drift-diffusion process on the high-dimensional orthant $\{h > 0\}$, the question is even more pressing: Does the process reach the boundary or is the degeneracy of the mobility as $h \downarrow 0$, which translates into a degeneracy of the diffusion near the boundary of $\{h > 0\}$, strong enough to prevent reaching the boundary? The fact that the boundary may be reached has been already recognized in [**30**], where also an (uncontrolled) fix has been proposed. For a rigorous analysis of a given discretization, we need a multi-dimensional version of a Feller test. One main insight of this chapter is that such a Feller test can be carried out with help of the entropy mentioned above. It shows that for the

[2] which means that the profile is constant in one direction, so that the space variable x is one-dimensional
[3] by which we mean the physical space variable x, and not the state-space variable h
[4] often in form of preservation of non-negativity if the interest was in film spreading and (partial) wetting

physical mobility considered in this chapter, and in the case of $1+1$-dimensions, the numerical mobility, which was introduced in [**54**, Section 5] in order to prevent touch-down in the deterministic case, does also prevent touch-down in the presence of thermal noise (cf. Theorem 3.8). However, in Section 3.10 we provide evidence, through analysis of the path-space rate functional of the continuum stochastic thin-film equation, that the absence of touch-down maybe an artifact of discretization - for the continuum system touch-down is unlikely only for $m \geq 8$ (cf. Proposition 3.13).

The use of entropy estimates to construct non-negative solutions to the (deterministic) thin-film equation goes back to the original work [**16**, p.190, (4.12)], proving the existence of non-negative solutions for mobility exponents $1 < m < 4$ (see Assumption 3.6) and preservation of positivity for $m \geq 4$. Subsequently, these estimates were refined by means of so-called α-entropy estimates in [**14**, p.182, Proposition 2.1] and [**17**, p.99, (4.8) - (4.13)], which allowed to deduce the preservation of positivity for $m \geq \frac{7}{2}$. A generalization of the existence of non-negative solutions to multiple space dimension was given in [**52**] and extended to a wider range of mobility exponents in [**27**, p.324, Proposition 2.2]. Localized forms of α-entropy estimates were subsequently introduced in [**15**, Section 4] in $1+1$ dimensions and [**18**, p.422, Theorem 3.1] in higher space dimensions and in [**28**] used to prove upper bounds on the propagation of the support of solutions. Backward weighted entropy estimates have been introduced in [**42**, Section 3] and [**43**, p.3142, Lemma 11] to prove lower bounds on propagation rates. Also in the context of stochastic thin-film equations (with spatially regular noise) entropy estimates have been used in order to derive a-priori estimates and the existence of non-negative solutions [**44**, p.423, Proposition 4.3] and [**29**, p.20, Lemma 4.3].

As has been already mentioned, for the discretized thin-film equation the use of entropy estimates, which rely on an appropriate discretization of the mobility, dates back to [**54**, Section 5] in the case of a finite element discretization, and to [**121**, p.529, Proposition 3.1] in the case of a finite difference discretization. In the discrete case the corresponding entropy estimates have a stronger effect yielding positivity already for $m \geq 2$ in case of the two aforementioned discretizations. In this chapter we transfer the discretization and entropy estimate of [**54**] to the stochastic setting and get positivity for the scheme for $m \geq 3$ (cf. (3.8)).

3.2. State of the art

In [**53**, Section 2.3], the authors make the ansatz of an (infinite-dimensional) SDE in Itô form with a drift term given by[5] the deterministic thin-film operator (cf. [**53**, (36)]), and seek a noise term such that the process satisfies detailed balance with respect to the

[5]just, i. e. there is no Itô correction term

associated Gibbs measure (cf. [**53**, (21)]). They carry this out on the level of a finite-difference discretization in space, based on centered finite differences (cf. [**53**, p.1269]) which allows to use a local numerical mobility function (cf. [**53**, (29b)]). Thanks to this simple structure[6] they find that this is the case, provided the multiplicative noise involves the exact square root of the numerical mobility function (cf. [**53**, (33)]). However in this case, it is easy to see that the process does touch-down (cf. Section 3.9).

When it comes to actual simulations, [**53**] departs from this somewhat academic spatial discretization: They treat the noise term, which due to its conservative and multiplicative nature has the structure of a scalar conservation law with nonlinear and heterogeneous (in fact, rough) drift, via a finite volume discretization with an upwind scheme (cf. [**53**, (63),(64)]). The upwind scheme preserves non-negativity. For the deterministic term, they however use the numerical mobility introduced in [**54**] (cf. [**53**, (B.3)]), which is rather based on a lumped finite element interpretation (cf. [**53**, p.1275]). Again, at least on the purely deterministic level, this ensures non-negativity. Using two different, and nonlocal, numerical mobility functions however destroys the structure of exact detailed balance. The authors acknowledge this deficiency (cf. [**53**, p.1278]), mentioning that the deviation from detailed balance is vanishing (of first order) in the grid size. However, it is well-known that in the case of a singular SPDE, two different spatial discretizations, while both nominally first-order consistent, may lead to order-one different solutions (cf. [**61**]).

In [**37**, Sections 2 and 4], the authors repeat the derivation of the infinite-dimensional SDE of [**53**], but obtain it in the limit of fully correlated noise in the wall-normal direction for the long-wave/lubrication approximation of the so-called fluctuating hydrodynamics equations (cf. [**80**, §88, (88.6)-(88.18)]). Following [**53**], the authors make, essentially, an identical observation, that a finite difference discretization of the associated stochastic thin-film equation is formally reversible with respect to the associated Gibbs measure if and only if the multiplicative noise is given by the square root of the associated mobility.

Again, for the purposes of numerical simulations, [**37**] departs from the finite-difference discretization and instead proposes a spectral collocation method. The idea is to carry out the differentiation operations by decomposing the solution in terms of the eigenfunctions of the covariance operator of the noise, while treating the numerical mobility in a similar manner to the finite-difference discretization (cf. [**37**, Section 5.1, (71)-(72b)]). While this may have some structural advantages, it suffers from the drawback that it is unclear, and possibly untrue, that the spectral discretization satisfies detailed balance. Furthermore, it is also unclear if this scheme preserves the positivity of the film height.

[6]where there is no difference between the Itô and Stratonovich form

In recent years, the existence of probabilistically weak solutions to the stochastic thin-film equation has been considered in a sequence of works. In all of these works the noise term is spatially regularized. In [**44**], the authors constructed weak solutions for the case of quadratic mobility, relying on a conjoining-disjoining pressure term, and noise interpreted in Itô sense. In [**23**] more general mobilities were treated depending on a non-conservative source term. Both works require the initial condition to be strictly positive. For quadratic mobility and noise in Stratonovich sense, this restriction was lifted in [**47**]. The case of cubic mobility without additional conjoining-disjoining pressure term was recently treated in [**29**]. Recently, these results were extended to $2+1$ dimensions in [**86**] and [**106**].

3.3. The thin-film equation as a formal gradient flow

As discussed in Section 1.3.1 the thin-film equation is a gradient flow (cf. [**94**, p.2092 ff.]); we recall it here. In $1+1$ dimensions the equation takes the form

$$\partial_t h + \partial_x (M(h)\partial_x^3 h) = 0 \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}\,, \tag{3.1}$$

where h is the film height and M is called the mobility. In the following discussion, we tacitly think of $h > 0$ – this chapter does not address partial wetting, which would require more modelling assumptions at the contact line, like the equilibrium contact angle, possibly in conjunction with additional dissipation. Equation (3.1) is based on a lubrication approximation of a fluids equation, like Darcy or Stokes (cf. [**50, 73**]) and is a fourth order and possibly degenerate parabolic partial differential equation. The mobility $M(h)$ depends on the dissipation mechanism (e.g. Stokes vs. Darcy) and the boundary condition (e.g. no-slip vs. Navier) for the fluid velocity. Often, it is assumed that the mobility follows a power law, i.e. $M(h) \propto h^m$ for some $m \geq 0$. For example, Stokes with no-slip boundary conditions gives rise to $M(h) \propto h^3$ and this is also the most relevant case. Stokes with Navier slip leads to $M(h) \propto h^2$ for h below the slip length, and Darcy yields $M(h) \propto h$.

In this chapter, we make the convenient assumption that the solution h of (3.1) is 1-periodic. Since we clearly have conservation of mass, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_0^1 h \,\mathrm{d}x = 0,$$

we choose as the configuration space

$$\mathcal{M} := \Big\{ h : \mathbb{R} \to \mathbb{R} : h \text{ 1-periodic}, h > 0, \int_0^1 h \,\mathrm{d}x = 1 \Big\}.$$

The thin-film equation on $\mathcal{M}$ is driven by capillarity in the form of the Dirichlet energy

$$E(h) := \frac{1}{2}\int_0^1 (\partial_x h)^2 \,\mathrm{d}x \tag{3.2}$$

and limited by viscosity as described by the metric tensor [7] [8]

$$g_h\Big(\dot{h},\dot{h}\Big) := \inf_j \bigg\{ \int_0^1 \frac{j^2}{M(h)}\,\mathrm{d}x : \partial_x j + \dot{h} = 0 \bigg\} \tag{3.3}$$

where $\dot{h} \in T_h\mathcal{M}$, and the tangent space is given by

$$T_h\mathcal{M} = \bigg\{ \dot{h} : \mathbb{R} \to \mathbb{R} : \dot{h} \text{ 1-periodic}, \int_0^1 \dot{h}\,\mathrm{d}x = 0 \bigg\}.$$

For $M(h) = h$, this metric tensor corresponds to the infinitesimal metric in the $2-$Wasserstein distance (cf. [**13**, p.384, (35)-(36)] and [**95**, p.111]).

Hence, it is natural to expect that the thin-film equation has the structure of a gradient flow, i.e. that (3.1) can formally be written as

$$\partial_t h = -\nabla E(h).$$

This can be understood in the following way. The energy functional E gives rise to a differential defined as

$$\mathrm{diff} E|_h.\dot{h} := \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} E\Big(h + s\dot{h}\Big) \tag{3.4}$$

for $h \in \mathcal{M}$ and $\dot{h} \in T_h\mathcal{M}$, and we can define a gradient via the Riemannian structure for all $h \in \mathcal{M}$ as the unique element $\nabla E(h) \in T_h\mathcal{M}$ satisfying

$$\mathrm{diff} E|_h.\dot{h} = g_h\Big(\nabla E(h), \dot{h}\Big) \tag{3.5}$$

for all $\dot{h} \in T_h\mathcal{M}$. Hence, the gradient flow formulation $\partial_t h = -\nabla E(h)$ means that we have

$$\mathrm{diff} E|_h.\dot{h} + g_h\Big(\partial_t h, \dot{h}\Big) = 0 \tag{3.6}$$

for all $h \in \mathcal{M}$ and $\dot{h} \in T_h\mathcal{M}$. More precisely, by considering the Euler–Lagrange equation for (3.3), we have

$$g_h\Big(\dot{h},\dot{h}\Big) = \int_0^1 M(h)(\partial_x f)^2 \,\mathrm{d}x, \tag{3.7}$$

---

[7] for which, by polarization, it is enough to specify the quadratic part

[8] Note that $\partial_x j + \dot{h} = 0$ determines j up to an additive constant so that the infimum is taken on a single parameter. We opted for this representation because it extends verbatim to the higher dimensional case and will play a crucial role in the discretization.

where the 1−periodic f is such that $\dot{h} + \partial_x(M(h)\partial_x f) = 0$. By polarization of (3.7) and integration by parts we indeed obtain (3.6):

$$g_h\Big(\partial_t h, \dot{h}\Big) = \int_0^1 \dot{h}\partial_x^2 h \,\mathrm{d}x \overset{(3.2),(3.4)}{=} -\mathrm{diff}E|_h.\dot{h}.$$

Choosing $\dot{h} = \partial_t h$ in (3.6) we recover the energy dissipation identity characteristic of gradient flows

$$\frac{\mathrm{d}}{\mathrm{d}t}E(h) = -g_h(\partial_t h, \partial_t h) = -\int_0^1 M(h)(\partial_x^3 h)^2 \,\mathrm{d}x \leq 0\,.$$

Often, the energy has further contributions next to the one coming from capillarity (cf. (3.2)) giving for instance rise to a disjoining pressure. In fact, the choice of the energy functional will not be important for Section 3.6 and Section 3.7 and so if not otherwise stated we will not further specify E.

However, following [**16**, p.188, (4.3)] we define the function s as a solution to the equation $s'' = \frac{1}{M}$ and then for E being the Dirichlet energy this yields another Lyapunov functional

$$S(h) := \int_0^1 s(h)\,\mathrm{d}x \tag{3.8}$$

called entropy in the mathematical literature, and the following entropy estimate

$$\frac{\mathrm{d}}{\mathrm{d}t}S(h) = -\int_0^1 (\partial_x^2 h)^2 \,\mathrm{d}x \leq 0 \tag{3.9}$$

holds. This estimate will play a major role in Section 3.8.

The preservation of positivity can also be interpreted geometrically in the sense that the evolution on the configuration space $\mathcal{M}$ does not touch its boundary $\partial\mathcal{M}$.

3.4. Thermodynamically consistent introduction of fluctuations

3.4.1. Invariant measure on configuration space and the associated reversible dynamics. In agreement with the standard equilibrium thermodynamics, we postulate that the invariant measure on configuration space of the stochastic dynamics is given by the Gibbs measure

$$\mathrm{d}\nu(h) = \frac{1}{Z}e^{-\beta E(h)}\,\mathrm{d}h \tag{3.10}$$

for some $\beta > 0$, which up to the Boltzmann factor is the inverse temperature, and a normalization constant Z. Here one thinks of $\mathrm{d}h$ as a uniform measure on the configuration space $\mathcal{M}$. In the special case where the energy functional is the Dirichlet energy (cf. (3.2)), the measure (3.10) looks similar to the classical Wiener measure. This relation, though, is not quite correct due to the following three reasons. First of all, we are on a periodic domain and, secondly, we have the additional constraint $\int_0^1 h\,\mathrm{d}x = 1$. Finally, the restriction to the orthant $\{h > 0\}$ is the major difference.

Hence we have to think of (3.10) as a Gaussian measure conditioned to be non-negative, i.e.

$$\mathrm{d}\nu(h) = \frac{1}{Z}\mathbb{1}\{h > 0\}\,\mathrm{d}\mu(h) \tag{3.11}$$

where μ is the so-called Gaussian free field, i.e. the stationary Gaussian measure with covariance operator given by $\left(-\beta\partial_x^2\right)^{-1}$ and conditioned on the spatial average being 1. We will refer to the measure ν on $\mathcal{M}$ as the conservative Brownian excursion due to its reminiscence to the classical Brownian excursion from stochastic analysis. Notice, however, that unlike in the case of the classical Brownian excursion, the set $\{h \geq 0\}$ we are conditioning on is not a null set with respect to the measure μ. In other words, the conservative Brownian excursion (3.11) is absolutely continuous with respect to the Gaussian free field, and it is well known that the latter is supported on $C^{\frac{1}{2}-}$-functions, and hence so is ν.

In the case of zero Dirichlet boundary data, the Brownian bridge conditioned to non-negative functions $\mathrm{d}\widetilde{\nu}(h) = \frac{1}{Z}\mathbb{1}_{\{h\geq 0\}}\,\mathrm{d}\mu(h)$ corresponds to the law of the Brownian excursion, which in turn is the law of the 3d Bessel bridge (cf. [**118**, p.205, Theorem 3]). As a consequence, the transience of the 3d Brownian motion implies that $\widetilde{\nu}$ is supported on positive functions. This repulsive effect of the boundary $\partial\mathcal{M}$ is called entropic repulsion. Entropic repulsion in discrete systems and interface models has been analyzed, for example, in [**33**]. Brownian excursion with fixed average has been realized as an invariant measure of an SPDE in [**119**].

We note in passing that in $2+1$-dimensions, the Gaussian measure would be related to the two-dimensional Gaussian free field, so that in view of the latter's ultraviolet logarithmic divergence, the conditioning on $h > 0$ is (borderline) singular; hence the nature of the Gibbs measure is unclear in this case.

We now turn to the stochastic dynamics. We follow the standard Ansatz that the time evolution of the law ν_t – which we will assume to be absolutely continuous with respect to the invariant measure ν – of the stochastic thin-film equation is described by the following Fokker–Planck equation in variational form, i.e. we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathcal{M}} \zeta\,\mathrm{d}\nu_t = -\frac{1}{\beta}\int_{\mathcal{M}} g(\nabla\zeta, \nabla f_t)\,\mathrm{d}\nu \tag{3.12}$$

for all sufficiently nice test functions ζ and where $f_t := \frac{\mathrm{d}\nu_t}{\mathrm{d}\nu}$. It is obvious from (3.12) that ν is indeed invariant. The symmetry of the so-called Dirichlet form on the r.h.s. of (3.12) implies that the generator $\mathcal{L}$, which is defined as the representation of the Dirichlet form w.r.t. $L^2(\mathrm{d}\nu)$, is symmetric. This in turn yields that the stochastic process is reversible, meaning that the invariant measure on path space is invariant under reversing the time direction. As we will see later, this ansatz will ensure that the dynamics obey the detailed balance condition known from thermodynamics.

3.4.2. Renormalization of the thin-film equation with thermal noise. In [**30**, (4)] it has been suggested that the thin-film equation with thermal noise is given by

$$\partial_t h + \partial_x \Big(M(h) \partial_x^3 h \Big) = \partial_x \Big(\sqrt{M(h)} \xi \Big) \tag{3.13}$$

where ξ denotes space-time white noise. In the course of this chapter, it will become apparent that (3.13) arises from (3.12). First, we explain why equation (3.13) is singular as an SPDE which means that there are nonlinear terms which are not well-defined a priori in a classical sense. This is in contrast to versions of the thin-film equation driven by a less singular (and thus less physical) noise than white noise, for which a well-posedness theory exists, see the discussion in Section 2.

As a consequence of the characterization of the invariant measure on configuration space in Section 3.4.1, we expect typical solutions h of the thin-film equation with thermal noise to have spatial regularity in the Hölder class $C^{\frac{1}{2}-}$ and not better. Hence, the product $M(h)\partial_x^3 h$ appearing in the thin-film operator is the product of a function in $C^{\frac{1}{2}-}$ and a distribution in the negative Hölder space[9] $C^{-\frac{5}{2}-}$ and thus ill-defined (and more than just border-line since $(\frac{1}{2}-) + (-\frac{5}{2}-) = -2-$).

Moreover, we encounter a similar difficulty in the multiplicative noise term that formally is given by $\partial_x(\sqrt{M(h)}\xi)$: Since the effective dimension for our fourth-order parabolic operator in one space dimension is $4 + 1 = 5$, ξ is in the negative Hölder class $C^{-\frac{5}{2}-}$ (which can be defined as $\partial_t C^{\frac{3}{2}-} + \partial_x^3 C^{\frac{1}{2}-}$, where space-time Hölder spaces are defined w. r. t. to the anisotropic fourth-order parabolic Carnot-Carathéodory norm). Hence the product $\sqrt{M(h)}\xi$ has the same singular nature as the product $M(h)\partial_x^3 h$. This similarity in the degree of singularity is reminiscent of quasi-linear second-order equations (cf. [**97**]). We stress that these difficulties are unrelated to the degeneracy[10] of M.

Hence, the thin-film equation with thermal noise is in need of a renormalization, a pressing and attractive topic for the theory of singular SPDE. In this chapter, we do not further address this issue for several reasons: 1) In 1+1-space dimensions, as mentioned above, the invariant measure is not in need of a renormalization. Hence the situation is better than in case of the well-studied stochastic quantization equation[11]. The invariant

[9] see a couple of sentences below for a definition
[10] meaning that $M(0) = 0$
[11] which comes in form of the Allen-Cahn equation driven by space-time white noise

measure for the latter equation[12] is in need of a renormalization for space dimensions ≥ 2 (and renormalizable in dimensions < 4). 2) In this chapter, we focus on structural properties of spatial discretizations that can be rigorously addressed without a well-posedness theory for the continuum limit. 3) A simple but typical scaling argument suggests that our problem is renormalizable in 1+1-space dimensions. Indeed, zooming in on small length and time scales through

$$x = \ell \hat{x}, \quad t = \ell^4 \hat{t}, \quad h = 1 + \ell^{\frac{1}{2}} \hat{h}, \quad \xi = \ell^{-\frac{5}{2}} \hat{\xi}, \tag{3.14}$$

where the rescaling of ξ is such that $\hat{\xi}$ is another instance of space-time white noise, and where 1 could be replaced by any positive constant, the equation (3.13) turns into

$$\partial_{\hat{t}} \hat{h} + \partial_{\hat{x}} (M(1 + \ell^{\frac{1}{2}} \hat{h}) \partial_{\hat{x}}^3 \hat{h}) = \partial_{\hat{x}} (\sqrt{M(1 + \ell^{\frac{1}{2}} \hat{h})} \hat{\xi}),$$

from which we learn that on small scales, the non-linearity fades away [13]. A similar computation shows that in 2+1-space dimensions the stochastic thin-film equation is critical, i.e. the rescaling (3.14) leaves the equation (3.13) invariant and hence the nonlinear terms persist on small scales.

There is a fourth point that we would like to make. Although at first sight the singular nature of the equation is very far from borderline, it is better than expected in some specific cases. As is common in the deterministic rigorous treatment, one could rewrite the non-linearity in the thin-film operator in a less singular way:

$$M(h) \partial_x^3 h = \partial_x^3 \overline{M}(h) - \frac{3}{2} \partial_x \Big(M'(h) (\partial_x h)^2 \Big) + \frac{1}{2} M''(h) (\partial_x h)^3$$

where $\overline{M}$ is the antiderivative of M. Of course the terms $(\partial_x h)^2$ and $(\partial_x h)^3$ are still singular but if we choose the following ansatz for renormalization which is inspired by the ϕ^4-model

$$(\partial_x h)^2 \to (\partial_x h)^2 - C, \quad (\partial_x h)^3 \to (\partial_x h)^3 - 3C \partial_x h$$

the divergent constant C drops out since by the chain rule

$$\begin{aligned} &- \frac{3}{2} \partial_x \Big(M'(h) \Big((\partial_x h)^2 - C \Big) \Big) + \frac{1}{2} M''(h) \Big((\partial_x h)^3 - 3C \partial_x h \Big) \\ &= - \frac{3}{2} \partial_x \Big(M'(h) (\partial_x h)^2 \Big) + \frac{1}{2} M''(h) (\partial_x h)^3. \end{aligned}$$

While this argument suggests that the non-linearity $M(h) \partial_x^3 h$ is less singular than expected, we now argue that the non-linearity $\sqrt{M(h)} \xi$ can be completely avoided in case of linear mobility, i.e. $M(h) = h$. It is well known (cf. [**115**, p.74, Theorem 2.18])

[12] also known as ϕ^4 model in quantum field theory

[13] this discussion obviously ignores additional difficulties that may arise from the degeneracy of the mobility

that for linear mobility under the change of variables $h \mapsto X$ where X is the inverse distribution function of h, i.e.

$$z = \int_0^{X(z)} h(x)\,\mathrm{d}x, \tag{3.15}$$

the metric tensor transforms as

$$g_h\Big(\dot{h}, \dot{h}\Big) = \int_0^1 \dot{X}^2\,\mathrm{d}z = g_X\Big(\dot{X}, \dot{X}\Big).$$

The Dirichlet energy transforms according to

$$E(X) = \frac{1}{2}\int_0^1 \frac{\Big(\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z)\Big)^2}{\Big(\frac{\mathrm{d}}{\mathrm{d}z}X(z)\Big)^5}\,\mathrm{d}z.$$

Hence the deterministic dynamics amount to the L^2-gradient flow of E, which is seen to assume the form

$$\partial_t X = \frac{1}{4}\partial_z^3(\partial_z X)^{-4} - \frac{5}{8}\partial_z\Big(\partial_z(\partial_z X)^{-2}\Big)^2$$

and then (3.12) can be seen to translate into

$$\partial_t X = \frac{1}{4}\partial_z^3(\partial_z X)^{-4} - \frac{5}{8}\partial_z\Big(\partial_z(\partial_z X)^{-2}\Big)^2 + \xi \tag{3.16}$$

where ξ is space-time white noise. The first term on the right hand side of (3.16) is well-defined since $\partial_z X$ behaves like h (cf. (3.75)), and a non-linearity in the Hölder continuous h is still harmless. For the second term on the right hand side of (3.16), we notice that it is a "KPZ-like" term followed by a derivative. Since the renormalization constant for the KPZ equation does not depend on the space variable (cf. [**46**, p.223, Theorem 15.1]) we might expect that in this case it is annihilated by the outer derivative. Thus, one may expect that the leading order counter terms are zero and one obtains only higher order counter terms.

We will comment further on the possible structure of a renormalizing counter term in Remark 3.12, once we introduce both our discretization and the central difference discretization. Furthermore, in the numerical experiments performed in Section 3.11.4 we observe that the two-point[14] distribution functions of the two discretizations we are considering in this chapter converge to the same object. This provides some numerical evidence for the our guess that equation (3.13) is less singular than expected, even for $M(h) = h^3$.

[14]in time

3.5. Discretization

A numerical treatment requires a discretization. From the Fokker–Planck equation in its variational form (4.2) we learn that it is determined by the triple $(\mathcal{M}, g, E)$, which hence we need to discretize. For the function space $\mathcal{M}$, we choose a Finite Element discretization. More precisely, we fix $N \in \mathbb{N}$ and denote the equidistant partition of the torus by $\{x_i\}_{i=1,\ldots,N}$. Then we denote by P_1 the space of 1-periodic, continuous, and piecewise linear (with respect to the equidistant partition) functions and we set

$$\mathcal{M}_N := \mathcal{M} \cap P_1 = \Big\{ h \in P_1 : h > 0, \int_0^1 h \,\mathrm{d}x = 1 \Big\},$$

which then comes with a canonical tangent bundle $T\mathcal{M}_N$. For the functional E, we make a conformal Ansatz by restricting to $\mathcal{M}_N$. This gives rise to a discretized conservative Brownian excursion ν_N according to (3.10). Finally, we need to specify a metric tensor on $T\mathcal{M}_N \otimes T\mathcal{M}_N$. A natural discretization of the metric tensor would be its restriction to the space $\mathcal{M}_N$. However, we will not consider this discretization in this chapter for reasons explained in Remark 3.7.

3.6. Introducing coordinates

In Section 3.4.1 we have already seen how a gradient flow structure, as determined by a Riemannian manifold $(\mathcal{M}, g)$ and a function E, gives rise to a stochastic process via the Fokker–Planck equation (cf. (3.12)). In this section, we aim to write this process in Itô form. To this end, we need to introduce coordinates. Let $\mathcal{M}$ be a differentiable Riemannian manifold with boundary, equipped with a Riemannian metric g, and assume that we have a global chart

$$(\varphi^\alpha)_\alpha : \mathcal{M} \to \Delta,$$

where Δ is an open subset of $\mathbb{R}^N$ with coordinates enumerated by $\alpha = 1, \ldots, N$. Moreover, we think of $\mathcal{M}$ as equipped with a probability measure ν. Then, these data give rise to a Fokker–Planck equation in variational form (cf. (3.12)) which describes the time evolution of the probability measure ν_t which we assume to be absolutely continuous with respect to ν. Hence (3.12) gives rise to a Markovian stochastic process on $\mathcal{M}$ of which ν is the invariant measure. By the symmetry of the right hand side of (3.12), the resulting process on path space is reversible.

The chart $(\varphi^\alpha)_\alpha$ allows to pull back functions from Δ to $\mathcal{M}$ and thus to push forward measures from $\mathcal{M}$ to Δ. For notational convenience we will not distinguish between $\zeta \circ \varphi$ and ζ, between $f_t \circ \varphi$ and f_t, and between $\varphi \# \nu_t$ and ν_t, and will write h^α instead of $\varphi^\alpha(h)$. A quick calculation shows that Radon-Nikodym derivatives transform like functions; in particular, the relation $\mathrm{d}\nu_t = f_t \,\mathrm{d}\nu$ lifts from $\mathcal{M}$ to Δ. By the usual duality, we define the gradient of φ^α as the unique element $\nabla\varphi^\alpha(h) \in T_h\mathcal{M}$ satisfying

$$\mathrm{diff}\varphi^\alpha|_h.\dot{h} = g_h\Big(\nabla\varphi^\alpha(h), \dot{h}\Big)$$

for all $\dot{h} \in T_h\mathcal{M}$ (cf. (3.5)). While here, we think of the metric tensor as a bilinear form on tangent vectors, it is now convenient to consider its dual, a bilinear form on co-tangent vectors like differentials. The coordinate representation of this dual metric tensor is given by

$$g^{\alpha\alpha'}(h) = \mathrm{diff}\varphi^\alpha|_h.\nabla\varphi^{\alpha'}(h). \tag{3.17}$$

The upper indices indicate the 2 contra-variant nature of the dual metric tensor. In fact, seen as a matrix, it is the inverse of the metric tensor $g_{\alpha\alpha'}(h)$ (cf. (3.81)). Then by 3.82, we get

$$g(\nabla\zeta, \nabla f_t) = g^{\alpha\alpha'}\partial_\alpha\zeta\partial_{\alpha'}f_t.,$$

where from now on we will use the Einstein convention of summing over repeated indices if not otherwise stated. Hence, we end up with the Fokker–Planck equation in variational form on Δ, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Delta \zeta\,\mathrm{d}\nu_t = -\frac{1}{\beta}\int_\Delta g^{\alpha\alpha'}\partial_\alpha\zeta\partial_{\alpha'}f_t\,\mathrm{d}\nu \tag{3.18}$$

for all sufficiently nice test functions ζ. Without much loss of generality, we assume that ν is given by

$$\mathrm{d}\nu(h) = \frac{1}{Z_\beta}e^{-\beta E(h)}\,\mathrm{d}h$$

for some function $E : \Delta \to \mathbb{R}$ where $\mathrm{d}h$ denotes the Lebesgue measure on Δ. For brevity we set $\rho_\infty := \frac{1}{Z_\beta}e^{-\beta E}$. Then we apply the divergence theorem which yields the following equation for f_t

$$\begin{cases} \rho_\infty\partial_t f_t = \frac{1}{\beta}\partial_\alpha\Big(g^{\alpha\alpha'}\rho_\infty\partial_{\alpha'}f_t\Big) & \text{on} \quad \mathbb{R}_{>0}\times\Delta, \\ n_\alpha g^{\alpha\alpha'}\partial_{\alpha'}f_t = 0 & \text{on} \quad \mathbb{R}_{>0}\times\partial\Delta, \end{cases} \tag{3.19}$$

where $n = (n_\alpha)_\alpha$ denotes the outer normal of the boundary $\partial\Delta$. Moreover, considering the probability density ρ_t defined through

$$\rho_t := f_t\rho_\infty$$

we see by (3.19) and the Leibniz rule that ρ_t solves the Fokker–Planck equation

$$\begin{cases} \partial_t\rho_t = \partial_\alpha\Big(g^{\alpha\alpha'}\Big(\frac{1}{\beta}\partial_{\alpha'}\rho_t + \rho_t\partial_{\alpha'}E\Big)\Big) & \text{on} \quad \mathbb{R}_{>0}\times\Delta, \\ n_\alpha g^{\alpha\alpha'}\Big(\frac{1}{\beta}\partial_{\alpha'}\rho_t + \rho_t\partial_{\alpha'}E\Big) = 0 & \text{on} \quad \mathbb{R}_{>0}\times\partial\Delta. \end{cases} \tag{3.20}$$

Note that (3.20) can be seen as a continuity equation for the probability density with the probability flux $J(\rho)$ being defined in components as

$$J^{\alpha}(\rho) := g^{\alpha\alpha'}\Big(\frac{1}{\beta}\partial_{\alpha'}\rho + \rho\partial_{\alpha'}E\Big).$$

Then not only is ρ_∞ the stationary solution of (3.20) but in fact we have that

$$J(\rho_\infty) \equiv 0$$

which corresponds to the detailed balance condition (cf. [**100**, p.119, (4.97)]). Instead of describing the evolution of the law through (3.20), we can use the duality between measures and continuous functions to compute the evolution of observables u of the process. Indeed, by computing the formal adjoint of (3.20), we can read off the following backward Kolmogorov equation:

$$\begin{aligned}\partial_t u_t &= \frac{1}{\beta}\partial_\alpha\Big(g^{\alpha\alpha'}\partial_{\alpha'}u_t\Big) - \partial_\alpha u_t g^{\alpha\alpha'}\partial_{\alpha'}E \\ &= \frac{1}{\beta}g^{\alpha\alpha'}\partial_{\alpha\alpha'}u_t + \partial_\alpha u_t\Big(\frac{1}{\beta}\partial_{\alpha'}g^{\alpha'\alpha} - g^{\alpha\alpha'}\partial_{\alpha'}E\Big)\end{aligned} \tag{3.21}$$

in Δ equipped with the boundary conditions on $\partial\Delta$

$$n_\alpha g^{\alpha\alpha'}\partial_{\alpha'}u_t = 0. \tag{3.22}$$

Note that the right hand side of (3.21) is the generator of the associated diffusion process. Thus, we can use (3.21) to identify the stochastic process h_t^α arising from (3.18). Indeed, its drift is given by $-\Big(\frac{1}{\beta}\partial_{\alpha'}g^{\alpha'\alpha} - g^{\alpha\alpha'}\partial_{\alpha'}E\Big)(h_t)$ and its diffusion matrix by $\frac{1}{\beta}g^{\alpha\alpha'}(h_t)$. This gives rise to the following stochastic differential equation in Itô form (cf. [**93**, p.126, Theorem 7.3.3 and p.152, Theorem 8.4.3])

$$\mathrm{d}h_t^\alpha = \Big(-g^{\alpha\alpha'}(h_t)\partial_{\alpha'}E(h_t) + \frac{1}{\beta}\partial_{\alpha'}g^{\alpha'\alpha}(h_t)\Big)\,\mathrm{d}t + \sigma^\alpha_{\alpha'}(h_t)\sqrt{\frac{2}{\beta}}\,\mathrm{d}W_t^{\alpha'}, \tag{3.23}$$

where $\sigma^\alpha_{\alpha'}$ denotes any matrix satisfying $g^{\alpha\alpha'} = \sum_{\alpha''=1}^N \sigma^\alpha_{\alpha''}\sigma^{\alpha'}_{\alpha''}$, and W_t is a standard Wiener process. Furthermore, the no-flux boundary conditions in (3.22) correspond to reflecting boundary conditions in (3.23) (cf. [**68**, p.222, Theorem 7.1]). The main purpose of this subsection was to elucidate the emergence of the Itô-correction term $\frac{1}{\beta}\partial_{\alpha'}g^{\alpha'\alpha}$.

3.7. The Grün–Rumpf metric

In [**54**, Section 5] the authors have introduced a discretization of the deterministic thin-film equation in a way such that a discrete version of the entropy estimate (3.9) holds; see Lemma (3.5). They propose a finite element discretization and in particular introduce a specific discretization of the mobility. As it turns out, the latter can be interpreted as a mixed finite element discretization with lumping of the metric tensor (3.3); see Definition (3.1). At the same time, the authors of [**121**] have considered a finite difference discretization of (3.1) with a similar discretization of the mobility as [**54**] that preserves the entropy estimate.

As has been discussed in the last section, the space $\mathcal{M}_N$ is the configuration space for the discretized stochastic thin-film equation. Any function in P_1, and thus also any $h \in \mathcal{M}_N$, is uniquely determined by its values at the nodal points $\{x_i\}_{i=1,\ldots,N}$. This gives rise to a natural chart

$$\left(\varphi^i\right)_i : \mathcal{M}_N \to \Delta_N$$

where for $h \in \mathcal{M}_N$ we have

$$h = \varphi^i(h)\hat{\varphi}_i. \tag{3.24}$$

Here Δ_N is the N-simplex defined as

$$\Delta_N := \left\{ h \in \mathbb{R}^N : h^i > 0, \frac{1}{N}\sum_{i=1}^{N} h^i = 1 \right\}$$

and for $i = 1, \ldots, N$ we define $\hat{\varphi}_i$ to be the unique piecewise linear and continuous function such that we have

$$\hat{\varphi}_i(x_j) = \delta_{ij}.$$

The family $(\hat{\varphi}_i)_{i=1,\ldots,N}$ is of course known as the hat basis in finite elements. As in

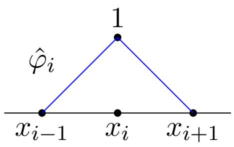


FIGURE 1. An element $\hat{\varphi}^i$ of the hat basis.

Section 4.1 we will write h^i instead of $\varphi^i(h)$. We denote by P_0 the space of piecewise constant functions and we note that $T_h\mathcal{M}_N := T_h\mathcal{M} \cap P_1$.

We now turn to the discretization of (3.3); in suitable coordinates it amounts to a reinterpretation of the metric considered in Section 5 of [**54**], see Remark 3.3. For the discretization of (3.3), following the strategy of first discretizing and then periodizing leads to a simpler result, and we shall follow it here. Hence Definition 3.1 is phrased with the unit torus replaced by $\mathbb{R}$[15].

[15]with the abuse of keeping the notation $\mathcal{M}_N$

Definition 3.1 (*Grün-Rumpf* metric)**.** Let $h \in \mathcal{M}_N$ and $\dot{h} \in T_h\mathcal{M}_N$. We define a metric tensor on $T_h\mathcal{M}_N \otimes T_h\mathcal{M}_N$ via

$$g_h(\dot{h}, \dot{h}) :=$$
$$\inf_j \left\{ \int_{\mathbb{R}} \frac{j^2}{M(h)} \,\mathrm{d}x : j \in P_0, \int_{\mathbb{R}} j \partial_x \zeta \,\mathrm{d}x = \frac{1}{N} \sum_{i \in \mathbb{Z}} \dot{h}^i \zeta^i \;\; \forall \zeta \in P_1 \text{ compactly supported} \right\}.$$

Remark 3.2. As mentioned earlier (3.1) is a mixed finite element discretization with lumping of (3.3). By a mixed discretization, we mean that we are not just discretizing the configuration space but also the space of fluxes, i.e. we require $j \in P_0$. Moreover, lumping means that instead of the L^2-inner product $\int_{\mathbb{R}} \dot{h}\zeta \,\mathrm{d}x$ we use the ℓ^2-inner product $\frac{1}{N}\sum_{i\in\mathbb{Z}} \dot{h}^i \zeta^i$.

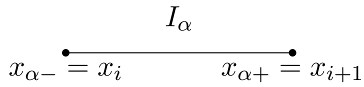


FIGURE 2. Relation of the intervals $(I_\alpha)_\alpha$ and the nodal points $\{x_i\}_i$.

Now we come to the choice of coordinates. In order to obtain a simpler expression of the metric tensor it is better to introduce another basis than the hat basis. For any α let $\overline{\varphi}_\alpha \in P_1$ be given by (see Figure (2))

$$\overline{\varphi}_\alpha := N^{\frac{3}{2}}(\hat{\varphi}_{\alpha+} - \hat{\varphi}_{\alpha-}). \tag{3.25}$$

We call the family $(\overline{\varphi}_\alpha)_{\alpha=1,\ldots,N}$ the *zigzag* basis [16]. Now we can introduce another set of coordinates given by the chart

$$(\varphi^\alpha)_\alpha : \mathcal{M}_N \to \mathbb{R}^N$$

where[17]

$$h = \varphi^\alpha(h)\overline{\varphi}_\alpha + 1. \tag{3.26}$$

Here 1 denotes the constant function with that value. Again for simplicity, instead of writing $\varphi^\alpha(h)$ we write h^α.

We note that by the relation (3.26), for every $h \in \mathcal{M}_N$ the induced basis on $T_h\mathcal{M}_N$ is given by the *zigzag* basis. Hence in these coordinates, the metric tensor (3.1) takes the form

$$g_{\alpha\alpha'}(h) := g_h(\overline{\varphi}_\alpha, \overline{\varphi}_{\alpha'}).$$

[16] As is easily seen it holds that $\sum_{\alpha=1}^N \overline{\varphi}_\alpha = 0$ and thus the *zigzag* basis is not really a basis. This issue is resolved by requiring that $\sum_{\alpha=1}^N h^\alpha = 0$.

[17] The image of $(\varphi^\alpha)_\alpha$ is also affine linear.

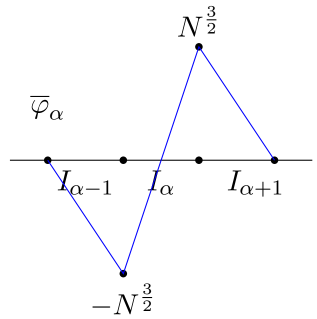


FIGURE 3. An element $\overline{\varphi}_\alpha$ of the *zigzag basis*.

Note that for any α we have

$$\frac{1}{N}\sum_{i\in\mathbb{Z}}(\overline{\varphi}_\alpha)^i\zeta^i = \sqrt{N}\Big(\zeta^{\alpha+}-\zeta^{\alpha-}\Big).$$

Similarly, we compute

$$\int_{\mathbb{R}} j\partial_x\zeta \,\mathrm{d}x = \sum_{\alpha'\in\mathbb{Z}} j_{\alpha'}\Big(\zeta^{\alpha'+}-\zeta^{\alpha'-}\Big)$$

where $j=\sum_{\alpha\in\mathbb{Z}} j_\alpha \mathbb{1}_{I_\alpha}$. Hence we see that an admissible choice is $j=\sqrt{N}\mathbb{1}_{I_\alpha}$ and since any other choice only differs by an additive constant this is already the optimal choice and this yields

$$g_{\alpha\alpha'}(h) = ⨍_{I_\alpha}\frac{1}{M(h)}\,\mathrm{d}x\ \delta_{\alpha\alpha'}.$$

As in Section 4.1 we denote the dual metric associated to (3.1) by $\Big(g^{\alpha\alpha'}(h)\Big)_{\alpha,\alpha'}$ and, since $g^{\alpha\alpha''}(h)g_{\alpha''\alpha'}(h)=\delta^\alpha_{\alpha'}$ (see (3.81)) we have

$$g^{\alpha\alpha'}(h) = \left(⨍_{I_\alpha}\frac{1}{M(h)}\,\mathrm{d}x\right)^{-1}\delta^{\alpha\alpha'}. \tag{3.27}$$

Having derived this discretization, we again impose a periodic data structure on the discrete level.

Remark 3.3. On every interval I_α the expression (3.27) is the harmonic mean of the mobility $M(h)$ and thus we recover the discretization proposed in [**54**, Section 5].

As mentioned in the last section, the discretization of the energy is just the restriction of E to the space P_1. Then according to (3.23) this specific discretization gives rise to the following SDE

$$\mathrm{d}h^\alpha_t = \left(-g^{\alpha\alpha'}(h_t)\partial_{\alpha'}E(h_t)+\frac{1}{\beta}\partial_{\alpha'}g^{\alpha'\alpha}(h_t)\right)\mathrm{d}t+\sigma^\alpha_{\alpha'}(h_t)\sqrt{\frac{2}{\beta}}\,\mathrm{d}W^{\alpha'}_t. \tag{3.28}$$

Definition 3.4. Restricting the derivative ∂_x to P_1 yields a linear operator $\partial_x : P_1 \to P_0$. We denote the matrix representation of this linear operator with respect to the hat basis on P_1 and the basis $(\mathbb{1}_{I_\alpha})_\alpha$ on P_0 by $A = (A_i^\alpha)_i^\alpha$, i.e. we have

$$A_i^\alpha b^i = N\left(b^{\alpha+} - b^{\alpha-}\right) \tag{3.29}$$

for all vectors $\left(b^i\right)^i$. Moreover, its transpose is given by

$$\left(A^T\right)_\alpha^i = A_i^\alpha$$

Now we pass from α-coordinates to i-coordinates. To this end, we compute

$$h \overset{(3.26)}{=} h^\alpha \overline{\varphi}_\alpha + 1 \overset{(3.25)}{=} N^{\frac{3}{2}} h^\alpha (\hat{\varphi}_{\alpha+} - \hat{\varphi}_{\alpha-}) + 1 = \sqrt{N} h^\alpha \left(A^T\right)_\alpha^i \hat{\varphi}_i + 1.$$

Thus by (3.24) we obtain the formula

$$h^i = \sqrt{N}\left(A^T\right)_\alpha^i h^\alpha + 1. \tag{3.30}$$

Then (3.B.2) and the chain rule yield

$$\partial_\alpha = \sqrt{N}\left(A^T\right)_\alpha^i \partial_i. \tag{3.31}$$

Hence, by applying (3.30) to (3.28) and (3.31) only to the first drift term, we end up with the following SDE in i-coordinates

$$\begin{aligned} \mathrm{d}h_t^i = &\left(-N\left(A^T\right)_\alpha^i g^{\alpha\alpha'}(h_t)\left(A^T\right)_{\alpha'}^j \partial_j E(h_t) + \frac{\sqrt{N}}{\beta}\left(A^T\right)_\alpha^i \partial_{\alpha'} g^{\alpha'\alpha}(h_t)\right) \mathrm{d}t \\ &+ \left(A^T\right)_\alpha^i \sigma_{\alpha'}^\alpha (h_t) \sqrt{\frac{2N}{\beta}}\, \mathrm{d}W_t^{\alpha'} \end{aligned} \tag{3.32}$$

subject to reflecting boundary conditions. It is easy to see that the Itô-correction term in the discrete thin-film equation with thermal noise (3.32) does in general not vanish, see (3.84) for the case $M(h) = h^3$.

3.8. Positivity of the scheme

As it turns out, the *Grün–Rumpf* metric is the right discretization in order to preserve positivity. From now on it will be important that the energy functional is the Dirichlet energy (3.2). In view of Definition 3.4, the restriction of E to $\mathcal{M}_N$ assumes the form

$$E(h) = \frac{1}{2N}\sum_{\alpha=1}^{N}((Ah)^\alpha)^2 = \frac{1}{2N} h^j A_j^\alpha \delta_{\alpha\alpha'} A_k^{\alpha'} h^k$$

and hence

$$\partial_j E(h) = \frac{1}{N} A_j^\alpha \delta_{\alpha\alpha'} A_k^{\alpha'} h^k. \tag{3.33}$$

Plugging this in the first drift term of (3.32) yields

$$-\left(A^T\right)^i_\alpha g^{\alpha\alpha'}(h_t)\left(A^T\right)^j_{\alpha'} A^{\alpha''}_j \delta_{\alpha''\alpha'''} A^{\alpha'''}_k h^k_t.$$

Instead of viewing $\partial_j E$ as a covector it makes sense to regard it as a vector. To this end, we contract the metric $g^{\alpha\alpha'}$ with respect to the ambient Euclidean metric, i.e.

$$g^{\alpha\alpha'} = g^\alpha_\gamma \delta^{\gamma\alpha'}$$

and this yields

$$g^{\alpha\alpha'}\left(A^T\right)^j_{\alpha'} A^{\alpha''}_j \delta_{\alpha''\alpha'''} A^{\alpha'''}_k h^k_t = g^\alpha_{\alpha'} A^{\alpha'}_j \left(A^T\right)^j_{\alpha''} A^{\alpha''}_k h^k_t = g^\alpha_{\alpha'}\left(AA^T A h_t\right)^{\alpha'}_k. \tag{3.34}$$

Furthermore we specify $\sigma^\alpha_{\alpha'}(h)$ to be the square-root of $g^\alpha_{\alpha'}(h)$ and from now on we will write $\sqrt{g}^\alpha_{\alpha'} := \sigma^\alpha_{\alpha'}$.

Combining (3.32) and (3.34) we end up with the following SDE

$$\begin{aligned} \mathrm{d}h^i_t = &\left(-\left(A^T\right)^i_\alpha g^\alpha_{\alpha'}(h_t) A^{\alpha'}_j \left(A^T\right)^j_{\alpha''} A^{\alpha''}_k h^k_t + \frac{\sqrt{N}}{\beta}\left(A^T\right)^i_\alpha \partial_{\alpha'} g^{\alpha'\alpha}(h_t)\right)\mathrm{d}t \\ &+ \left(A^T\right)^i_\alpha \sqrt{g}^\alpha_{\alpha'}(h_t)\sqrt{\frac{2N}{\beta}}\,\mathrm{d}W^{\alpha'}_t. \end{aligned} \tag{3.35}$$

Introducing the abbreviations $G^{-1}(h) := (g^\alpha_{\alpha'}(h))^\alpha_{\alpha'}$ and $\sqrt{G}^{-1}(h) := \left(\sqrt{g}^\alpha_{\alpha'}(h)\right)^\alpha_{\alpha'}$ as well as the (rescaled) divergence-operator in α-coordinates

$$\left(\overline{D}\cdot\Sigma\right)^\alpha := \frac{1}{\sqrt{N}}\partial_{\alpha'}\Sigma^{\alpha'\alpha}$$

for some matrix field $\Sigma = \left(\Sigma^{\alpha'\alpha}\right)^{\alpha'\alpha}$ we see that (3.35) can be written in matrix form as

$$\mathrm{d}h_t = \left(-A^T G^{-1}(h_t) AA^T A h_t + \frac{N}{\beta} A^T \overline{D}\cdot G^{-1}(h_t)\right)\mathrm{d}t + A^T\sqrt{G}^{-1}(h_t)\sqrt{\frac{2N}{\beta}}\,\mathrm{d}W_t. \tag{3.36}$$

The following table provides the connection to the continuum case:

discrete	continuum
$G^{-1}(h)$	$M(h)$, see (3.27)
$\sqrt{G}^{-1}(h)$	$\sqrt{M(h)}$
A	∂_x, see Definition 3.4
A^T	$-\partial_x$
$\sqrt{N}\frac{dW_t}{dt}$	ξ.

For the last claim let $f_1(t), \dots, f_N(t)$ be compactly supported. A quick computation shows that

$$\mathbb{E}\left[\left(\int_0^\infty \frac{1}{N} f_\alpha(t)\sqrt{N}\frac{\mathrm{d}W^\alpha}{\mathrm{d}t}\,\mathrm{d}t\right)^2\right] = \frac{1}{N}\mathbb{E}\left[\left(\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t} f_\alpha(t) W_t^\alpha\,\mathrm{d}t\right)^2\right] = \frac{1}{N}\sum_{i=1}^N \int_0^\infty f_i^2(t)\,\mathrm{d}t.$$

Thus we obtain the following continuum analogs of (3.36):

discrete	continuum
$A^T G^{-1}(h) A A^T A h$	$\partial_x(M(h)\partial_x^3 h)$
$A^T \sqrt{G}^{-1}(h)\sqrt{\frac{2N}{\beta}}\frac{\mathrm{d}W_t}{\mathrm{d}t}$	$\partial_x\left(\sqrt{M(h)}\sqrt{\frac{2}{\beta}}\xi\right)$.

This confirms the form (3.13) of the SPDE. We will comment on the continuum form of the Itô correction term (3.36) in (3.12).

We will now turn our discussion to the entropy S. Recall that s is chosen such that $s'' = \frac{1}{M}$. For $h \in \Delta_N$ we write $s(h) := \left(s(h^i)\right)^i$. The choice of the metric tensor (3.1) is based on the fact that it satisfies the crucial identity

$$G^{-1}(h) A s'(h) = Ah \tag{3.37}$$

which is the discrete analog of

$$M(h)\partial_x s'(h) = \partial_x h.$$

By formally letting $\beta \to \infty$ in (3.36), we recover the *Grün–Rumpf* discretization of the deterministic thin-film equation

$$\frac{\mathrm{d}}{\mathrm{d}t} h_t = -A^T G^{-1}(h_t) A A^T A h_t. \tag{3.38}$$

In [**54**] the authors have used the identity (3.37) to show the following entropy estimate

Proposition 3.5. [**54**, p.129, Lemma 5.1] *Let h_t be a solution to* (3.38)*. We define the discrete entropy*[18] *via*

$$S(h) := \frac{1}{N}\sum_{i=1}^N s(h^i)$$

where s is chosen such that $s'' = \frac{1}{M}$. Then we have the identity

$$\frac{\mathrm{d}}{\mathrm{d}t} S(h_t) = -\frac{1}{N}\sum_{i=1}^N \left(\left(A^T A h_t\right)^i\right)^2.$$

Recall that we are particularly interested in the case $M(h) = h^3$.

[18] Notice that the discrete entropy is the lumped version of (3.8)

Assumption 3.6. We assume that for some $0 \leq m < \infty$ we have

$$L := \sup_{h \in (0,\infty)} \frac{M(h)}{h^m} < \infty. \tag{3.39}$$

From now on, for $m \geq 2$ we specifically set

$$s(h) := \int_h^\infty \int_{h'}^\infty \frac{1}{M(h'')} \,\mathrm{d}h''.$$

Using Proposition (3.5) it is easy to see that if the mobility satisfies assumption (3.6) the deterministic scheme preserves positivity for $m \geq 2$.

Remark 3.7. In terms of the configuration space positivity means that the flow h_t does not touch the boundary of the manifold $\mathcal{M}_N$ but stays in the open orthant $\{h > 0\}$. In fact, one can show that the distance (induced by the metric tensor (3.1)) between the boundary and the interior of $\mathcal{M}_N$ is finite if and only if $m < 3$, see (3.11) for the case of $N = 2$. Hence, by energy dissipation, any gradient flow with respect to the metric tensor (3.1) preserves positivity for $m \geq 3$. In case of the Dirichlet energy as the energy functional the entropy estimate (3.5) upgrades this threshold to $m \geq 2$.
On the other hand, it can be seen that the restriction of the metric tensor (3.3) to $T\mathcal{M}_N \otimes T\mathcal{M}_N$ induces a distance that is finite to the boundary iff $m < 5$.
The main result in this chapter transfers the entropy estimate (3.5) to the stochastic setting.

Theorem 3.8. *Let h_t be a solution to* (3.32) *such that the initial condition h_0 satisfies $\mathbb{E}[S(h_0)] < \infty$ and the mobility M satisfies Assumption 3.6 for $m \geq 3$, then the following identity holds*

$$\mathbb{E}[S(h_t)] + \int_0^t \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left(\left(A^T A h_r\right)^i\right)^2\right] \mathrm{d}r = \mathbb{E}[S(h_0)] + \frac{2N^3}{\beta} t. \tag{3.40}$$

Let $T > 0$. If, moreover, for $p < \infty$ we have that $\mathbb{E}[S^p(h_0)] < \infty$, then

$$\mathbb{E}\left[\left(\sup_{0\leq r\leq T} S(h_r)\right)^p\right]^{\frac{1}{p}} \leq \begin{cases} C\left(\left(\mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3 T}{\beta}\right)^{\frac{m-3}{m-2}} + \frac{N^{3+\frac{1}{m-2}} T}{\beta}\right)^{\frac{m-2}{m-3}} & \text{for } m > 3 \\ C\left(\mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + 1\right) e^{C\frac{N^4 T}{\beta}} & \text{for } m = 3 \end{cases}$$

for some constant C only depending on p, m and L.

Proof. By assumption, the process h_t satisfies the SDE

$$\mathrm{d}h_t = b(h_t)\,\mathrm{d}t + \sigma(h_t)\,\mathrm{d}W_t$$

where the drift b and the diffusion matrix σ are given according to (3.36). We set $\mathcal{M}_N^R := \{h \in \mathcal{M}_N : S(h) \leq R\}$ for some R. Notice that thanks to $m \geq 2$ it holds that

$h \in \mathcal{M}_N^R$ implies that h is strictly bounded away from 0. It is clear that there exist Lipschitz extensions $\overline{b}$ of b and $\overline{\sigma}$ of σ to all of $\mathbb{R}^{N+1}$ such that

$$\overline{b}|_{\mathcal{M}_N^R} = b|_{\mathcal{M}_N^R} \text{ and } \overline{\sigma}|_{\mathcal{M}_N^R} = \sigma|_{\mathcal{M}_N^R} \tag{3.41}$$

as well as a smooth extension $\overline{S}$ of the entropy S such that

$$\overline{S}|_{\mathcal{M}_N^R} = S|_{\mathcal{M}_N^R}. \tag{3.42}$$

Then we consider the process

$$\mathrm{d}\overline{h}_t = \overline{b}(\overline{h}_t)\,\mathrm{d}t + \overline{\sigma}(\overline{h}_t)\,\mathrm{d}W_t.$$

We apply Itô's formula (cf. [**103**, p.222, Theorem 3.3] and [**100**, p.67, Lemma 3.2]) to $\overline{S}(\overline{h}_t)$ which yields

$$\overline{S}(\overline{h}_t) = \overline{S}(h_0) + \int_0^t \overline{\mathcal{L}}\,\overline{S}(\overline{h}_s)\,\mathrm{d}s + \sqrt{\frac{2N}{\beta}} \int_0^t \partial_i \overline{S}(\overline{h}_s)\overline{\sigma}_\alpha^i(\overline{h}_s)\,\mathrm{d}W_s^\alpha \tag{3.43}$$

where $\overline{\mathcal{L}}$ denotes the generator of the process $\overline{h}_t$. Moreover, we define the stopping time

$$\tau_R := \inf\{t \geq 0 : S(h_t) > R\}.$$

By definition, we have that $\overline{h}_t = h_t$ for $t \leq \tau_R$ and thus by (3.43), (3.41) and (3.42) we get

$$\begin{aligned} S(h_{t\wedge\tau_R}) = S(h_0) &+ \int_0^{t\wedge\tau_R} \mathcal{L}S(h_s)\,\mathrm{d}s \\ &+ \sqrt{\frac{2N}{\beta}} \int_0^{t\wedge\tau_R} \partial_i S(h_s)\Big(A^T\sqrt{G}^{-1}(h_s)\Big)_\alpha^i \mathrm{d}W_s^\alpha. \end{aligned} \tag{3.44}$$

Here $\mathcal{L}$ denotes the generator of (3.36); according to (3.21), which we postprocess by (3.31), we have for any sufficiently nice function f

$$\mathcal{L}f = N\frac{1}{\beta}\partial_i\Big(\Big(A^TG^{-1}A\Big)^{ij}\partial_j f\Big) - N\partial_i f\Big(A^TG^{-1}A\Big)^{ij}\partial_j E.$$

Then, we compute using (3.33)

$$\begin{aligned} \mathcal{L}S(h) &\overset{(3.37)}{=} \frac{1}{\beta}\partial_i\Big(A^TAh\Big)^i - \frac{1}{N}\sum_{i=1}^N\Big(\Big(A^TAh\Big)^i\Big)^2 \\ &\overset{(3.4)}{=} \frac{2N^3}{\beta} - \frac{1}{N}\sum_{i=1}^N\Big(\Big(A^TAh\Big)^i\Big)^2. \end{aligned} \tag{3.45}$$

We now consider the martingale in (3.44)

$$X_t := \sqrt{\frac{2}{\beta N}}\int_0^{t\wedge\tau_R}\sum_{i=1}^N s'(h_{s\wedge\tau_R}^i)\Big(A^T\sqrt{G}^{-1}(h_{s\wedge\tau_R})\Big)_\alpha^i \mathrm{d}W_s^\alpha\,,$$

and note that, for $T > 0$ and $p < \infty$, the Burkholder–Davis–Gundy inequality (cf. [**103**, p.161, Corollary 4.2]) yields

$$\mathbb{E}\left[\sup_{0\le s\le t}|X_s|^p\right]^{\frac{1}{p}} \lesssim_p \mathbb{E}\left[\langle X_t\rangle^{\frac{p}{2}}\right]^{\frac{1}{p}},$$

where

$$\langle X_t\rangle = \frac{2}{\beta N}\int_0^{t\wedge\tau_R}\sum_{i=1}^N s'\left(h^i_{s\wedge\tau_R}\right)\left(A^TG^{-1}(h_{s\wedge\tau_R})A\right)^i_j s'\left(h^j_{s\wedge\tau_R}\right)\mathrm{d}s$$

is the quadratic variation of X. Here and from now on $\lesssim$ is equivalent to $\le C$ for some universal constant C that only depends on p, m, L. The integrand can be rewritten as follows

$$\begin{aligned}\sum_{i=1}^N s'\left(h^i_{s\wedge\tau_R}\right)\left(A^TG^{-1}(h_{s\wedge\tau_R})A\right)^i_j s'\left(h^j_{s\wedge\tau_R}\right) &\overset{(3.37)}{=} \sum_{\alpha=1}^N (Ah_{s\wedge\tau_R})^\alpha (As'(h_{s\wedge\tau_R}))^\alpha\\ &= \sum_{i=1}^N\left(A^TAh_{s\wedge\tau_R}\right)^i s'(h^i_{s\wedge\tau_R}).\end{aligned}$$

We estimate the second term in the above expression as

$$s'(h^i_{s\wedge\tau_R}) \overset{(3.39)}{\lesssim} \left(h^i_{s\wedge\tau_R}\right)^{1-m} \lesssim \left(\sum_{j=1}^N\left(h^j_{s\wedge\tau_R}\right)^{2-m}\right)^{\frac{m-1}{m-2}} \overset{(3.39)}{\lesssim} N^{\frac{m-1}{m-2}}S^{\frac{m-1}{m-2}}(h_{s\wedge\tau_R})$$

and hence by conservation of mass we arrive at

$$\begin{aligned}\sum_{i=1}^N s'\left(h^i_{s\wedge\tau_R}\right)\left(A^TG^{-1}(h_{s\wedge\tau_R})A\right)^i_j s'\left(h^j_{s\wedge\tau_R}\right) &\lesssim N^{\frac{m-1}{m-2}}S^{\frac{m-1}{m-2}}(h_{s\wedge\tau_R})\sum_{i=1}^N\left|\left(A^TAh_{s\wedge\tau_R}\right)^i\right|\\ &\lesssim N^{\frac{m-1}{m-2}+3}S^{\frac{m-1}{m-2}}(h_{s\wedge\tau_R}).\end{aligned}$$

Looking at (3.44) and collecting all the estimates yields

$$\begin{aligned}&\mathbb{E}\left[\left(\sup_{0\le s\le t} S(h_{s\wedge\tau_R})\right)^p\right]^{\frac{1}{p}}\\ &\lesssim \mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3t}{\beta} + \sqrt{\frac{N^{3+\frac{1}{m-2}}}{\beta}}\mathbb{E}\left[\left(\int_0^{t\wedge\tau_R} S^{\frac{m-1}{m-2}}(h_{s\wedge\tau_R})\,\mathrm{d}s\right)^{\frac{p}{2}}\right]^{\frac{1}{p}}\\ &\le \mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3t}{\beta} + \sqrt{\frac{N^{3+\frac{1}{m-2}}}{\beta}}\mathbb{E}\left[\left(\sup_{0\le s\le t} S(h_{s\wedge\tau_R})\int_0^t \sup_{0\le r\le s} S^{\frac{1}{m-2}}(h_{r\wedge\tau_R})\,\mathrm{d}s\right)^{\frac{p}{2}}\right]^{\frac{1}{p}}.\end{aligned}$$

Then, we use Young's inequality to the effect that

$$\mathbb{E}\left[\left(\sup_{0\le s\le t} S(h_{s\wedge\tau_R})\right)^p\right]^{\frac{1}{p}} \lesssim \mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3 t}{\beta} + \frac{N^{3+\frac{1}{m-2}}}{\beta}\mathbb{E}\left[\left(\int_0^t \sup_{0\le r\le s} S^{\frac{1}{m-2}}(h_{r\wedge\tau_R})\,\mathrm{d}s\right)^p\right]^{\frac{1}{p}}.$$

Finally, by using Minkowski's and Jensen's inequalities, we are left with

$$\mathbb{E}\left[\left(\sup_{0\le s\le t} S(h_{s\wedge\tau_R})\right)^p\right]^{\frac{1}{p}} \lesssim \mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3 t}{\beta} + \frac{N^{3+\frac{1}{m-2}}}{\beta}\int_0^t \mathbb{E}\left[\left(\sup_{0\le r\le s} S(h_{r\wedge\tau_R})\right)^p\right]^{\frac{1}{p(m-2)}}\mathrm{d}s. \tag{3.46}$$

By (3.17), for $m=3$ the integral inequality (3.46) yields

$$\mathbb{E}\left[\left(\sup_{0\le t\le T} S(h_{t\wedge\tau_R})\right)^p\right]^{\frac{1}{p}} \lesssim \left(\mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + 1\right)e^{C\frac{N^4 T}{\beta}} \tag{3.47}$$

for some constant C depending on m and L. On the other hand for $m>3$, by (3.17) we get

$$\mathbb{E}\left[\left(\sup_{0\le t\le T} S(h_{t\wedge\tau_R})\right)^p\right]^{\frac{1}{p}} \lesssim \left(\left(\mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3 T}{\beta}\right)^{\frac{m-3}{m-2}} + \frac{N^{3+\frac{1}{m-2}}T}{\beta}\right)^{\frac{m-2}{m-3}}. \tag{3.48}$$

We now argue that in the proof the stopping time was not necessary. By Chebyshev's inequality, we have

$$\mathbb{E}\left[\sup_{0\le t\le T\wedge\tau_R} S(h_t)\right] \ge R\,\mathbb{P}(\tau_R\le T)$$

and thus by envoking (3.47) respectively (3.48) we get

$$R\,\mathbb{P}(\tau_R\le T) \lesssim \begin{cases} (\mathbb{E}[S(h_0)]+1)e^{C\frac{N^4 T}{\beta}} & \text{for } m=3\\ \left(\left(\mathbb{E}[S^p(h_0)]^{\frac{1}{p}} + \frac{N^3 T}{\beta}\right)^{\frac{m-3}{m-2}} + \frac{N^{3+\frac{1}{m-2}}T}{\beta}\right)^{\frac{m-2}{m-3}} & \text{for } m>3. \end{cases}$$

Hence we have in either case

$$\lim_{R\to\infty}\mathbb{P}(\tau_R\le T) = 0 \tag{3.49}$$

and this proves the second assertion using Fatou's lemma. Finally, taking expectations in (3.44) and using (3.45) together with (3.49) gives the first assertion. □

As a direct consequence Theorem 3.8 yields

Corollary 3.9. *Let h_t be a solution to* (3.32) *such that the mobility $M(h)$ satisfies Assumption 3.6 for $m \geq 3$ and the initial datum satisfies $\mathbb{E}[S(h_0)] < \infty$. Then we have that*

$$\mathbb{P}(h > 0) = 1.$$

In particular, we do not have to impose the reflecting boundary condition for the SDE (3.32) if $m \geq 3$. The main selling point of our discretization is thus that we do not need to impose additional physics and/or rely on numerical tricks in the simulation in order to preserve positivity.

Remark 3.10. Although Theorem 3.8 yields positivity for $m \geq 3$ for every fixed $N \in \mathbb{N}$ the bound on the entropy grows with N. First of all, it is clear that (3.40) does not survive naively in the limit $N \to \infty$ since the term $\frac{2N^3 t}{\beta}$ will blow up. On the other hand, one can rearrange terms in the following way

$$\mathbb{E}[S(h_t)] - \mathbb{E}[S(h_0)] = \frac{2N^3}{\beta}t - \int_0^t \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left(\left(A^T A h_r\right)^i\right)^2\right] \mathrm{d}r. \tag{3.50}$$

The spatial increments of h_r behave like Brownian motion and hence the dissipation term $\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N \left(\left(A^T A h_r\right)^i\right)^2\right]$ scales like N^3 which shows that the scaling in N on the right hand side of (3.40) is natural and it is not unreasonable to expect that the right hand side of (3.50) converges for $N \to \infty$. On the other hand, at equilibrium the right hand side of (3.50) does not depend on the mobility but for $m \geq 5$ the left hand side is not finite in the continuum limit and thus we do not expect an equality like (3.40) to hold for $N \to \infty$.

Remark 3.11. We present an argument that the ranges $m < 3$ and $m \geq 3$ are qualitatively very different. To this end, for $N = 2$ we consider the associated Dirichlet form of the process, i.e. the right hand side of (3.18), namely

$$\int_0^2 \frac{1}{g(h)} \partial_h f(h) \partial_h \zeta(h) \, \mathrm{d}\nu(h)$$

where (cf. (3.83))

$$g(h) \sim h^{1-m}(2-h)^{1-m}.$$

We perform a change of variables $h \mapsto \hat{h}$ that is defined according to

$$\frac{\mathrm{d}\hat{h}}{\mathrm{d}h} = \sqrt{g}(h)$$

and we note that this yields the transformation

$$g^{-1}(h)\partial_h f(h)\partial_h \zeta(h) \to \partial_{\hat{h}} f(\hat{h}) \partial_{\hat{h}} \zeta(\hat{h}).$$

Then for $h \ll 1$ we have

$$\hat{h} \sim \begin{cases} \frac{2}{3-m} h^{\frac{3-m}{2}} & \text{for } m \neq 3 \\ \ln h & \text{for } m = 3. \end{cases}$$

For $2 - h \ll 1$ this holds similarly with $2 - h$ instead of h. Hence for $m < 3$ the configuration space for $\hat{h}$ is bounded and for $m \geq 3$ it is unbounded and therefore we do not need any boundary conditions. In fact, this heuristic is in the spirit of the Feller test (cf. [**70**, p.348, Theorem 5.29]) which also yields that the process touches the boundary of the configuration space for $m < 3$ and does not for $m \geq 3$. For this reason, the threshold $m = 3$ in Theorem 3.8 is sharp.

3.9. The central difference discretization

In this section we recall the finite-difference discretization used in [**37**] and compare it to the *Grün–Rumpf* discretization in the last section. We will argue that the finite-difference discretization has "touch–down" for any mobility $M(h)$, i.e. there is some $i = 1, \dots, N$ and some $t \geq 0$ such that $h_t^i = 0$.

By $C = \left(C_i^j\right)_i^j$ we denote the central difference matrix, i.e. we have for all vectors $(b^i)^i$

$$C_i^j b^i = N\left(b^{j+1} - b^{j-1}\right).$$

and, moreover, we let

$$G(h) := (g_{\alpha\alpha'}(h))_{\alpha\alpha'}, \quad g_{\alpha\alpha'}(h) := \frac{1}{M(h^\alpha)} \delta_{\alpha\alpha'}.$$

Then the finite-difference discretization of the SPDE (3.13) is the following SDE (cf. [**37**, p.591, (38)])

$$\mathrm{d}h_t = -C^T G^{-1}(h_t) C A^T A h_t \, \mathrm{d}t + C^T \sqrt{G}^{-1}(h_t) \sqrt{\frac{2N}{\beta}} \, \mathrm{d}W_t \tag{3.51}$$

which is supplemented with reflecting boundary conditions on $\partial\{h > 0\}$ and where the matrix A is given by (3.29). In [**37**, p.591-593] the authors check that the SDE (3.51) obeys the detailed balance condition which is largely due to the fact that

$$\sum_{j=1}^N \partial_j \left(C^T G^{-1}(h) C\right)_j^i = 0 \tag{3.52}$$

for all $h \in \mathbb{R}^N$. The term on the left hand side of (3.52) is reminiscent of the Itô–correction term emerging in (3.36). In particular, the equation (3.51) has the same invariant measure as (3.36); see also Section 3.11.2 for further numerical evidence on this.

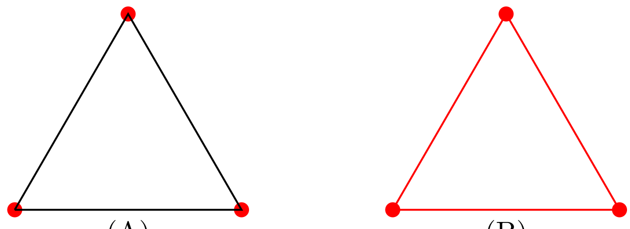


FIGURE 4. The configuration space $\mathcal{M}_3$ for the two discretizations: central difference on the left (A) and *Grün–Rumpf* on the right (B). The edges and corners where the diffusion matrix degenerates are colored in red. As can be seen from the figure, the central difference discretization does not degenerate orthogonal to the $d = 1$ codimension subsets of $\mathcal{M}_3$, while the *Grün–Rumpf* discretization degenerates on the whole boundary.

We will now give an argument that the process h_t defined by (3.51) touches down. The boundary $\partial\mathcal{M}_N$ can be decomposed into several sets of lower codimension. We call the sets of codimension 1 the faces of the simplex, i.e. the sets of the form $F_N^i := \overline{\mathcal{M}_N} \cap \{h^i = 0, h^j > 0, j \neq i\}$ for $i = 1, \dots, N$. Obviously, the hyperplane containing F_N^i is orthogonal to the unit vector e_i. Note that the quadratic variation of h_t^i is given by $\int_0^t \Big(C^T G^{-1}(h_t)C\Big)_{ii} \,\mathrm{d}t$. Then we see that the matrix $C^T G^{-1} C$ does not degenerate in the direction orthogonal to the faces since

$$\Big(C^T G^{-1}(h)C\Big)_{ii} = N^2\Big(M(h^{i-1}) + M(h^{i+1})\Big) > 0$$

for $h \in F_N^i$ and hence the quadratic variation stays positive even on F_N^i. This suggests that this discretization of the stochastic thin-film equation indeed features touch-down and we also observe this phenomenon numerically, see Section 3.11.3. Notice that on the other hand in case of the *Grün–Rumpf* discretization, the corresponding diffusion matrix $C^T G^{-1} C$ does degenerate in the direction orthogonal to the faces. We provide a small schematic for $N = 3$ in Fig. 4 to demonstrate these features of the two discretizations.

Remark 3.12 (The Itô-correction term). Consider the continuum stochastic thin-film equation in Stratonovich form with cut-off noise ξ^N (i.e. cutting off at the N^{th} Fourier mode):

$$\partial_t h = -\partial_x(M(h)\partial_x^3 h) + \sqrt{\frac{2}{\beta}}\partial_x\Big(\sqrt{M(h)} \circ \xi^N\Big).$$

It is fairly straightforward to check (cf. [**113**, Equation 2.5]) that the same SPDE can be written down in Itô form as follows

$$\partial_t h = -\partial_x(M(h)\partial_x^3 h) + \frac{N}{8\beta}\partial_x\left(\frac{(M'(h))^2}{M(h)}\partial_x h\right) + \sqrt{\frac{2}{\beta}}\partial_x\Big(\sqrt{M(h)}\xi^N\Big).$$

The above situation closely mimics the one in our scenario: We have presented two spatial discretizations of the thin-film equation with thermal noise and they differ from each other by the correction term

$$\frac{N}{\beta} A^T \overline{D} \cdot G^{-1}(h_t) \,.$$

The reader can convince themselves, that as N goes to ∞, the above expression formally converges to

$$-\frac{N}{\beta}\partial_x\Big((M(h))^2\partial_x\Big(\frac{M'(h)}{(M(h))^2}\Big)\Big) = \frac{N}{\beta}\partial_x\Bigg(\Big(2\frac{(M'(h))^2}{M(h)} - M''(h)\Big)\partial_x h\Bigg)^{19}.$$

For the case of power law mobilities $M(h) = h^m$, one can check that the two correction terms are the same, up to a multiplicative constant. This observation is consistent with the finding of [**62**] in which the authors discuss how different spatial discretizations of the stochastic Burgers equation can differ by terms which are analogous to the Itô-to-Stratonovich correction for SDEs. It would not be unreasonable to expect that such a term plays a role in renormalization as a possible counter term.

3.10. Touch-down for the continuum system

The open question of whether the deterministic thin-film equation with cubic mobility preserves positivity, is related to the degeneracy of the mobility when the film height approaches zero. In fact, in the case of high mobility exponent $m \geq \frac{7}{2}$, it has been shown that indeed strict positivity is preserved (cf. [**14**, p.194 , Theorem 4.1, (iii)]), while the opposite has been shown for $m < \frac{1}{2}$ in [**14**, p.198, Theorem 6.1].

In this section, we would like to discuss the same question (touch-down vs. positivity) for the continuum thin-film equation with thermal noise. We address this question through the associated large deviations rate functional of the continuum system. Before proceeding, we note that the entropic repulsion exhibited by the conservative Brownian excursion defined in Section 3.4.1 is a purely energetic phenomenon. As such, it is independent of the degeneracy of the mobility and is thus orthogonal to the discussion of touch-down which will be presented in this section.

There is a well-known connection between the large deviation principle for a microscopic reversible Markov process and the (appropriate) gradient flow structure of its mean-field limit (cf. [**31, 87**]). It is classical that for a reversible stochastic perturbation of a (finite-dimensional, but Riemannian) gradient flow, the rate functional I is given in terms of the metric tensor g and the energy function E (see, for example, [**45**, Chapter 4, Section

[19]for the specific case of $m = 3$ one can see this from the explicit form of the Itô-correction term provided in (3.85)

3, Theorem 3.1]): For a given time horizon $[0,T], T>0$, I_T is the following functional on the space of all paths $[0,T]\ni t\mapsto h_t\in\mathcal{M}$

$$\begin{aligned} I_T(h) :=&\frac{1}{2}\int_0^T g_{h_t}\left(\frac{\mathrm{d}h_t}{\mathrm{d}t}+\nabla E(h_t),\frac{\mathrm{d}h_t}{\mathrm{d}t}+\nabla E(h_t)\right)\mathrm{d}t\\ =&\frac{1}{2}\int_0^T g_{h_t}\left(\frac{\mathrm{d}h_t}{\mathrm{d}t},\frac{\mathrm{d}h_t}{\mathrm{d}t}\right)\mathrm{d}t+\frac{1}{2}\int_0^T g_{h_t}\left(\nabla E(h_t),\nabla E(h_t)\right)\mathrm{d}t\\ &+E(h_T)-E(h_0). \end{aligned} \tag{3.53}$$

Formally, (3.53) extends to infinite-dimensional situations like ours: While the SPDE might require a renormalization, the rate functional often does not (cf. [**66**]) – and can be analyzed rigorously (cf. [**74**]). We take this route in order to give a heuristic argument that touch-down is generic for power-law mobilities[20] $M(h)=h^m$ with mobility exponents $m<8$ and constitutes an extremely unlikely event for $m\geq 8$. To this end, we assume that the small-noise/high temperature large deviations rate functional I_T for (3.13) is given by (3.53) with g_{h_t} defined as in (3.3)[21], E given by the Dirichlet energy (3.2), and the gradient ∇E defined by duality as in (3.5). We first present our result for $m<8$, where we argue that touch-down is a generic phenomenon using an upper bound for the rate functional obtained via a self-similar ansatz.

Proposition 3.13. *Assume $M(h)=h^m$ for some $m<8$ and fix $T>0$. Then, there exists a curve $[-T,0]\ni t\mapsto h_t\in\mathcal{M}$ such that*

$$I_T(h)<\infty,\quad \min_{x\in\mathbb{R}} h_{-T}>0,\quad\text{and}\quad \min_{x\in\mathbb{R}} h_0=0\,.$$

Proof. For the sake of convenience, we present the proof only for the range $1<m<8$. For any curve $[-T,0]\ni t\mapsto h_t\in\mathcal{M}$, we can write the rate functional as follows

$$\begin{aligned} I_T(h) =&\frac{1}{2}\int_{-T}^0 g_{h_t}(\partial_t h_t,\partial_t h_t)\,\mathrm{d}t+\frac{1}{2}\int_{-T}^0 g_{h_t}(\nabla E(h_t),\nabla E(h_t))\,\mathrm{d}t\\ &+E(h_0)-E(h_{-T})\,. \end{aligned} \tag{3.54}$$

Note that we can apply Cauchy–Schwarz and Young's inequality to obtain the bound

$$\begin{aligned} |E(h_0)-E(h_{-T})| &= \left|\int_{-T}^0 g_{h_t}(\partial_t h_t,\nabla E(h_t))\,\mathrm{d}t\right|\\ \leq&\frac{1}{2}\int_{-T}^0 g_{h_t}(\partial_t h_t,\partial_t h_t)\,\mathrm{d}t+\frac{1}{2}\int_{-T}^0 g_{h_t}(\nabla E(h_t),\nabla E(h_t))\,\mathrm{d}t\,. \end{aligned} \tag{3.55}$$

[20] we consider power-law mobilities for convenience. One would expect the same result to hold with more general mobilities under the appropriate upper and lower bounds on the mobility.

[21] in the sequel, for the sake of simplicity, we will consider the metric g_h (and the equation) on $\mathbb{R}$. It can be defined in the natural way as in (3.3).

This leaves us with

$$I_T(h) \leq \int_{-T}^{0} g_{h_t}(\partial_t h_t, \partial_t h_t)\,\mathrm{d}t + \int_{-T}^{0} g_{h_t}(\nabla E(h_t), \nabla E(h_t))\,\mathrm{d}t.$$

We now consider the following self-similar ansatz

$$h_t(x) = (-t)^{\eta\gamma}\hat{h}(x(-t)^{-\eta}), \quad \hat{h}(\hat{x}) = (\hat{x}^2+1)^{\frac{\gamma}{2}},$$

with $\eta > 0$ and $0 < \gamma < 1$. Then,

$$\lim_{t\uparrow 0} h_t(x) = |x|^\gamma.$$

We thus have that

$$h_t(x) = (x^2 + (-t)^{2\eta})^{\frac{\gamma}{2}}\,.$$

Note now that, from the definition of the metric tensor (3.3),

$$\int_{-T}^{0} g_{h_t}(\partial_t h_t, \partial_t h_t)\,\mathrm{d}t = \int_{-T}^{0}\int_{\mathbb{R}} \frac{j_t^2}{h_t^m}\,\mathrm{d}x\,\mathrm{d}t\,,$$

where $j = j_t$ is a time-dependent flux field satisfying

$$\partial_t h_t + \partial_x j_t = 0\,.$$

It turns out that j_t also has a simple structure in self-similar variables. Indeed, it can be written as

$$j_t(x) = (-t)^{\eta\gamma+\eta-1}\hat{j}(x(-t)^{-\eta})\,,$$

where

$$\hat{j}(\hat{x}) = -\eta\gamma\int_0^{\hat{x}} (y^2+1)^{\frac{\gamma}{2}-1}\,\mathrm{d}y\,.$$

We then have that

$$\begin{aligned}\int_{-T}^{0} g_{h_t}(\partial_t h_t, \partial_t h_t)\,\mathrm{d}t &= \int_{-T}^{0} (-t)^{\eta\gamma(2-m)+2\eta-2}\int_{\mathbb{R}} \frac{\hat{j}^2(x(-t)^{-\eta})}{\hat{h}^m(x(-t)^{-\eta})}\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_{-T}^{0} (-t)^{\eta\gamma(2-m)+3\eta-2}\int_{\mathbb{R}} \frac{\hat{j}^2(\hat{x})}{\hat{h}^m(\hat{x})}\,\mathrm{d}\hat{x}\,\mathrm{d}t\,.\end{aligned}$$

For the integrability of the time-dependent term in the integrand we require that

$$\eta\gamma(2-m) + 3\eta > 1\,. \tag{3.56}$$

On the other hand, for the space-dependent term in the integrand we note that $|\hat{j}|(\hat{x}) \lesssim 1 + (\hat{x}^2+1)^{\frac{\gamma-1}{2}}$ and $\hat{h}(\hat{x}) = (\hat{x}^2+1)^{\frac{\gamma}{2}}$. It follows that for the integrability of this term it is sufficient to have

$$-m\gamma < -1\,. \tag{3.57}$$

We now turn our attention to the second term in (3.54). We compute

$$\partial_x^3 h_t = (-t)^{\eta(\gamma-3)}\hat{h}'''(x(-t)^{-\eta})\,.$$

Using the definition of the metric tensor (3.3) and of the gradient ∇E (3.5), we obtain

$$\begin{aligned}\int_{-T}^{0} g_{h_t}(\nabla E(h_t), \nabla E(h_t))\,\mathrm{d}t &= \int_{-T}^{0}\int_{\mathbb{R}} h_t^m(\partial_x^3 h_t)^2\,\mathrm{d}x\,\mathrm{d}t\\ &= \int_{-T}^{0}(-t)^{\eta\gamma(m+2)-6\eta}\int_{\mathbb{R}}\hat{h}^m(x(-t)^{-\eta})(\hat{h}'''(x(-t)^{-\eta}))^2\,\mathrm{d}x\,\mathrm{d}t\\ &= \int_{-T}^{0}(-t)^{\eta\gamma(m+2)-5\eta}\int_{\mathbb{R}}\hat{h}^m(\hat{x})(\hat{h}'''(\hat{x}))^2\,\mathrm{d}\hat{x}\,\mathrm{d}t\,.\end{aligned}$$

For the integrability of the time-dependent term in the above expression, it is sufficient to have

$$\eta(\gamma(m+2)-5) > -1\,. \tag{3.58}$$

On the other hand, note that $\left(\hat{h}^m(\hat{h}''')^2\right)(\hat{x}) \lesssim (\hat{x}^2+1)^{\frac{m\gamma}{2}+\gamma-3}$. Thus, for the integrability of the space-dependent term we require

$$m\gamma + 2\gamma - 6 < -1\,. \tag{3.59}$$

We first note that (3.57) can be reduced to

$$\frac{1}{m} < \gamma < 1\,,$$

if $1 < m < 8$. On the other hand, (3.59) is equivalent to the following condition

$$\gamma < \frac{5}{2+m}\,. \tag{3.60}$$

The remaining conditions (3.56) and (3.58) can be reformulated as

$$3-\gamma(m-2) > \frac{1}{\eta} > 5-\gamma(m+2)\,. \tag{3.61}$$

Note that if (3.60) is satisfied then $5-\gamma(m+2)$ is always larger than 0. On the other hand, $3-\gamma(m-2) > 5-\gamma(m+2)$ if and only if $\gamma > 1/2$. Thus, we can choose γ such that

$$\max\left(\frac{1}{2}, \frac{1}{m}\right) < \gamma < \min\left(1, \frac{5}{2+m}\right)\,,$$

for all $1 < m < 8$. We can then choose $\eta > 0$ so that (3.61) is satisfied. Thus, for these choices of η and h we have $I_T(h) < \infty$, and the result follows. □

We now turn to the case $m \geq 8$ where we argue that touch-down is an extremely rare event by obtaining an ansatz-free diverging (as $h \to 0$) lower bound for the rate functional. For simplicity, we restrict ourselves to paths $[0,T] \ni t \mapsto h_t$ that start at $h_0 \equiv 1$.

Proposition 3.14. *Assume $M(h) = h^m$ for some $m \geq 8$. Then, for any path $[0,T] \ni t \mapsto h_t \in \mathcal{M}$ starting from $h_0 \equiv 1$, the rate function I_T diverges in the following quantitative sense*

$$T^{\frac{1}{4}} I_T(h) \gtrsim \begin{cases} \sup_{x\in\mathbb{R}} \left(\ln \frac{1}{h_T} - 1 + h_T\right)_+ & m = 8 \\ \sup_{x\in\mathbb{R}} \left(\frac{1}{h_T^{\frac{m}{8}-1}} - 1\right)_+^2 & m > 8 \end{cases}, \tag{3.62}$$

as $\inf_{x\in\mathbb{R}} h_T \to 0$[22] *where the implicit constant in $\gtrsim$ depends only on m.*

Proof. We note first that the second identity in (3.53) yields the following inequality

$$E(h_t) \leq I_T(h) + E(h_0)\,, \tag{3.63}$$

for all $t \in [0,T]$. Note that in view of (3.3) we learn from (3.53) that there exists a time-dependent flux field $j = j_t(x)$ satisfying the continuity equation

$$\partial_t h_t + \partial_x j_t = 0\,, \tag{3.64}$$

such that the dissipation is controlled as

$$\frac{1}{2}\int_0^T \int_{\mathbb{R}} \frac{j_t^2}{h_t^m}\,\mathrm{d}x\,\mathrm{d}t \leq I_T(h) + E(h_0)\,. \tag{3.65}$$

We again monitor some "*entropy*" $\int_{\mathbb{R}} s(h_t)\,\mathrm{d}x$ along the path, where $s = s(h)$ is now defined via

$$\begin{gathered} s(1) = s'(1) = 0, \quad s(h) = 0 \text{ for } h \geq 1, \\ s''(h) = \frac{1}{h^{\frac{m}{2}}} \text{ for } h < 1. \end{gathered} \tag{3.66}$$

Since by (3.64)

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} s(h_t)\,\mathrm{d}x = \int_{\mathbb{R}} s''(h_t) j_t \partial_x h_t\,\mathrm{d}x\,,$$

we obtain from (3.66) and by Cauchy-Schwarz in the x-variable

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} s(h_t)\,\mathrm{d}x\right|^2 \leq \int_{\mathbb{R}} \frac{j_t^2}{h_t^m}\,\mathrm{d}x \int_{\mathbb{R}} (\partial_x h_t)^2\,\mathrm{d}x\,,$$

Thus, by (3.63) and (3.65),

$$\int_0^T \left|\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} s(h_t)\,\mathrm{d}x\right|^2 \mathrm{d}t \leq 2(I_T(h) + E(h_0))^2\,.$$

[22]although we present the result for $\mathbb{R}$ an essentially identical argument should also work for the torus

By integration and Cauchy-Schwarz in the t-variable, this yields

$$\frac{1}{\sqrt{2T}}\left|\int_{\mathbb{R}} s(h_T)\,\mathrm{d}x - \int_{\mathbb{R}} s(h_0)\,\mathrm{d}x\right| \le I_T(h) + E(h_0)\,.$$

Appealing once more to (3.63) this entails

$$\frac{1}{\sqrt{2T}}\int_{\mathbb{R}} s(h_T)\,\mathrm{d}x + E(h_T) \le \frac{1}{\sqrt{2T}}\int_{\mathbb{R}} s(h_0)\,\mathrm{d}x + 2E(h_0) + 2I_T(h)\,.$$

For our special initial data $h_0 \equiv 1$ and in view of (3.66), this simplifies to

$$\frac{1}{\sqrt{2T}}\int_{\mathbb{R}} s(h_T)\,\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}} (\partial_x h_T)^2\,\mathrm{d}x \le 2I_T(h)\,. \tag{3.67}$$

We note now that

$$s(h) \gtrsim \left(\frac{1}{h^{\frac{m}{4}-1}} - 1\right)_+^2.$$

Thus, (3.67) implies by Cauchy-Schwarz in the x-variable

$$\int_{\mathbb{R}} \left(\frac{1}{h_T^{\frac{m}{4}-1}} - 1\right)_+ |\partial_x h_T|\,\mathrm{d}x \lesssim T^{\frac{1}{4}} I_T(h)\,. \tag{3.68}$$

For $m = 8$, the left hand side of the above expression is equal to $\int_0^1 \left|\partial_x(\ln\frac{1}{h_T} + h_T - 1)_+\right|\mathrm{d}x$. Since the spatial average of h_T is equal to one, $(\ln\frac{1}{h_T} + h_T - 1)_+$ must vanish in at least one point. Thus, the left hand side of (3.68) controls $\sup_{x\in\mathbb{R}}(\ln\frac{1}{h_T} + h_T - 1)_+$. This establishes the first item in (3.62); the second item follows similarly. □

We conclude this section by showing that a curve with finite rate functional also has the expected regularity in time. While this is a priori unrelated to non-negativity of the film, we will see that we can use this scale-invariant regularity estimate to obtain a strengthening of Proposition 3.14 in Corollary 3.16, where the mobility exponent $m = 8$ again plays a special role.

Proposition 3.15. *Assume $M(h) = h^m$ for some $m \ge 0$ and consider a curve $[0,\infty) \ni t \mapsto h_t \in \mathcal{M}$ such that*

$$\overline{I}(h) := \frac{1}{2}\int_0^\infty g_{h_t}(\partial_t h_t, \partial_t h_t)\,\mathrm{d}t + \frac{1}{2}\int_0^\infty g_{h_t}(\nabla E(h_t), \nabla E(h_t))\,\mathrm{d}t + E(h_0) < \infty\,.$$

Then, h_t is locally Hölder continuous in time with exponent $\frac{1}{8}$. Furthermore, it satisfies the following scale-invariant estimate

$$|h_t(x) - h_s(y)| \lesssim \overline{I}^{\frac{1}{2}}(h)\Big(\min\{h_t(x), h_s(y)\}^{\frac{m}{8}}|t-s|^{\frac{1}{8}} + |x-y|^{\frac{1}{2}}\Big)\,,$$

for all $x, y \in \mathbb{R}$ and $|t-s| \ll \overline{I}^{-4}(h)\min\{h_t(x), h_s(y)\}^{8-m}$, where the implicit constants in $\lesssim$, $\ll$ depend only on m.

Proof. To start with, we consider the case where $[0,\infty) \ni t \mapsto h_t$ is such that $h_0(0) \le 1$ and $\overline{I}(h) \le 3$. Note that this along with (3.55) implies that

$$\sup_{t\in[0,\infty)} E(h_t) \le \overline{I}(h) \le 3\,, \tag{3.69}$$

which in turn implies that h_t is $\frac{1}{2}$-Hölder continuous in space for all $t \ge 0$ with the bound

$$|h_t(x) - h_t(y)| \lesssim \overline{I}^{\frac{1}{2}}(h)|x-y|^{\frac{1}{2}}\,. \tag{3.70}$$

We now fix a smooth compactly supported nonnegative function φ which is strictly positive in $(-1,1)$ and satisfies $\int_{\mathbb{R}} \varphi \,\mathrm{d}x = 1$ and $\varphi(x) \le 1$. We then define

$$F_t := \int_{\mathbb{R}} \varphi h_t \,\mathrm{d}x\,.$$

We then have

$$\frac{\mathrm{d}}{\mathrm{d}t} F_t = \int_{\mathbb{R}} \varphi' j_t \,\mathrm{d}x\,,$$

where $j_t = j_t(x)$ is a time-dependent flux field which solves

$$\partial_t h_t + \partial_x j_t = 0\,.$$

Dividing and multiplying by $h_t^{\frac{m}{2}}$ and then applying the Cauchy–Schwarz inequality in space, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} F_t \le \Big(\int_{\mathbb{R}} (\varphi')^2 h_t^m \,\mathrm{d}x\Big)^{\frac{1}{2}} a(t)$$

where

$$a(t) := \left(\int_{\mathbb{R}} \frac{j_t^2}{h_t^m} \,\mathrm{d}x\right)^{\frac{1}{2}}.$$

For the first term on the right hand side of the above expression, we have the following bound

$$\begin{aligned}\int_{\mathbb{R}} (\varphi')^2 h_t^m \,\mathrm{d}x \le& \sup_{x\in\mathbb{R}} (\varphi')^2 \Big|\int_{-1}^1 h_t^m \,\mathrm{d}x\Big| \\ \lesssim& \left|\int_{-1}^1 \left(\min_{x\in[-1,1]} h_t(x) + \int_{x_*}^x \partial_y h_t(y)\,\mathrm{d}y\right)^m \mathrm{d}x\right|,\end{aligned}$$

where $x_* = \mathrm{argmin}_{x\in[-1,1]} h_t(x)$. Using (3.69) and Jensen's inequality and the fact that φ is strictly positive in $(-1,1)$, we obtain

$$\begin{aligned}\int_{\mathbb{R}} (\varphi')^2 h_t^m \,\mathrm{d}x \lesssim& \left(\min_{x\in[-1,1]} h_t(x) + \left|\int_{-1}^1 \Big(\int_{x_*}^x \partial_y h_t(y)\,\mathrm{d}y\Big)\,\mathrm{d}x\right|\right)^m \\ \lesssim& (F_t+1)^m\,.\end{aligned}$$

This leaves us with

$$\frac{\mathrm{d}}{\mathrm{d}t}F_t \lesssim (1+F_t)^{\frac{m}{2}}a(t)\,.$$

We can now use the fact $\int_0^\infty a^2(t)\,\mathrm{d}t \le 2\overline{I}(h) \le 6$ along with the Cauchy–Schwarz and Young inequalities, to rewrite the above inequality as

$$F_t \lesssim 1 + F_0 + t + \int_0^t F_s^m\,\mathrm{d}s\,.$$

We thus obtain for $t \le 1$ (cf. Lemma 3.17)

$$F_t \lesssim \Big((1+t+F_0)^{1-m} + (1-m)t\Big)^{\frac{1}{1-m}}\,,$$

if $m \neq 1$ and

$$F_t \lesssim (1+F_0)e^{Ct}\,,$$

if $m = 1$ for some constant $C > 0$. In either of the two cases, we have that $F_t \le 3$ for all $0 < t \le t_*$ for some $t_* > 0$ depending on m, as long as F_0 is finite, which itself holds true since (3.69) and $h_0(0) \le 1$ imply

$$F_0 \le \int_{-1}^1 h_0\,\mathrm{d}x \lesssim 1\,.$$

We can then use Jensen's inequality and (3.69) to obtain

$$h_t(x) = \min_{x\in[-1,1]} h_t(x) + \int_{x_*}^x \partial_y h_t\,\mathrm{d}y \lesssim 1\,, \tag{3.71}$$

for all $0 < t \le t_*$ and $x \in [-1/2, 1/2]$.

By the shift-invariance[23] of $\overline{I}$, we may check the time regularity of h at some fixed point, say $x, t = 0$. Define $\varphi_\varepsilon(\cdot) := \varepsilon^{-1}\varphi(\varepsilon^{-1}\cdot)$. Then, for any $0 \le t \le t_*$, we can use (3.70) to obtain

$$|h_t(0) - h_0(0)| \lesssim \varepsilon^{\frac{1}{2}} + \left|\int_0^t \int_{\mathbb{R}} \varphi^\varepsilon \partial_s h_s\,\mathrm{d}x\,\mathrm{d}s\right|.$$

As before, we use the fact that h_t satisfies the continuity equation (3.64) with time-dependent flux field $j_t = j_t(x)$ to obtain

$$|h_t(0) - h_0(0)| \lesssim \varepsilon^{\frac{1}{2}} + \left|\int_0^t \int_{\mathbb{R}} \varepsilon^{-2}\varphi'(x/\varepsilon) j_s\,\mathrm{d}x\,\mathrm{d}s\right|.$$

Dividing and multiplying by $h_t^{\frac{m}{2}}$ as before and applying the Cauchy–Schwarz and Young inequalities, we obtain

$$|h_t(0) - h_0(0)| \lesssim \varepsilon^{\frac{1}{2}} + \varepsilon^{-\frac{9}{2}} \int_0^t \int_{-\varepsilon}^{\varepsilon} (\varphi'(x/\varepsilon))^2 h_s^m\,\mathrm{d}x\,\mathrm{d}s + \varepsilon^{\frac{1}{2}} \int_0^t \int_{\mathbb{R}} \frac{j_s^2}{h_s^m}\,\mathrm{d}x\,\mathrm{d}s\,. \tag{3.72}$$

---

[23] $\overline{I}$ is not truly shift invariant, but we simply use the fact that $\overline{I}(\tau_{y,s}h) \le 2\overline{I}(h)$ with $\tau_{y,s}h_t = h_{t+s}(\cdot + x)$

For the second term on the right hand side of the above expression, we rescale in x and use (3.71), to obtain

$$\varepsilon^{-\frac{9}{2}} \int_0^t \int_{-\varepsilon}^{\varepsilon} (\varphi'(x/\varepsilon))^2 h_s^m \,\mathrm{d}x\,\mathrm{d}s \lesssim \varepsilon^{-\frac{7}{2}} t\,.$$

For the third term on the right hand side of (3.72) we simply apply the bound (3.65) and use the fact that the $\overline{I}(h)$ is bounded to arrive at

$$\varepsilon^{\frac{1}{2}} \int_0^t \int_{\mathbb{R}} \frac{j_s^2}{h_s^m} \,\mathrm{d}x\,\mathrm{d}s \lesssim \varepsilon^{\frac{1}{2}}\,.$$

This leaves us with

$$|h_t(0) - h_0(0)| \lesssim \varepsilon^{\frac{1}{2}} + \varepsilon^{-\frac{7}{2}} t\,.$$

Choosing $\varepsilon = t^{\frac{1}{4}}$ and applying (3.70), we obtain

$$|h_t(x) - h_0(0)| \lesssim |t|^{\frac{1}{8}} + |x|^{\frac{1}{2}}\,,$$

for $(t,x) \in [0,t_*) \times \mathbb{R}$.

We can now rescale to recover the corresponding estimate for an arbitrary $[0,\infty) \ni t \mapsto h_t \in \mathcal{M}$ with $\overline{I}(h) < \infty$. To this end, we introduce

$$\hat{h}_{\hat{t}}(\hat{x}) = \lambda h_t(x)\,, \hat{x} = \mu x\,, \hat{t} = \nu t$$

for some $\lambda, \nu, \mu > 0$ to be chosen later. Under this choice of scaling, we have

$$E(h_t) = \int_{\mathbb{R}} (\partial_x h_t)^2 \,\mathrm{d}x = \mu\lambda^{-2} E(\hat{h}_{\hat{t}})\,,$$

and

$$\frac{1}{2}\int_0^\infty \int_{\mathbb{R}} \frac{j_t^2}{h_t^m} \,\mathrm{d}x\,\mathrm{d}t = \nu\mu^{-3}\lambda^{m-2} \frac{1}{2}\int_0^\infty \int_{\mathbb{R}} \frac{\hat{j}_{\hat{t}}^2}{\hat{h}_{\hat{t}}^m} \,\mathrm{d}\hat{x}\,\mathrm{d}\hat{t}\,,$$

where $j_t = j_t(x)$ is as before and $\hat{j}_{\hat{t}} = \hat{j}_{\hat{t}}(\hat{x})$ satisfies

$$\partial_{\hat{t}} \hat{h}_{\hat{t}} + \partial_{\hat{x}} \hat{j}_{\hat{t}} = 0\,.$$

Furthermore, the remaining term in $\overline{I}$ scales as

$$\frac{1}{2}\int_0^\infty \int_{\mathbb{R}} \left(\partial_x^3 h_t\right)^2 h_t^m \,\mathrm{d}x\,\mathrm{d}t = \lambda^{-m-2}\mu^5\nu^{-1} \frac{1}{2}\int_0^\infty \int_{\mathbb{R}} \left(\partial_{\hat{x}}^3 \hat{h}_{\hat{t}}\right)^2 \hat{h}_{\hat{t}}^m \,\mathrm{d}\hat{x}\,\mathrm{d}\hat{t}\,.$$

Since we may assume, without loss of generality, that $h_0(0) > 0$, we make the following choices

$$\lambda = \frac{1}{h_0(0)}\,, \mu = \lambda^2 \overline{I}(h)\,, \nu = \mu^3 \lambda^{2-m} \overline{I}(h)\,.$$

It follows that $\overline{I}(\hat{h}) \leq 3$, and $\hat{h}_0(0) = 1$. We thus have

$$\begin{aligned}|h_t(x) - h_0(0)| =&\lambda^{-1}|\hat{h}_{\hat{t}}(\hat{x}) - 1| \lesssim \lambda^{-1}\Big(\nu^{\frac{1}{8}}|t|^{\frac{1}{8}} + \mu^{\frac{1}{2}}|x|^{\frac{1}{2}}\Big)\\ \lesssim& h_0(0)\Big(\overline{I}^{\frac{1}{2}}(h)h_0^{\frac{m-8}{8}}(0)|t|^{\frac{1}{8}} + h_0^{-1}(0)\overline{I}^{\frac{1}{2}}(h)|x|^{\frac{1}{2}}\Big)\,,\end{aligned}$$

for all $0 \leq t \leq \overline{I}^{-4}(h)h_0^{8-m}(0)t_*$ and $x \in \mathbb{R}$. ☐

Corollary 3.16. *Let $m \geq 8$ and let $t \mapsto h_t \in \mathcal{M}$ satisfy $\overline{I}(h) < \infty$. Assume that, for some $x \in \mathbb{R}$, h_0 is almost touching down, i.e. $h_0(x) \ll 1$. Then, for all $t \geq 0$ such that $h_t(x) = 1$ it holds that*

$$t \gtrsim \begin{cases}\overline{I}^{-4}(h)h_0^{8-m}(x) & \text{for}\, m > 8\\ \overline{I}^{-4}(h)\ln(h_0^{-1}(x)) & \text{for}\, m = 8\,,\end{cases}$$

where the implicit constant in $\gtrsim$ depends only on m.

Proof. The dependence on x does not play any role in the proof since the argument we will present is pointwise in space. We will thus omit it for the rest of the proof. Moreover, we will set the implicit constants in $\lesssim$ in Proposition 3.15 to 1. By Proposition 3.15, we have for $0 \leq t \leq \overline{I}^{-4}(h)h_0^{8-m}$

$$|h_t - h_0| \leq \overline{I}^{\frac{1}{2}}(h)h_0^{\frac{m}{8}}t^{\frac{1}{8}} \leq h_0.$$

Then, we set $\tau_0 := 0$ and $\tau_1 := \overline{I}^{-4}(h)h_0^{8-m}$ and we observe that we have

$$h_{\tau_1} \leq 2h_0.$$

Inductively, we define $\tau_k := \tau_{k-1} + \overline{I}^{-4}(h)h_{\tau_{k-1}}^{8-m}$ for $k \in \mathbb{N}$. Then, it holds that

$$\tau_k = \overline{I}^{-4}(h)\sum_{i=0}^{k-1} h_{\tau_i}^{8-m}$$

as well as (using Proposition 3.15)

$$h_{\tau_i} \leq 2^i h_0. \tag{3.73}$$

Choosing $n := \lceil \log_2(h_0^{-1})\rceil$ we have $h_t \geq 1$ only if $t \geq \tau_n$. Note that if $m \geq 8$, we can apply (3.73) to obtain $h_{\tau_i}^{8-m} \geq 2^{i(8-m)}h_0^{8-m}$. This tells us that

$$\begin{aligned}\tau_n =&\overline{I}^{-4}(h)\sum_{i=0}^{n-1} h_{\tau_i}^{8-m}\\ \geq&\overline{I}^{-4}(h)\log_2(h_0^{-1})\,,\end{aligned}$$

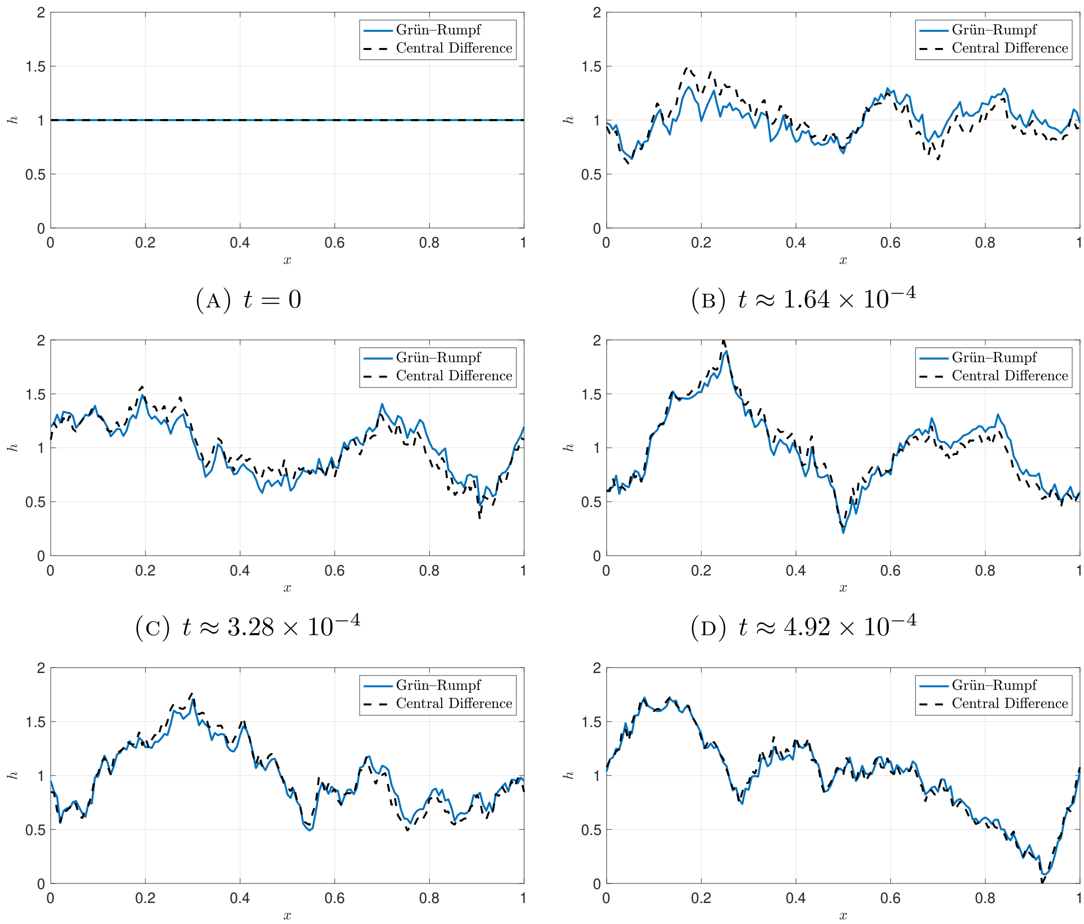
for $m = 8$. The case $m > 8$ can be derived in an essentially identical manner. ☐

3.11. Numerical experiments

3.11.1. Description of the time-stepping scheme. We describe here the time-stepping scheme for the SDE (3.36) with the *Grün–Rumpf* metric as described in Section 3.7. The central difference discretization (cf. Section 3.9) is treated in an identical manner. For our simulations, we rely on a semi-implicit Euler–Maruyama method which treats the noise, Itô-correction term, and metric tensor in (3.36) explicitly but treats the rest of the drift in an implicit manner. With $\Delta t > 0$ denoting the time step, the scheme can be described as follows

$$\begin{cases} h_0 &= h \in \mathcal{M}_N \\ h_{k+1} &= \Big(\mathrm{Id} + \Delta t A^T G^{-1}(h_k) A A^T A\Big)^{-1} \Big[h_k + \frac{\Delta t N}{\beta} A^T \overline{D} \cdot G^{-1}(h_k) \\ & \quad + \sqrt{\frac{2N\Delta t}{\beta}} A^T \sqrt{G}^{-1}(h_k) W_k\Big] \end{cases} \tag{3.74}$$

for all $k \in \mathbb{N}$, where h_k denotes the vector of film heights at the nodal points $(x_i)_i$ and at time $k\Delta t$ and $(W_k)_k$ is a sequence of independent $\mathcal{N}(0, I)$-distributed random vectors. We refer the reader to Section 3.C where we provide numerically stable expressions for the inverse metric and the Itô-correction term. For the specific choice of $M(h) = h^3$ the inverse metric $G^{-1}(h_k)$ is computed at each time step using (3.83) and the Itô-correction term $A^T \overline{D} \cdot G^{-1}(h_k)$ using (3.85). Since G^{-1} is a diagonal matrix, its square root can be computed explicitly. Due to the semi-implicit nature of the time-stepping scheme, in each step we have to compute the inverse of $\mathrm{Id} + \Delta t A^T G^{-1}(h_k) A A^T A$ which we do using the `MATLAB` function `mldivide`, which itself uses a Cholesky decomposition to perform the required matrix inversion.



(A) $t = 0$

(B) $t \approx 1.64 \times 10^{-4}$

(C) $t \approx 3.28 \times 10^{-4}$

(D) $t \approx 4.92 \times 10^{-4}$

(E) $t \approx 6.56 \times 10^{-4}$

(F) $t \approx 8.2 \times 10^{-4}$

FIGURE 5. Snapshots of the film height for the *Grün–Rumpf* and central difference discretizations at equally spaced time increments (time goes from (A) $\to$ (F)) for the same realization of the noise. As can be seen from the figures, the central difference discretization touches down (at $t_* \approx 8.2 \times 10^{-4}$, see (F)) while the *Grün–Rumpf* discretization stays away from the boundary. The simulations were performed with the following parameters: $N = 150$, $\Delta t = 10^{-10}$, $\beta = 1$, and $h_0 \equiv 1$.

3.11.2. Invariance of the measure ν_N. In this subsection, we perform some numerical experiments to check the invariance of the measure ν_N. We start by describing below a simple numerical procedure to sample from ν_N.

Algorithm 1: Sampling from ν_N

Result: Realization of ν_N

Sample discrete spatial white noise at temperature β^{-1}, i.e. a random N-dimensional vector of i.i.d. $\mathcal{N}(0, \beta^{-1} N \times \mathrm{Id})$-distributed random variables dW_N;

Project onto average zero vectors: $\mathrm{d}W_N^0 = \mathrm{d}W_N - N^{-1} \sum_i \mathrm{d}W_{N,i}$;

Integrate to get a discrete Brownian bridge: $W_{N,1}^0 = 0$, $W_{N,i}^0 = W_{N,i-1}^0 + N^{-1}\,\mathrm{d}W_{N,i-1}^0$;

Project onto average 1 vectors: $W_N = W_N^0 - N^{-1}\Big(\sum_i W_{N,i}^0\Big) + 1$;

if $\exists i$ *s.t.* $W_{N,i} < 0$ **then**

  reject;

else

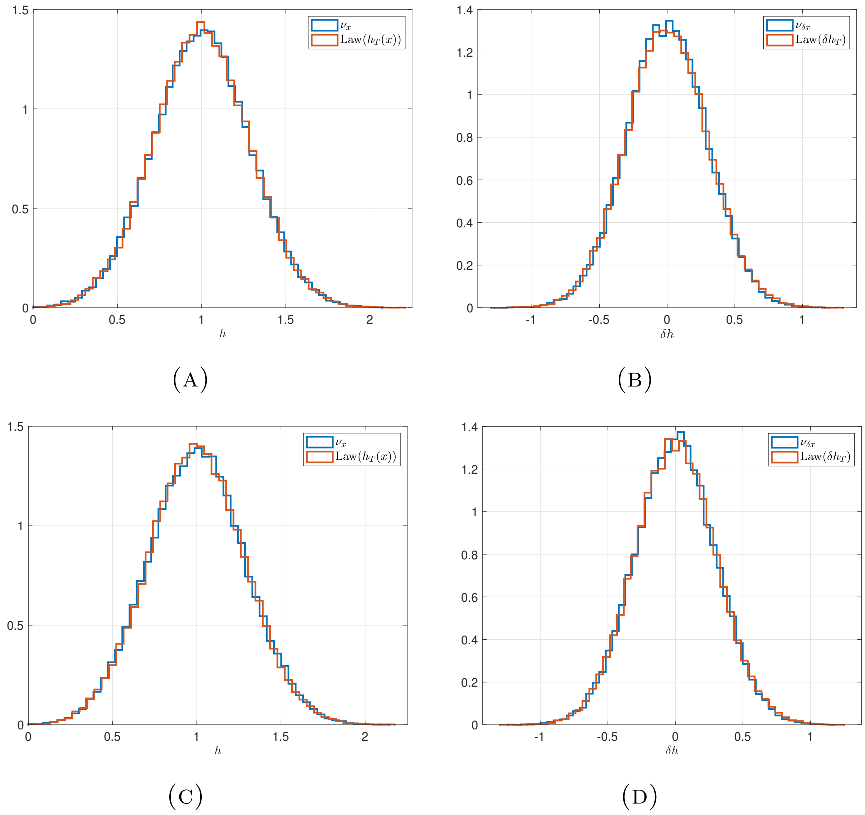
  accept;

end

We now integrate in time starting from $h_0 \equiv 1$ according to the semi-implicit Euler–Maruyama algorithm described in (3.74) up to some large time $T \gg \Delta t$. Repeating this procedure, we obtain a large number of samples, $M \gg 1$, of the process at time $t = T$ which we compare to the samples of ν_N generated by Algorithm 1. Note that T needs to be chosen to be larger than the typical relaxation time (to the invariant measure) of both discretizations. We found that $T = 10^{-3}$ works well for this purpose. We compare both the single-point distributions and the two-point correlations, i.e. the law of $\delta h_T = h_T(x + \delta x) - h_T(x)$ for some $\frac{1}{N} =: \Delta x \ll \delta x \ll 1$. Due to the stationarity (in space) of the invariant measure the choice of $x \in [0, 1]$ is irrelevant. We present the results of this experiment in Fig. 6.

3.11.3. Positivity, exit times, and entropic repulsion. As shown in Theorem 3.8, under appropriate conditions on the initial datum, the *Grün–Rumpf* discretization stays away from the boundary $\partial\mathcal{M}_N$. On the other hand, one expects (see the discussion in Section 3.9) the central difference discretization to touch the boundary with probability 1. We provide some numerical evidence for these features of the two discretizations in Fig. 5. Indeed, for the same realization of the noise, the *Grün–Rumpf* discretization stays away from 0, while the central difference discretization touches down.

We can provide stronger numerical evidence for the fact that the central difference discretization touches down by computing the mean exit time from $\mathcal{M}_N$ of the associated process. If this quantity is finite, this implies that the central difference discretization leaves $\mathcal{M}_N$, i.e. touches down, almost surely. Let $h_t^{h_0}$ be a solution of the central difference discretization of the stochastic thin-film equation (3.51) with initial condition



(A) (B)

(C) (D)

FIGURE 6. Plots of the histograms for $M = 1000$ samples of the single-point statistics and two-point correlations of the film height, i.e. h_T and $\delta h_T = h_T(x+\delta x)-h_T(x)$, for the *Grün–Rumpf* ((A),(B)) and the central difference ((C),(D)) discretizations compared to the reference measure, the conservative Brownian excursion ν_N. The simulations were carried out with the following parameters: $N = 50$, $\Delta t = 10^{-10}$, $\beta = 1$, $T = 10^{-3}$, $\delta x = 0.1$, and $h_0 \equiv 1$.

$h_0 \in \mathcal{M}_N$. Then, we define the exit time of $h_t^{h_0}$ from the interior to be

$$\tau(h_0) := \inf\Big\{t \geq 0 : h_t^{h_0} \notin \mathcal{M}_N\Big\}.$$

We take $h_0 \equiv 1$ and set $\tau := \tau(1)$. Then, we sample τ by running a Monte-Carlo simulation of (3.51) according to the time-stepping scheme described in (3.74). This time, instead of imposing reflecting boundary conditions, we stop the simulation as soon as we reach the boundary $\partial\mathcal{M}_N$, i.e. when the film touches down. Fig. 7 shows the behavior of the mean exit time as N grows. In particular, it seems that the mean exit time is finite and remains bounded as N tends to infinity. In the final part of this subsection, we study numerically the positivity properties of the continuum conservative Brownian

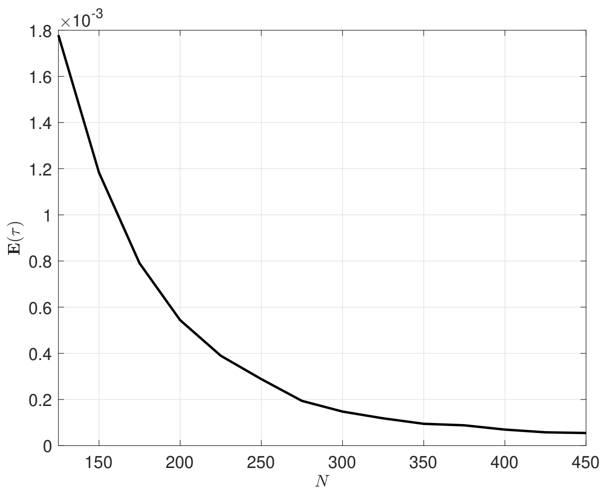


FIGURE 7. The dependence of the mean exit time of the central difference discretization on N. The simulations were performed with the following parameters: $\Delta t = 10^{-10}$, $\beta = 1$, $M = 100$, and $h_0 \equiv 1$.

excursion ν, i.e. its entropic repulsion. As has been mentioned before, our conservative Brownian excursion is qualitatively similar to the classical Brownian excursion from stochastic analysis. Moreover, it is known that the classical Brownian excursion features an entropic repulsion, in the sense that the single point distribution decays to 0 at 0. In fact, one can compute the single point statistics for the classical Brownian excursion $(Y_t)_{t\geq 0}$ explicitly (cf. [**103**, p.463]): For fixed $t \geq 0$ and $x, y > 0$ such that $Y_0 = x$ and $Y_T = y$ a.s., it takes the form

$$p_t^{x,y}(z) = \frac{T}{t(T-t)} z \frac{I_{\frac{1}{2}}(\frac{xz}{t}) I_{\frac{1}{2}}(\frac{zy}{T-t})}{I_{\frac{1}{2}}(\frac{xy}{T})} e^{-\frac{x^2+z^2}{2t}} e^{-\frac{z^2+y^2}{2(T-t)}} e^{\frac{x^2+y^2}{2T}}$$

where $I_{\frac{1}{2}}$ is the modified Bessel function of the first kind of order $\frac{1}{2}$. Notice that for $z \ll 1$, it holds that $I_{\frac{1}{2}}(z) \sim z^{\frac{1}{2}}$. From the above expression, it is clear that the distribution decays to 0 quadratically as $z \to 0$. In Fig. 8 we see that the single point distribution of our conservative Brownian excursion for $N \gg 1$ also exhibits quadratic decay at 0.

3.11.4. Convergence of the two discretizations. As mentioned earlier in the chapter, two different discretizations of a singular SPDE can converge to different limiting objects (cf. [**61**]). Thus, it would not be unreasonable to expect that the *Grün–Rumpf* and central difference discretizations of the thin-film equation with thermal noise have different continuum limits. However, numerical evidence seems to indicate that, at least started at equilibrium, the path space measures of the two discretizations converge to the same object.

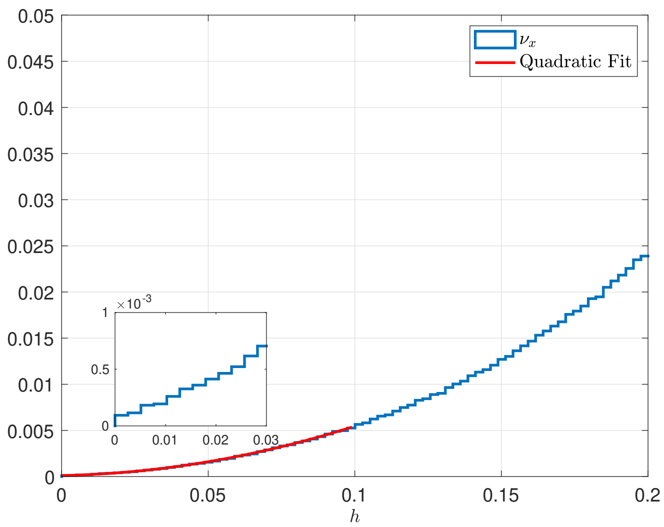


FIGURE 8. The entropic repulsion of the continuum conservative Brownian excursion ν as observed through the single point statistics of ν_N for N large ($= 2000$) obtained from $M = 2 \times 10^5$ samples. The single point distribution (in blue) decays quadratically as $h \to 0$ as can be seen by comparing it to the fitted curve (in red) $p(h) \approx 0.4704 \times h^2$. The zoomed-in version of the histogram exhibits the fact that entropic repulsion is a feature of the continuum invariant measure; for finite but large N the single point density is positive but small at 0.

We check this by sampling from ν_N using Algorithm 1 and then integrating in time with $h_0 \sim \nu_N$ to some final time T. Repeating this process, we obtain a large number, $M \gg 1$, of samples. We can then compute the two-point (in time) distributions of both discretizations, i.e. the joint law of h_t and $h_{t+\delta t}$ for some $\Delta t \ll \delta t \ll T$, for different values of N. One then observes that, as N increases, the two discretizations seem to converge to each other. Note that since we start our simulations at the invariant measure and the underlying process is reversible the choice of $t \geq 0$ is irrelevant. We present the results of these experiments in Figs. 9 and 10.

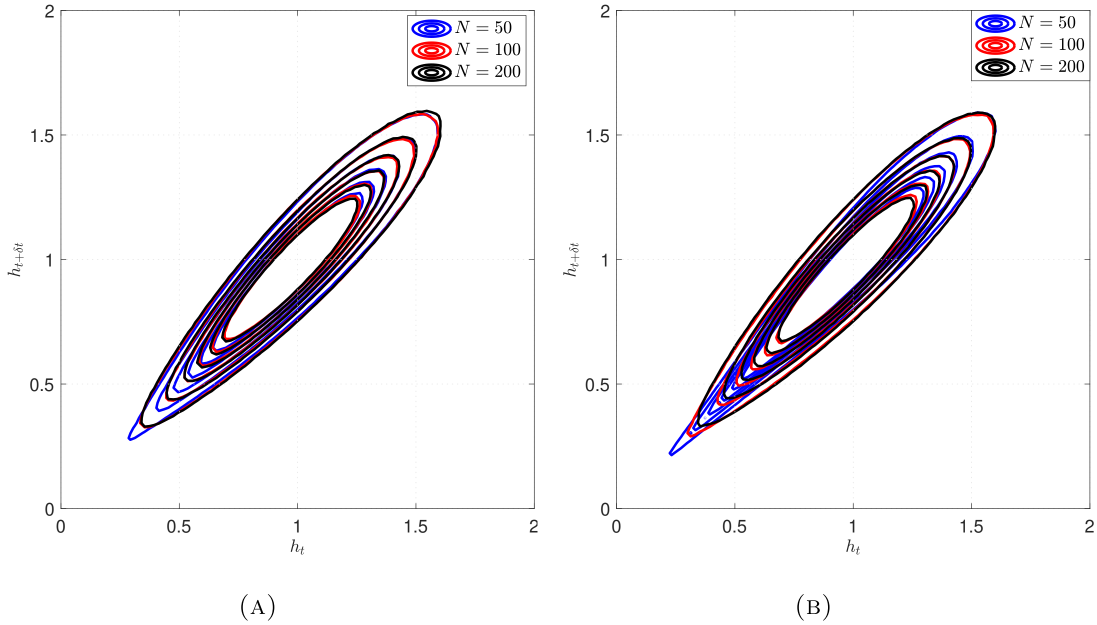
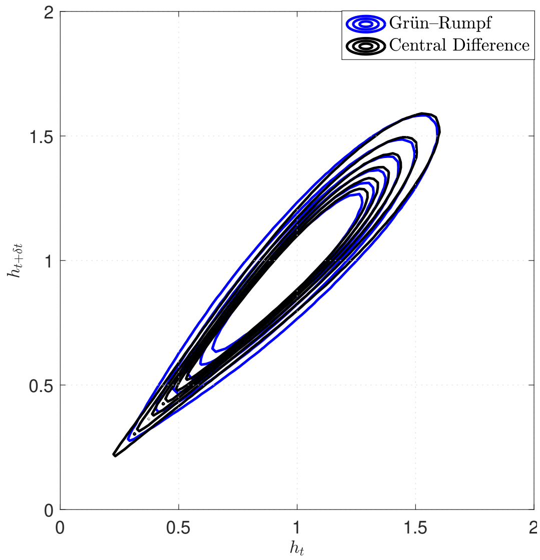
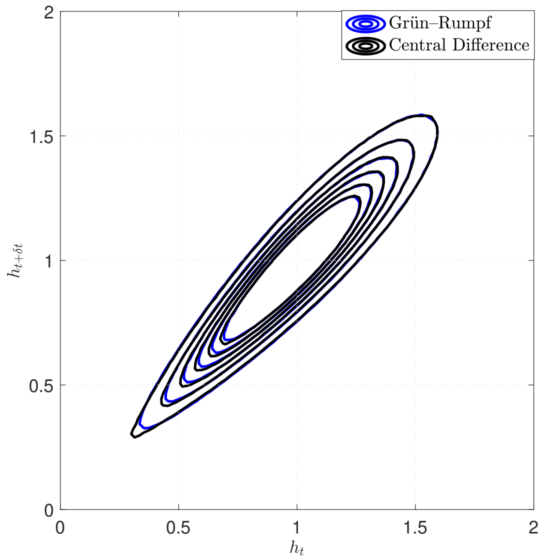


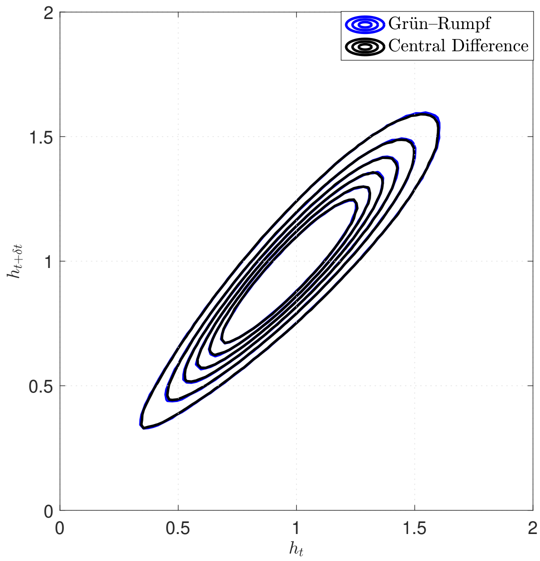
FIGURE 9. Level sets of the two-point (in time) distributions, i.e. the joint distributions of h_t and $h_{t+\delta t}$, for (A) the *Grün–Rumpf* and (B) the central difference discretizations for $N = 50, 100, 200$.



(A)



(B)



(C)

FIGURE 10. Comparisons of the level sets of the two-point (in time) distributions of the the *Grün–Rumpf* and the central difference discretizations for (A) $N = 50$, (B) $N = 100$, and (C) $N = 200$.

Appendix

3.A. The thin-film equation with linear mobility in Lagrangian coordinates

Let

$$z = \int_0^{X(z)} h(x)\,\mathrm{d}x \tag{3.75}$$

then taking the derivative twice with respect to z of (3.75) yields

$$1 = h(X(z))\frac{\mathrm{d}}{\mathrm{d}z}X(z) \tag{3.76}$$

as well as

$$0 = \partial_x h(X(z))\left(\frac{\mathrm{d}}{\mathrm{d}z}X(z)\right)^2 + h(X(z))\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z). \tag{3.77}$$

Multiplying (3.77) with $h(X(z))^2$ and invoking (3.76) we end up with

$$\partial_x h(X(z)) = -h(X(z))^3\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z). \tag{3.78}$$

Hence we compute for the Dirichlet energy

$$\begin{aligned}
E(h) &:= \frac{1}{2}\int_0^1 (\partial_x h)^2\,\mathrm{d}x = \frac{1}{2}\int_0^1 (\partial_x h(X(z)))^2\frac{\mathrm{d}}{\mathrm{d}z}X(z)\,\mathrm{d}z \\
&\overset{(3.78)}{=} \frac{1}{2}\int_0^1 \left(h(X(z))^3\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z)\right)^2\frac{\mathrm{d}}{\mathrm{d}z}X(z)\,\mathrm{d}z \\
&\overset{(3.76)}{=} \frac{1}{2}\int_0^1 \frac{\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z)\right)^2}{\left(\frac{\mathrm{d}}{\mathrm{d}z}X(z)\right)^5}\,\mathrm{d}z \\
&=: E(X).
\end{aligned}$$

Moreover, for some δX we compute

$$\mathrm{diff}E|_X.\delta X = \frac{1}{2}\int_0^1 2\frac{\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z)}{\left(\frac{\mathrm{d}}{\mathrm{d}z}X(z)\right)^5}\frac{\mathrm{d}^2}{\mathrm{d}z^2}(\delta X(z)) - 5\frac{\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2}X(z)\right)^2}{\left(\frac{\mathrm{d}}{\mathrm{d}z}X(z)\right)^6}\frac{\mathrm{d}}{\mathrm{d}z}(\delta X(z))\,\mathrm{d}z$$

$$= \int_0^1 \left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} \left(\frac{\frac{\mathrm{d}^2}{\mathrm{d}z^2} X(z)}{\left(\frac{\mathrm{d}}{\mathrm{d}z} X(z)\right)^5} \right) + \frac{5}{2} \frac{\mathrm{d}}{\mathrm{d}z} \left(\frac{\left(\frac{\mathrm{d}^2}{\mathrm{d}z^2} X(z)\right)^2}{\left(\frac{\mathrm{d}}{\mathrm{d}z} X(z)\right)^6} \right) \right) \delta X(z) \,\mathrm{d}z.$$

This, as usual, gives rise to the L^2-gradient flow

$$\begin{aligned} \partial_t X &= -\partial_z^2 \left(\frac{\partial_z^2 X}{(\partial_z X)^5} \right) - \frac{5}{2} \partial_z \left(\frac{(\partial_z^2 X)^2}{(\partial_z X)^6} \right) \\ &= \frac{1}{4} \partial_z^3 (\partial_z X)^{-4} - \frac{5}{8} \partial_z \Big(\partial_z (\partial_z X)^{-2} \Big)^2. \end{aligned}$$

3.B. Computing the change of coordinates

3.B.1. The dual metric in coordinates. Let the setting be as in the beginning of Section 3.6. As usual, we define the musical isomorphism via

$$T^*\mathcal{M} \to T\mathcal{M}, \omega \to \omega^\sharp$$

where

$$\omega.\dot{h} = g\Big(\omega^\sharp, \dot{h}\Big)$$

for all $\dot{h} \in T\mathcal{M}$. This gives rise to the dual metric g' on $T^*\mathcal{M} \otimes T^*\mathcal{M}$ via

$$g'(\omega, \omega') := g\Big(\omega^\sharp, \omega'^\sharp\Big) \tag{3.79}$$

for all $\omega, \omega' \in T^*\mathcal{M}$. Let $g'^{\alpha\alpha'}$ and $g_{\alpha\alpha'}$ be the representation of g' respectively g in the coordinates $(\varphi^\alpha)_\alpha$ and let ℓ, ℓ' be covectors and τ, τ' be vectors that are related by

$$\ell_\alpha = g_{\alpha\alpha'} \tau^{\alpha'}, \quad \ell'_\alpha = g_{\alpha\alpha'} \tau'^{\alpha'}. \tag{3.80}$$

Then by definition of (3.79) and by (3.80), we have

$$g_{\alpha\alpha'} \tau^\alpha \tau'^{\alpha'} = g'^{\alpha\alpha'} \ell_\alpha \ell'_{\alpha'}$$

and thus we see that $g'^{\alpha\alpha''} g_{\alpha''\alpha'} = \delta_{\alpha\alpha'}$ such that finally

$$g'^{\alpha\alpha'} = g^{\alpha\alpha'}. \tag{3.81}$$

Moreover, by (3.79) and (3.81), we see that for ζ, ζ' sufficiently smooth functions on $\mathcal{M}$ we have

$$g(\nabla\zeta, \nabla\zeta') = g'(\mathrm{diff}\zeta, \mathrm{diff}\zeta') = g^{\alpha\alpha'} \partial_\alpha \zeta \partial_{\alpha'} \zeta'. \tag{3.82}$$

3.B.2. Explicit formulae for partial derivatives. For some function $f : \mathcal{M}_N \to \mathbb{R}$ we have

$$\partial_i f(h) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} f(h + \varepsilon \hat{\varphi}_i)$$

as well as

$$\partial_\alpha f(h) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} f(h + \varepsilon\overline{\varphi}_\alpha).$$

3.C. Computation of the numerical mobility

We restrict ourselves to mobility functions of the form $M(h) = h^m$. Then we compute

$$\begin{aligned}
g_{\alpha\alpha}(h) &= \frac{1}{m-1}\frac{(h^{\alpha-})^{1-m} - (h^{\alpha+})^{1-m}}{h^{\alpha+} - h^{\alpha-}} \\
&= \frac{1}{m-1}\frac{1}{h^{\alpha+} - h^{\alpha-}}\frac{(h^{\alpha+})^{m-1} - (h^{\alpha-})^{m-1}}{(h^{\alpha-})^{m-1}(h^{\alpha+})^{m-1}} \\
&= \frac{1}{m-1}\frac{1}{h^{\alpha+} - h^{\alpha-}}\frac{\sum_{k=1}^{\infty}\binom{m-1}{k}(h_{\alpha-})^{m-1-k}(h^{\alpha+} - h^{\alpha-})^k}{(h^{\alpha-})^{m-1}(h^{\alpha+})^{m-1}} \\
&= \frac{1}{m-1}\frac{\sum_{k=1}^{\infty}\binom{m-1}{k}(h^{\alpha-})^{m-1-k}(h^{\alpha+} - h^{\alpha-})^{k-1}}{(h^{\alpha-})^{m-1}(h^{\alpha+})^{m-1}}.
\end{aligned}$$

In particular, this yields for $m = 3$

$$g_{\alpha\alpha}(h) = \frac{1}{2}\frac{h^{\alpha-} + h^{\alpha+}}{(h^{\alpha-})^2(h^{\alpha+})^2}$$

and hence

$$g^{\alpha\alpha}(h) = 2\frac{(h^{\alpha-})^2(h^{\alpha+})^2}{h^{\alpha-} + h^{\alpha+}}. \tag{3.83}$$

Moreover, for the Itô-correction term we are left with computing

$$\partial_{\alpha'} g^{\alpha'\alpha}(h) = -g^{\gamma\gamma'}(h)\partial_\gamma g_{\gamma'\alpha'}(h) g^{\alpha'\alpha}(h)$$

and using (3.B.2) we compute the derivative of the metric tensor via

$$\partial_\gamma g_{\gamma'\alpha'}(h) = -\delta_{\gamma'\alpha'}\fint_{I_{\alpha'}} \frac{M'(h)}{M(h)^2}\overline{\varphi}_\gamma \,\mathrm{d}x.$$

By the diagonal structure of $g(h)$ it is enough to compute

$$\partial_\alpha g_{\alpha\alpha}(h) = N^{\frac{3}{2}}\frac{\frac{1}{M(h^{\alpha+})} + \frac{1}{M(h^{\alpha-})} - 2g_{\alpha\alpha}(h)}{h^{\alpha+} - h^{\alpha-}} \tag{3.84}$$

where we used integration by parts which in the case $m = 3$ yields

$$\begin{aligned}
\partial_\alpha g_{\alpha\alpha}(h) &= N^{\frac{3}{2}}\frac{(h^{\alpha-})^{-3} + (h^{\alpha+})^{-3} - \frac{h^{\alpha-}+h^{\alpha+}}{(h^{\alpha-})^2(h^{\alpha+})^2}}{h^{\alpha+} - h^{\alpha-}} \\
&= N^{\frac{3}{2}}\frac{(h^{\alpha-})^{-2}((h^{\alpha-})^{-1} - (h^{\alpha+})^{-1}) + (h^{\alpha+})^{-2}((h^{\alpha+})^{-1} - (h^{\alpha-})^{-1})}{h^{\alpha+} - h^{\alpha-}}
\end{aligned}$$

$$= N^{\frac{3}{2}}\left(\frac{1}{(h^{\alpha-})^3 h^{\alpha+}} - \frac{1}{(h^{\alpha+})^3 h^{\alpha-}}\right).$$

Hence, for $m = 3$, we have

$$\partial_{\alpha'} g^{\alpha'\alpha}(h) = N^{\frac{3}{2}} 4h^i h^{i+1} \frac{h^i - h^{i+1}}{h^i + h^{i+1}}. \tag{3.85}$$

3.D. An integral inequality

Lemma 3.17. *Let u(t) be positive and bounded for $t \in [0, T]$. Let $0 \le \gamma < \infty$. Then, if*

$$u(t) \le u(0) + Ct + C\int_0^t u^\gamma(s)\,\mathrm{d}s \tag{3.86}$$

for some constant C, we have for $\gamma = 1$

$$u(t) \le (u(0) + 1)e^{Ct}$$

and for $\gamma \neq 1$

$$u(t) \le \Big((u(0) + CT)^{1-\gamma} + (1-\gamma)Ct\Big)^{\frac{1}{1-\gamma}}.$$

Proof. For $\gamma = 1$ we note that we can write (3.86) as

$$u(t) + 1 \le u(0) + 1 + C\int_0^t u(s) + 1\,\mathrm{d}s$$

and then apply Gronwall's inequality to get the assertion.
If $\gamma < 1$ then we set $X(t) := \int_0^t u^\gamma(s)\,\mathrm{d}s$ and hence

$$\frac{\mathrm{d}}{\mathrm{d}t}X(t) = u^\gamma(t) \overset{(3.86)}{\le} (u(0) + Ct + CX(t))^\gamma$$

which implies

$$\frac{\mathrm{d}}{\mathrm{d}t}(u(0) + CT + CX(t)) \le C(u(0) + CT + CX(t))^\gamma. \tag{3.87}$$

The differential inequality (3.87) further yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(u(0) + CT + CX(t))^{1-\gamma} \le C_\gamma$$

for $C_\gamma := (1-\gamma)C$ and since $X(0) = 0$ we have by integrating that

$$(u(0) + CT + CX(t))^{1-\gamma} \le (u(0) + CT)^{1-\gamma} + C_\gamma t.$$

By taking the inverse and appealing again to the assumption (3.86) we get the desired estimate. □

CHAPTER 4

Gradient bound for the φ^4_2-model

In this chapter we consider the massive φ^4_2-model on a two-dimensional torus of fixed size $L > 0$. The main theorem proves a certain gradient bound for the Markov semigroup which for large enough mass $m > 0$ implies exponential contraction in a certain weak norm. The proof is based on energy estimates for the linearized equation together with a stopping time argument which is inspired by [**21**] and relies on the strong Markov property of the Gaussian noise. As a corollary, using the approach of Bakry and Émery, we can show a local spectral gap inequality for the Markov semigroup which by ergodicity implies a spectral gap inequality for the massive φ^4_2-measure for large enough mass.

This chapter is based on the article [**78**] which is joint work with Pavlos Tsatsoulis.

4.1. Introduction

We consider the dynamic φ^4_2-model on the torus $\mathbb{T}^2 = \mathbb{R}^2/L\mathbb{Z}^2$ of fixed size $L > 0$ given by

$$\begin{cases} (\partial_t - \Delta + m)u = -u^3 + 3\infty u + \sqrt{2}\xi \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{T}^2 \\ u\big|_{t=0} = f, \end{cases} \tag{4.1}$$

where $m > 0$ is a positive mass, ξ denotes space-time white noise and f is a suitable initial condition. The infinite counter term $+3\infty u$ on the r.h.s. of (4.1) is reminiscent of renormalization (see Section 4.2 below) since the SPDE is singular due to the roughness of ξ.

This model serves as a toy example in the stochastic quantization of Euclidean quantum field theories. As explained in Section 1.4, it describes the natural reversible dynamics of the φ^4_2-measure formally given by

$$\mathrm{d}\nu(u) = \frac{1}{Z}\exp\left\{-\int_{\mathbb{T}^2}\mathrm{d}x\left(\frac{1}{2}|\nabla u(x)|^2 + \frac{1}{4}|u(x)|^4 - \frac{3\infty}{2}|u(x)|^2\right)\right\}\mathrm{d}u. \tag{4.2}$$

The construction of (4.2) was one of the first achievements in quantum field theory and goes back to Nelson (cf. [**92**]). Alternatively, Parisi and Wu in [**99**] proposed the use of (4.1) in order to construct and sample via MCMC methods the measure (4.2). A first attempt to implement this approach was made by Da Prato and Debussche in [**25**]. Later, along with the development of regularity structures (cf. [**59**]) and paracontrolled

distributions (cf. [**58**]), (4.1) was studied extensively by many authors (cf. [**91, 105, 104, 111, 63, 90, 56, 57**]). These results justified rigorously the connection of the singular dynamics and the measure in the sense of Parisi and Wu.
In the current chapter we study the regularization properties of the Markov semigroup $(P_t)_{t\geq 0}$ associated to (4.1) (see (4.15) below for the definition) through gradient-type estimates. Gradient-type estimates of Markov semigroups are important in the study of functional inequalities, e.g. spectral gap (or infinite dimensional Poincaré) and log-Sobolev inequalities, and transportation inequalities (cf. [**72, 8, 9, 22**]). These estimates usually require some convexity assumption, see for example [**22**, Property (**H.C.K.**), p. 232 and p. 235]. In the case of (4.1) convexity is destroyed by the presence of the infinite counter term $-3\infty u$ and at first glance it is unclear whether any type of such estimates can be derived. The argument we present here allows us to prove the following gradient estimate for the semigroup $(P_t)_{t\geq 0}$.

Theorem 4.1. *Let $(P_t)_{t\geq 0}$ be the Markov semigroup associated to* (4.1) *and $\kappa \in (0,1)$. For every $q > 1$ and $\varepsilon < 1-\kappa$ there exists $m_* \equiv m_*(\varepsilon, q, L) > 0$ such that*

$$\|DP_tF(f)\|_{L^2_x} \leq C(t\wedge 1)^{-\frac{\kappa+\varepsilon}{2}} e^{-(m-m_*)t} \left(P_t\|DF\|^q_{H^{-\kappa}_x}(f)\right)^{\frac{1}{q}}, \tag{4.3}$$

for every cylindrical functional F, $t > 0$, $f \in C^{-\alpha_0}$ and an implicit constant $C \equiv C(\varepsilon, \kappa, q, L) < \infty$ which is uniform in f and m. In the case $\kappa = 0$ the estimate holds for $\varepsilon = 0$ and a universal constant C which is independent of L.

Replacing L^2_ω-norm on the r.h.s. by an L^1_ω-norm yields the strong gradient estimate [**9**, Theorem 3.2.4]. The main difference is that the strong gradient estimate implies the log-Sobolev inequality, while (4.3) implies the (weaker) spectral gap inequality (cf.[**22**, Section 1] and [**9**, Sections 4 and 5]). Note that in contrast to the classical literature here we insist on a gradient estimate where the r.h.s. depends on the $H^{-\kappa}_x$-norm, allowing for κ arbitrarily close to 1. This is almost in line with the behaviour of the Gaussian free field in dimension 2 where the *carré du champ* is given by the H^{-1}_x-inner product or, equivalently, its *Cameron-Martin space* is given by H^1_x. As an immediate consequence (4.3) implies exponential contraction for $m > m_*$ in the following sense,

$$\sup_{\|h\|_{L^2_x}\leq 1} \sup_{\|DF\|_{H^{-\kappa}_x}\leq 1} |P_tF(f+h) - P_tF(f)| \leq C(t\wedge 1)^{-\frac{\kappa}{2}-\varepsilon} e^{-(m-m_*)t}, \tag{4.4}$$

where the second supremum is taken over all cylindrical functionals F.

In recent years gradient-type estimates of the form (4.3) have seen a rise in popularity. Starting with the work of Bakry–Émery (cf. [**8**]) it has become a vast research topic to relate these estimates to lower bounds of the Ricci curvature of the associated manifold. Since the interpretation of the heat flow on a manifold as a formal gradient flow with respect to the entropy on the Wasserstein space (cf. [**95**]), the notion of displacement convexity of the entropy is also closely related to lower bounds of the Ricci curvature (cf. [**96**]). This relationship can be associated to exponential contraction of the heat flow with respect to the Wasserstein metric which in our case corresponds to (4.4). Indeed, in [**116**] it has been shown that in the finite-dimensional case all these notions are equivalent. In the infinite-dimensional setting we, for example, refer to [**38**]. A similar gradient estimate for the φ^4_2-model has been obtained in [**26**], using similar techniques. This estimate does not fit into the Bakry–Émery framework[1], but it has interesting implications for the Kolmogorov operator.

In order to prove (4.3) we study the linearized equation

$$\begin{cases} (\partial_t - \Delta + m) J^f_{0,t} h = -3(u^2 - \infty) J^f_{0,t} h \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{T}^2, \\ J^f_{0,t} h\big|_{t=0} = h, \end{cases} \tag{4.5}$$

for suitable initial condition h. In the absence of the counter term one easily obtains a contraction estimate for any $m > 0$ of the form

$$\| J^f_{0,t} h \|^2_{L^2_x} \leq e^{-2mt} \| h \|^2_{L^2_x}, \tag{4.6}$$

which in turn implies the strong gradient estimate (cf. [**71**, Lemma 2.1]) where the same dynamics are considered in the 1-dimensional setting on the whole space[2]. To deal with the counter term we appeal to the Da Prato–Debussche decomposition (see Section 4.2 below), understanding $u^2 - \infty$ as

$$u^2 - \infty = v^2 + 2v\dot{\mathsf{I}} + \ddot{\mathsf{V}} + c_{t,\infty}, \tag{4.7}$$

where $\dot{\mathsf{I}}$ is the solution to the stochastic heat equation (4.10) with zero initial data, $\ddot{\mathsf{V}}$ its second Wick power defined in (4.5) and $c_{t,\infty}$ the constant defined in (4.14)[3]. The idea is to treat the lower order terms in (4.7), namely $2v\dot{\mathsf{I}} + \ddot{\mathsf{V}} + c_{t,\infty}$, as drift terms and absorb them to the mass m. Due to the lack of the required exponential integrability, in order to obtain a meaningful gradient estimate we restart the noise every time the Wick powers exceed a certain barrier using a stopping time argument in the spirit of Cass–Litterer–Lions (cf. [**21**]) for rough differential equations (see Section 4.3.2 below). This argument allows us to bypass the problem of exponential integrability of the Wick powers. Instead, we need to study the exponential integrability of the counting process $N(t)$ of the number of restarts to reach time t which due to the strong Markov property has exponential tails

[1] it is not a L^2-based estimate

[2] Using a post-processing of (4.6) as in Proposition 4.11 below one can upgrade the L^2_x-estimate to an $H^{-\kappa}_x$-estimate for $\kappa \in [0, 1)$ in the case of the torus.

[3] The constant $c_{t,\infty}$ appears due to the fact that we insist on using Wick powers of $\dot{\mathsf{I}}$ which at time $t = 0$ vanish. This is just a technical convenience but not necessary in our approach.

(see Proposition 4.8 below). A crucial ingredient to our approach is the *"coming down from infinity"* property of v first obtained in [**111**, Proposition 3.7] (see also [**90, 89, 56**] for up-to-date results on "coming down from infinity"), which ensures that the estimates on $N(t)$ do not depend on the initial data f, therefore, covering uniformly the whole time interval $[0, \infty)$. As a result of the stopping time argument we prove the following L^2_x-estimate for every $p < \infty$,

$$\mathbb{E}\Big[\|J^f_{0,t}\|^p_{L^2_x \to L^2_x}\Big]^{\frac{1}{p}} \leq Ce^{-(m-m_*)t}$$

for some $m_* > 0$ and $C < \infty$ uniformly in f, see Proposition 4.9 . Using a simple post-processing we can upgrade the above estimate to

$$\mathbb{E}\Big[\|J^f_{0,t}\|^p_{L^2_x \to H^\kappa_x}\Big]^{\frac{1}{p}} \leq C(t \wedge 1)^{-\frac{\kappa+\varepsilon}{2}} e^{-(m-m_*)t}, \tag{4.8}$$

see Proposition 4.11.

As we already mentioned earlier, the motivation to study gradient-type estimates for Markov semigroups comes from applications on functional inequalities. As a consequence of (4.3) we derive a spectral gap inequality for the Markov semigroup $\{P_t\}_{t\geq 0}$ based on the celebrated method of Bakry–Émery. Due to the presence of the $H^{-\kappa}_x$-norm for κ arbitrarily close to 1 the carré du champ is almost optimal when compared to the small scale behaviour of the Gaussian free field in 2-dimensions on a torus of fixed size $L > 0$ (which plays the role of an infra-red cutoff).

Theorem 4.2. *Under the assumptions of Theorem 4.1 the following spectral gap inequality holds*

$$P_t F^2(f) - \big(P_t F(f)\big)^2 \leq C \int_0^t (s \wedge 1)^{-\kappa-\varepsilon} e^{-2(m-m_*)s} \,\mathrm{d}s \; P_t \|FG\|^2_{H^{-\kappa}_x}(f) \quad \nu\text{-a.s. in } f,$$

for every cylindrical functional F, $t > 0$ and implicit constant $C \equiv C(\varepsilon, \kappa, L) < \infty$ which is uniform in f and m. In the case $\kappa = 0$ the estimate holds for $\varepsilon = 0$ and a universal constant C which is independent of L.

Let us mention that a spectral gap-type inequality for the Markov semigroup generated by (4.1) has already been obtained in [**111**] in the total variational norm in $C^{-\alpha_0}$ based on a combination of the strong Feller property, a support theorem and the "coming down from infinity" property. The same holds in dimension 3 based on the results from [**64, 65, 90**]. Although the total variational norm is stronger than any Wasserstein metric, the results in [**111**] do not provide an estimate w.r.t. the L^2_x-derivative.

Using the ergodicity of P_t, see for example [**111**, Corollary 6.6], as a corollary we prove a spectral gap inequality for the φ^4_2-measure for large masses $m > m_*$.

Corollary 4.3. *Under the statement of Theorem 4.2 and the additional assumption $m > m_*$ the φ^4_2-measure satisfies the spectral gap inequality*

$$\mathbb{E}_\nu F^2 - \big(\mathbb{E}_\nu F\big)^2 \le C \frac{1}{(m-m_*)^{1-\kappa-\varepsilon} \wedge (m-m_*)} \mathbb{E}_\nu \|DF\|^2_{H_x^{-\kappa}}, \tag{4.9}$$

for every cylindrical functional F, where for $\kappa = 0$ the estimate holds for $\varepsilon = 0$.

Remark 4.4. We emphasize that in order to obtain (4.9) we need to choose m large enough and, in particular, $m > m_*$ to ensure that the spectral gap constant does not blow-up in the limit $t \nearrow \infty$. This is a technical restriction of the method presented here and it is rather unnatural in the case of the torus. On the other hand, such a condition would be natural in the whole plane regime, provided that the dependence of the implicit constant C and the mass m_* on L can be eliminated. As we already stated in Theorem 4.1, C does not depend on L for $\kappa = 0$ and it would be interesting to investigate whether the dependence of m_* on L can be eliminated as well to allow for a large scale analysis. At first sight this seems possible using suitable weighted norms (in the spirit of [**91, 56**]), but it is rather unclear whether one can derive meaningful estimates in this direction.

Spectral gap inequalities are a convenient tool which quantifies ergodicity. When it comes to applications beyond the study of long time behavior, they have been used in the context of stochastic homogenization (cf. [**51**]) to obtain stochastic estimates on the corrector. In a similar spirit, spectral gap inequalities can be used as a tool in deriving stochastic estimates in the context of singular SPDEs (cf. [**81**] and [**67**, Section 5] for a simpler example).

While completing this thesis, the relevant work [**12**] appeared, which derives log-Sobolev inequalities for the φ^4-measure in dimensions 2 and 3 with carré du champs given by the L^2_x-norm. More precisely, the authors study approximations of the measure with ultraviolet and infra-red cutoffs and derive lower and upper bounds on the log-Sobolev constant independent of the cutoffs. Their approach is based on the machinery developed in [**11**] in combination with correlation inequalities. Although these results are optimal in the large scale regime and they imply the spectral gap inequality, the techniques presented here are more appropriate in the small scale regime.

4.1.1. Notation. For $\beta \in \mathbb{R}$ we set $C^\beta := B^\beta_{\infty,\infty}(\mathbb{T}^2)$ and the corresponding norm is denoted by $\|\cdot\|_\beta$. The space of arbitrarily smooth functions is accordingly denoted by C^∞. We set $L^p_x := L^p(\mathbb{T}^2)$ and $\|\cdot\|_{L^p_x}$ for the corresponding norm. Similarly, we use the same notation for $L^2_{t,x} := L^2(\mathbb{R}_+ \times \mathbb{T}^2)$ and $H^\alpha_x := H^\alpha(\mathbb{T}^2)$. Note that we have $L^p_x = B^0_{p,p}(\mathbb{T}^2)$ and $H^\alpha_x = B^\alpha_{2,2}(\mathbb{T}^2)$. The space $\mathcal{F}C^\infty_b$ denotes all cylindrical functions, i.e. for a distribution u we have $F \in \mathcal{F}C^\infty_b$ if there exists $n \in \mathbb{N}$, $\overline{F} \in C^\infty_b(\mathbb{R}^n)$ and

$h_i \in C^\infty(\mathbb{T}^2)$ for $i = 1, \dots, n$ such that $F(u) = \overline{F}(u(h_1), \dots, u(h_n))$ where we write $u(h) := \int_{\mathbb{T}^2} u(x)h(x)\,\mathrm{d}x$ for the natural pairing. Moreover, $a \wedge b := \min\{a, b\}$.

4.2. General framework

Contrary to Chapter 2, and as is custom for the φ^4_d-model, we denote by $\<1>_{0,t}$ the solution to the stochastic heat equation

$$\begin{cases} (\partial_t - \Delta + m)\<1> = \sqrt{2}\xi \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{T}^2 \\ \<1>|_{t=0} = 0, \end{cases} \tag{4.10}$$

which is explicitly given by

$$\<1>_{0,t}(\varphi) = \sqrt{2}\xi\Big(\mathbb{1}_{[0,t)} H_{t-\cdot} * \varphi\Big)$$

for all sufficiently nice test functions $\varphi : \mathbb{T}^2 \to \mathbb{R}$ where $(t, x) \mapsto H_t(x)$ denotes the heat kernel associated with the operator $(\partial_t - \Delta + m)$. We also denote by $\<2>_{0,t}$ and $\<3>_{0,t}$ its second and third Wick powers defined as the limits

$$\<2>_{0,t} := \lim_{\delta \searrow 0} \Big(\<1>^{(\delta)}_{0,t}\Big)^2 - c^{(\delta)}_{0,t}, \quad \<3>_{0,t} := \lim_{\delta \searrow 0} \Big(\<1>^{(\delta)}_{0,t}\Big)^3 - 3c^{(\delta)}_{0,t}\<1>^{(\delta)}_{0,t}, \tag{4.11}$$

where $c^{(\delta)}_{0,t} = \mathbb{E}\Big(\<1>^{(\delta)}_{0,t}(0)\Big)^2$, δ denotes some space mollification and the convergence takes place in $C^{-\alpha}$ for every $\alpha > 0$. For simplicity, we write $\<k>_{0,t}$, $k = 1, 2, 3$, to denote the collection of $\<1>_{0,t}$, $\<2>_{0,t}$, $\<3>_{0,t}$. We are only interested in the analytical properties of the Wick powers $\<k>_{0,t}$, $k = 1, 2, 3$, given by the next proposition.

Proposition 4.5. *Let $T > 0$. For any $k = 1, 2, 3$, $\alpha > 0$ and $p < \infty$ we have*

$$\mathbb{E}\Bigg[\sup_{0 \le t \le T} \Big\|\<k>_{0,t}\Big\|^p_{-\alpha}\Bigg]^{\frac{1}{p}} \le C \tag{4.12}$$

where the constant $C \equiv C(L, T, \alpha, p)$ does not depend on m, vanishes for $T \searrow 0$ and grows at most polynomially in T.

We postpone the proof of this proposition in the appendix, Section 4.C, where we present an alternative argument in the spirit of [**67**, Section 5] and [**81**] using the fact that the white noise ξ satisfies a spectral gap inequality. Note that we stress the independence of the constant C on m, which allows us to ensure that m_* in Theorem 4.1 is independent of m (in particular, θ in Proposition 4.8 can be chosen independently of m).

We interpret the solution u of (4.1) using the Da Prato–Debussche decomposition (cf. [**25**]), namely, we define $u_{0,t} := \<1>_{0,t} + v_{0,t}$, where $v_{0,t}$ solves

$$\begin{cases} (\partial_t - \Delta + m)v_{0,t} = -v^3_{0,t} - 3v^2_t\<1>_{0,t} - 3v_{0,t}\<2>_{0,t} - \<3>_{0,t} + 3c_{t,\infty}\Big(\<1>_{0,t} + v_{0,t}\Big) \\ v|_{t=0} = f, \end{cases} \tag{4.13}$$

and $f \in C^{-\alpha_0}$ for $\alpha_0 > 0$ sufficiently small. Let us remark on the constant $c_{t,\infty}$ which appears on the r.h.s. of (4.13). This is due to the fact that we renormalize the Wick powers via time dependent constants in order for them to vanish at time $t = 0$, although renormalization on the level of the dynamics is done via a time-independent constant $c^{(\delta)}_{0,\infty}$ to ensure that the resulting Markov processes is homogeneous in time. In the limit $\delta \searrow 0$ the difference between the two constants leads to

$$c_{t,\infty} := 2 \int_t^\infty H_{2s}(0)\, \mathrm{d}s \lesssim t^{-\frac{\beta}{2}} \tag{4.14}$$

for every $\beta > 0$. A crucial ingredient that we use in the sequel is the "coming down from infinity" for the solution $v_{0,t}$ to (4.13) which we include in the appendix, Section 4.D. We refer the reader to [**25, 91, 111**] for details on the global well-posedness of (4.13).

For $t \geq s$ we also consider the restarted processes $\Psi^{k}_{s,t}$, $k = 1, 2, 3$ which are defined via the solution to

$$\begin{cases} (\partial_t - \Delta + m)\Psi_{s,t} = \sqrt{2}\xi \\ \Psi_{s,t}|_{t=s} = 0, \end{cases}$$

and respectively via (4.11) with $\Psi^{(\delta)}_{0,t}$ replaced by $\Psi^{(\delta)}_{s,t}$. Note that $\{\Psi^{k}_{0,t}\}_{t\geq 0}$ and $\{\Psi^{k}_{s,t}\}_{t\geq s}$ are equal in law and $\{\Psi^{k}_{s,t}\}_{t\geq s}$ is independent of $\{\Psi^{k}_{0,t}\}_{t\in[0,s]}$.

Similarly, we consider $v_{s,t}$ which is defined as the solution to

$$\begin{cases} (\partial_t - \Delta + m)v_{s,t} = -v_{s,t}^3 - 3v_{s,t}^2 \Psi_{s,t} - 3v_{s,t}\Psi^{2}_{s,t} - \Psi^{3}_{s,t} + 3c_{t-s,\infty}\Big(\Psi_{s,t} + v_{s,t}\Big) \\ v_{s,t}|_{t=s} = u_s. \end{cases}$$

Note that all pathwise and stochastic estimates for $\Psi^{k}_{0,t}$, $k = 1, 2, 3$, and $v_{0,t}$ extend to $\Psi^{k}_{s,t}$, $k = 1, 2, 3$, and $v_{s,t}$. Especially, due to the "coming down from infinity" property pathwise estimates on $v_{s,t}$ do not depend on u_s.

4.3. Strategy of the proof

In this section we want to give an outline of the proof of Theorem 4.1. By [**111**, Theorem 4.2] for $f \in C^{-\alpha_0}$ we know that $\left\{u^f_{0,t}\right\}_{t\geq 0}$ is a Markov process with $u^f_{0,t}|_{t=0} = f$. In particular, for $F \in \mathcal{F}C^\infty_b$ the operator

$$P_t F(f) := \mathbb{E}\Big[F(u^f_{0,t})\Big] \tag{4.15}$$

yields a one-parameter semigroup. We denote by D the L^2-derivative, i.e. we have

$$DF(u) = \sum_{i=1}^n \partial_i \overline{F}(u(h_1), \ldots, u(h_n))h_i.$$

The implicit function theorem implies that the map

$$f \mapsto v^f$$

is differentiable and for any $h \in C^\infty$ it holds that $J^f_{0,t}h := v'_{0,t}(f).h$ is a (mild) solution of the equation

$$\begin{cases} (\partial_t - \Delta + m)J^f_{0,t}h = -3\Big(\big(v^f_{0,t}\big)^2 + 2v^f_{0,t}\Psi_{0,t} + \Psi^2_{0,t}\Big)J^f_{0,t}h + 3c_{t,\infty}J^f_{0,t}h \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{T}^2 \\ J^f_{0,0}h = h. \end{cases}$$

For a proof we refer to Section 4.F.

By definition, we have $u^f_{0,t} = \Psi_{0,t} + v^f_{0,t}$ and since $\Psi_{0,t}$ does not depend on the initial condition we can conclude that also $f \mapsto u^f$ is differentiable, i.e. there exists $u'(f) = v'(f) : X \to Y$ (cf. Section 4.F for the definition of the function spaces X and Y) such that

$$u^{f+h} - u^f - u'(f).h = v^{f+h} - v^f - v'(f).h = o(\|h\|_{C^{-\alpha_0}}).$$

Thus we can compute for any $t \geq 0$ and $F \in \mathcal{F}C^\infty_b$ using a simple Taylor expansion

$$\begin{aligned} &P_tF(f+h) - P_tF(f) \\ &= \mathbb{E}\Big[F(u^{f+h}_{0,t}) - F(u^f_{0,t})\Big] \\ &= \mathbb{E}\Bigg[\sum_{i=1}^n \partial_i\overline{F}(u^f_{0,t}(h_1), \ldots, u^f_{0,t}(h_n))(u^{f+h}_{0,t}(h_i) - u^f_{0,t}(h_i))\Bigg] + o(\|h\|_{C^{-\alpha_0}}) \\ &= \mathbb{E}\Bigg[\sum_{i=1}^n \partial_i\overline{F}(u^f_{0,t}(h_1), \ldots, u^f_{0,t}(h_n))(v'_{0,t}(f).h, h_i)_{L^2_x}\Bigg] + o(\|h\|_{C^{-\alpha_0}}). \end{aligned}$$

This shows that $f \mapsto P_tF(f)$ is differentiable and we have

$$(P_tF)'(f).h = \mathbb{E}\Big[\Big(DF(u^f_{0,t}), v'_{0,t}(f).h\Big)_{L^2_x}\Big] = \mathbb{E}\Big[\Big(DF(u^f_{0,t}), J^f_{0,t}h\Big)_{L^2_x}\Big]. \tag{4.16}$$

Moreover, by Proposition 4.11 we see that $(P_tF)'(f) : L^2_x \to \mathbb{R}$ is a bounded linear functional[4] and thus there exists $DP_tF(f) \in L^2_x$ such that

$$(P_tF)'(f).h = \int_{\mathbb{T}^2} DP_tF(f)(x)h(x)\,\mathrm{d}x$$

and in particular

$$\|DP_tF(f)\|_{L^2_x} = \sup_{\|h\|_{L^2} \leq 1} \big|(P_tF)'(f).h\big|. \tag{4.17}$$

[4]more specifically the extended operator initially defined on the dense subspace C^∞

By (4.16), (4.17) and the Hölder's inequality in probability for any $\kappa \geq 0$

$$
\begin{aligned}
\|DP_tF(f)\|_{L^2_x} &= \sup_{h\in C^\infty, \|h\|_{L^2_x}\leq 1} |(P_tF)'(f).h| \\
&\leq \mathbb{E}\Big[\big\|DF(u^f_{0,t})\big\|^q_{H^{-\kappa}_x}\Big]^{\frac{1}{q}} \left(\sup_{h\in C^\infty, \|h\|_{L^2}\leq 1} \mathbb{E}\Big[\big\|J^f_{0,t}h\big\|^p_{H^\kappa_x}\Big]\right)^{\frac{1}{p}} \\
&= \Big(P_t\|DF\|^q_{H^{-\kappa}_x}(f)\Big)^{\frac{1}{q}} \left(\sup_{h\in C^\infty, \|h\|_{L^2_x}\leq 1} \mathbb{E}\Big[\big\|J^f_{0,t}h\big\|^p_{H^\kappa_x}\Big]\right)^{\frac{1}{p}} \qquad (4.18)
\end{aligned}
$$

where in the first line we used that C^∞ is dense in L^2_x. Hence, in order to prove Theorem 4.1 we have to estimate the quantity

$$
\left(\sup_{h\in C^\infty, \|h\|_{L^2_x}\leq 1} \mathbb{E}\Big[\big\|J^f_{0,t}h\big\|^p_{H^\kappa_x}\Big]\right)^{\frac{1}{p}} \qquad (4.19)
$$

uniformly in the initial condition $f \in C^{-\alpha_0}$. In order to do this, we will proceed in three steps. The first step is to prove an L^2_x-energy estimate with the drawback that the implicit constant is random and moreover it is not clear that it is integrable. The second step – which is our core argument – shows that this constant is indeed integrable and moreover uniformly in the initial condition. The third step is a post-processing from L^2_x to H^κ_x for any $\kappa < 1$.
Before we embark in discussing our intermediate results, let us give the proof of our main theorem.

Proof of Theorem 4.1. For $\kappa \in (0, 1)$ the assertion follows from combining (4.18) and Proposition 4.11 below, which provides an estimate on (4.19). For $k = 0$ we apply Proposition 4.9. □

4.3.1. L^2_x-energy estimate. For $s \leq t$ and $h \in C^\infty$ we define $J_{s,t}h$ as the solution to the equation

$$
\begin{cases}
(\partial_t - \Delta + m)J_{s,t}h = -3\Big(v^2_{s,t} + 2v_{s,t}\stackrel{\bullet}{|}_{s,t} + \stackrel{\bullet\,\bullet}{\vee}_{s,t}\Big)J_{s,t}h + 3c_{t-s,\infty}J_{s,t}, \\
J_{s,t}h|_{t=s} = h.
\end{cases} \qquad (4.20)
$$

In order to ease notation we will suppress the dependence on the initial condition but we will always assume that $v_{s,t}|_{t=s} = u^f_s$.

The first step towards bounding (4.19) is a standard energy estimate in order to bound the L^2_x-norm of $J_{s,t}h$ with respect to the L^2_x-norm of h. From now on, all proofs are postponed to Section 4.5.

Proposition 4.6. *For all $s \le t' \le t$, $m > 0$ we have*

$$
\begin{aligned}
&\|J_{s,t}h\|^2_{L^2_x} \\
&+ \int_{t'}^{t} e^{-2m(t-r)+2\int_r^t g(s,r')\,\mathrm{d}r'} \|\nabla J_{s,r}h\|^2_{L^2_x}\,\mathrm{d}r + \int_{t'}^{t} e^{-2m(t-r)+2\int_r^t g(s,r')\,\mathrm{d}r'} \|v_{s,r}J_{s,r}h\|^2_{L^2_x}\,\mathrm{d}r \\
&\le e^{-2m(t-t')+2\int_{t'}^t g(s,r)\,\mathrm{d}r} \|J_{s,t'}h\|^2_{L^2_x},
\end{aligned}
\tag{4.21}
$$

where

$$
\begin{aligned}
g(s,t) := c\Big(&\big\|\Psi_{s,t}\big\|^2_{-\alpha} + \big\|\Psi_{s,t}\big\|^{\frac{2}{1+\alpha}}_{-\alpha} \|\nabla v_{s,t}\|_\infty^{\frac{2\alpha}{1+\alpha}} + \big\|\Psi_{s,t}\big\|^{\frac{2}{1-\alpha}}_{-\alpha} \\
&+ \big\|\Psi^{2}_{s,t}\big\|_{-\alpha} + \big\|\Psi^{2}_{s,t}\big\|^{\frac{2}{2-\alpha}}_{-\alpha} + c_{t-s,\infty}\Big)
\end{aligned}
$$

for some deterministic constant $c \equiv c(\alpha) < \infty$. In particular, we have

$$
\|J_{s,t}h\|^2_{L^2_x} \le e^{-2m(t-s)+2\int_s^t g(s,r)\,\mathrm{d}r} \|h\|^2_{L^2_x}. \tag{4.22}
$$

There are some important things we want to remark concerning Proposition 4.6. The first remark is that if it were not for the singular nature and the renormalization procedure involved the *error* term g would be zero and hence we would have a clean energy estimate. The second is that in order to prove Theorem 4.1 with an L^2_x-norm on the r.h.s. it is enough to consider (4.22) but since our goal is to achieve an H^κ_x-estimate it is crucial to use the additional information coming from (4.21), namely, the estimate on the gradient of $J_{s,r}h$ and the product $v_{s,r}J_{s,r}h$. The last and most important thing we want to remark makes the bridge to our next section. Notice that by Fernique's theorem the quantity $\|\Psi_{s,t}\|_{-\alpha}$ in $g(s,t)$ has Gaussian moments, whereas $\|\Psi^{2}_{s,t}\|_{-\alpha}$ has only exponential moments. Therefore, the pre-factor on the r.h.s. of (4.21) fails to be stochastically integrable. To overcome this problem we appeal to a stopping time argument, which we explain in the next section.

4.3.2. Stopping time argument and L^2_x-estimate. In order to bypass the issue of integrability of $e^{\int_s^t g(s,r)\,\mathrm{d}r}$ we appeal to probabilistic arguments inspired by [**21**]. More precisely, we restart the Wick powers Ψ^{k}, $k = 1, 2, 3$, each time they exceed a certain barrier. This allows us to replace $g(s,t)$ by a the length of the time interval times a deterministic constant times a counting processes $N(t)$, see (4.25) below. By choosing the length of the time interval small enough we can ensure the exponential integrabillity of the counting process $N(t)$, see Proposition 4.8. The drawback is the exponential factor $e^{m_* t}$ appearing in Theorem 4.1.

We define the stopping time

$$\widetilde{\tau}_1 := \inf\left\{ t \geq 0 : \sup_{k=1,2,3} \left\| \overset{\bullet\,\bullet}{\vee}{}^{k}_{0,t} \right\|_{-\alpha} \geq \eta \right\}$$

and for $\theta \in (0,1)$ we set

$$\tau_1 := \widetilde{\tau}_1 \wedge \theta.$$

The value of $\eta \equiv \eta(\alpha, L)$ will be fixed via

$$\sup_{\theta\in(0,1]} \mathbb{P}(\widetilde{\tau}_1 \leq \theta) < \frac{1}{4}. \tag{4.23}$$

This is possible due to Markov's inequality, (4.12) and the fact that $\theta < 1$ since

$$\mathbb{P}(\widetilde{\tau}_1 \leq \theta) \leq \mathbb{P}\left(\sup_{k=1,2,3} \sup_{t\leq\theta} \left\| \overset{\bullet\,\bullet}{\vee}{}^{k}_{0,t} \right\|_{-\alpha} \geq \eta \right) \leq \mathbb{P}\left(\sup_{k=1,2,3} \sup_{t\leq 1} \left\| \overset{\bullet\,\bullet}{\vee}{}^{k}_{0,t} \right\|_{-\alpha} \geq \eta \right) < \frac{1}{4}.$$

We inductively define a sequence of stopping times for $n > 1$ via

$$\widetilde{\tau}_n := \inf\left\{ t \geq \tau_{n-1} : \sup_{k=1,2,3} \left\| \overset{\bullet\,\bullet}{\vee}{}^{k}_{\tau_{n-1},t} \right\|_{-\alpha} \geq \eta \right\},$$

where $\left\{ \overset{\bullet\,\bullet}{\vee}{}^{k}_{s,t} \right\}_{t\geq s}$ denotes the process at time t restarted at time s, and

$$\tau_n := \tau_{n-1} + (\widetilde{\tau}_n - \widetilde{\tau}_{n-1}) \wedge \theta.$$

Furthermore, we define the standard filtration of σ-algebras for $t > 0$

$$\mathcal{F}_t := \sigma\Big(\xi(h) : h \in L^2_{t,x},\, \operatorname{supp} h \subset (0,t) \times \mathbb{T}^2 \Big).$$

We notice that since $\sigma\Big(\overset{\bullet}{|}_{0,\cdot\wedge t} \Big) \subset \mathcal{F}_t$ and the process $\overset{\bullet}{|}_{t,t+\cdot}$ is independent of $\overset{\bullet}{|}_{0,\cdot\wedge t}$ (cf. [**111**, Proposition 2.3]), by the strong Markov property for any stopping time τ the process $\overset{\bullet}{|}_{\tau,\tau+\cdot}$ is independent of $\mathcal{F}_\tau$. Since $\sigma\Big(\overset{\bullet\,\bullet}{\vee}{}^{k}_{0,\cdot\wedge t} \Big) \subset \mathcal{F}_t$, we have that for any $n \geq 1$

$$\widetilde{\tau}_n - \widetilde{\tau}_{n-1} \text{ is independent of } \mathcal{F}_{\widetilde{\tau}_{n-1}}$$

and thus

$$\overset{\bullet\,\bullet}{\vee}{}^{k}_{\tau_n,\tau_n+\cdot} \text{ is independent of } \mathcal{F}_{\tau_n}. \tag{4.24}$$

Let $t \leq \tau_1$. By the definition of τ_1 we know that $\left\| \overset{\bullet\,\bullet}{\vee}{}^{k}_{0,t} \right\|_{-\alpha} < \eta$ for all $k = 1,2,3$. Then by Proposition 4.6 and Lemma 4.21 for any $\varepsilon > 0$ we have that

$$\begin{aligned} \|J_{0,t}h\|^2_{L^2_x} &\leq e^{-2mt+2\int_0^t g(0,r)\,\mathrm{d}r} \|h\|^2_{L^2_x} \\ &\leq e^{-2mt+2c\left(\eta^2+\eta^{\frac{2}{1-\alpha}}+\eta^{\frac{2}{2-\alpha}}\right)t+c\eta^{\frac{2}{1+\alpha}}\int_0^t r^{\frac{2\alpha}{1+\alpha}(-1-\varepsilon)}\,\mathrm{d}r} \|h\|^2_{L^2_x} \end{aligned}$$

$$\leq e^{-2mt+2c\left(t+t^{\frac{1-\alpha(1+2\varepsilon)}{1+\alpha}}\right)}\|h\|^2_{L^2_x}$$

for some $c \equiv c(\alpha,\eta) < \infty$. For $\tau_{n-1} \leq t \leq \tau_n$ we have by Proposition 4.6 in the same manner

$$\|J_{\tau_{n-1},t}h\|^2_{L^2_x} \leq e^{-2m(t-\tau_{n-1})+2c\left((t-\tau_{n-1})+(t-\tau_{n-1})^{\frac{1-\alpha(1+2\varepsilon)}{1+\alpha}}\right)}\|J_{\tau_{n-2},\tau_{n-1}}h\|^2_{L^2_x}$$

and thus by induction we get that

$$\|J_{0,t}h\|^2_{L^2_x} \leq e^{-2mt+2c\left((t-\tau_{n-1})^{\frac{1-\alpha(1+2\varepsilon)}{1+\alpha}}+(t-\tau_{n-1})+\sum_{i=1}^{n-1}(\tau_i-\tau_{i-1})^{\frac{1-\alpha(1+2\varepsilon)}{1+\alpha}}+(\tau_i-\tau_{i-1})\right)}\|h\|^2_{L^2_x}.$$

From now on we set $\gamma := \frac{1-\alpha(1+2\varepsilon)}{1+\alpha}$. By introducing the following counting process

$$N(t) := \inf\{n \geq 1 : \tau_n \geq t\} \tag{4.25}$$

we furthermore estimate using $\tau_i - \tau_{i-1} \leq \theta$ for any $t \geq 0$

$$\|J_{0,t}h\|^2_{L^2_x} \leq e^{-2mt+2c\theta^\gamma N(t)}\|h\|^2_{L^2_x}. \tag{4.26}$$

Remark 4.7. Although we suppressed the dependence on the initial condition f to ease the notation, we should also point out that our estimates do not depend f. This is possible because of the “coming down from infinity" property (cf. Section 4.D), which allows us to ensure that the gradient estimate in Theorem 4.1 is uniform in f.

The above procedure boils down the problem of estimating $J_{0,t}h$ to showing exponential moment for $N(t)$. Since the sequence $\{\tau_n\}_{n\geq 1}$ has independent increments[5] we can expect this provided we choose θ small enough. This is the content of the next proposition, which is in the core of our argument, therefore we present the proof here.

Proposition 4.8. *Let $c \equiv c(\alpha,\eta) > 0$ as in* (4.26). *For all $p \geq 1$ there exists $\theta_0 \equiv \theta_0(\alpha,p,\eta) \in (0,1)$ which is independent of m such that for all $\theta \leq \theta_0$ and $t \geq 0$*

$$\mathbb{E}\Big[e^{pc\theta^\gamma N(t)}\Big]^{\frac{1}{p}} \leq Ce^{\frac{2\ln 2}{\theta}t},$$

where C is a universal constant uniform in L and m.

Proof. Let $n \geq 1$. The Markov inequality and (4.24) yield

$$\mathbb{P}(N(t) \geq n) = \mathbb{P}(\tau_n \leq t) = \mathbb{P}\left(\sum_{k=1}^n(\tau_k - \tau_{k-1}) \leq t\right)$$
$$= \mathbb{P}\Big(e^{-\frac{2\ln 2}{\theta}\sum_{k=1}^n(\tau_k-\tau_{k-1})} \geq e^{-\frac{2\ln 2}{\theta}t}\Big)$$

[5] at least if conditioned onto $\mathcal{F}_{\tau_{n-1}}$

$$\leq e^{\frac{2\ln 2}{\theta}t}\mathbb{E}\Big[e^{-\frac{2\ln 2}{\theta}\sum_{k=1}^{n}(\tau_k-\tau_{k-1})}\Big] = e^{\frac{2\ln 2}{\theta}t}\Big(\mathbb{E}\Big[e^{-\frac{2\ln 2}{\theta}\tau_1}\Big]\Big)^n. \tag{4.27}$$

Moreover, we estimate

$$\mathbb{E}\Big[e^{-\frac{2\ln 2}{\theta}\tau_1}\Big] \leq e^{-2\ln 2} + \mathbb{P}(\widetilde{\tau}_1 \leq \theta) \leq \frac{1}{4} + \mathbb{P}(\widetilde{\tau}_1 \leq \theta). \tag{4.28}$$

which combined with (4.23) yields

$$\mathbb{E}\Big[e^{-\frac{2\ln 2}{\theta}\tau_1}\Big] \leq \frac{1}{2}.$$

Finally, we have by (4.27) that

$$\mathbb{P}(N(t) \geq n) \leq 2^{-n}e^{\frac{2\ln 2}{\theta}t}$$

and the claim follows by choosing θ_0 small enough such that $\theta_0^\gamma < \frac{\ln 2}{cp}$. □

As an immediate consequence of Proposition 4.8 and (4.26) we obtain the following L^2_x-estimate.

Proposition 4.9. *For every $p \geq 1$ there exists $m_* \equiv m_*(\alpha, p, L) > 0$ such that for every $t \geq 0$,*

$$\mathbb{E}\Big[\|J_{0,t}\|^p_{L^2_x \to L^2_x}\Big]^{\frac{1}{p}} \leq Ce^{-(m-m_*)t},$$

for some universal constant $C < \infty$ which is uniform in m and L.

4.3.3. Upgrade from L^2_x to H^κ_x. In this section we upgrade the L^2_x-estimate in Proposition 4.9 to an H^κ_x-estimate.

The first step is to post-process Proposition 4.6 using (4.26).

Corollary 4.10. *For every $t' \leq t$ we have that*

$$\begin{aligned} &\|J_{0,t}h\|^2_{L^2_x} + \int_{t'}^t e^{-2m(t-s)}\|\nabla J_{0,s}h\|^2_{L^2_x}\,\mathrm{d}s + \int_{t'}^t e^{-2m(t-s)}\|v_{0,s}J_{0,s}h\|^2_{L^2_x}\,\mathrm{d}s \\ &\quad \leq e^{-2mt}\Big(e^{2c\theta^\gamma N(t')} + \int_{t'}^t e^{2c\theta^\gamma N(s)}g(0,s)\,\mathrm{d}s\Big)\|h\|^2_{L^2_x}. \end{aligned} \tag{4.29}$$

We can now upgrade Proposition 4.9 to H^κ_x.

Proposition 4.11. *Let $\kappa \in (0,1)$ and $p \geq 1$. For every $\alpha < \frac{1-\kappa}{5}$ there exists $m_* \equiv m_*(\alpha, p, L) > 0$ such that*

$$\mathbb{E}\Big[\|J_{0,t}\|^p_{L^2_x \to H^\kappa_x}\Big]^{\frac{1}{p}} \leq C(t \wedge 1)^{-\frac{k+5\alpha}{2}}e^{-(m-m_*)t},$$

for some constant $C \equiv C(p, \alpha, \kappa, L) < \infty$ which is uniform in f.

Here we need $\kappa < 1$ to ensure the integrability of the exponent when $t \searrow 0$. Moreover, we again crucially used the "coming down from infinity" property that ensures that the bound does not depend on the initial data f.

4.4. Spectral gap inequalities

In this section we give our main application of the gradient estimate Theorem 4.1. At the core of the argument lies the celebrated method of Bakry and Émery (cf. [**7, 8, 96**]) to prove log-Sobolev inequalities as well as spectral gap inequalities.
As was discussed in Section 1.5, by the convexity of the potential it is natural to expect that (4.1) satisfies a log-Sobolev inequality, but due to the singular nature of the equation we are only able to prove a spectral gap inequality. At this point we want to mention that in [**72**] it was shown that (4.1) does satisfy a log-Sobolev inequality when $d = 1$[6] with respect to L^2_x. In the following we also want to point out how the required renormalization procedure obstructs us from proving an log-Sobolev inequality. The first step is to show the following identity (cf. [**9**, p. 131, (3.1.21)]), the proof of of which can be found in the appendix, Section 4.E.

Proposition 4.12. *The following identity holds for every $t > 0$ and $F \in \mathcal{F}C_b^\infty$,*

$$P_t F^2(f) - (P_t F(f))^2 = 2 \int_0^t P_{t-s}\Big(\|DP_s F\|^2_{L^2_x}\Big)(f)\, \mathrm{d}s \quad \nu\text{-a.s. in } f. \tag{4.30}$$

We are now in position to prove Theorem 4.2 and Corollary 4.3.

Proof of Theorem 4.2 and Corollary 4.3. We apply Theorem 4.1 combined with (4.30) and the fact that P_t is a Markov semigroup yielding

$$P_t\Big(F^2\Big) - (P_t F)^2 \lesssim \int_0^t (s \wedge 1)^{-\kappa-\varepsilon} e^{-2(m-m_*)s}\, \mathrm{d}s\ P_t \|DF\|^2_{H_x^{-\kappa}}.$$

Finally, choosing $m > m_*$ and noting that

$$\int_0^\infty (s \wedge 1)^{-\kappa-\varepsilon} e^{-2(m-m_*)s}\, \mathrm{d}s \lesssim \frac{1}{(m-m_*)^{1-\kappa-\varepsilon}} \vee \frac{1}{m-m_*},$$

we appeal to ergodicity (cf. [**111**, p. 1241, Corollary 6.6]) letting $t \nearrow \infty$ to obtain (4.9). $\square$

4.5. Proof of intermediate statements

In this section we collect the proofs of the intermediate statements missing from the previous section.

Proof of Proposition 4.6. Testing the equation (4.20) with $J_{s,t}h$ yields

$$\begin{aligned} &\frac{1}{2}\partial_t \|J_{s,t}h\|^2_{L^2_x} + \|\nabla J_{s,t}h\|^2_{L^2_x} + m\|J_{s,t}h\|^2_{L^2_x} + 3\|v_{s,t}(J_{s,t}h)\|^2_{L^2_x} \\ &\quad = -6\Big(\<1>_{s,t}, v_{s,t}(J_{s,t}h)^2\Big)_{L^2_x} - 3\Big(\<2>_{s,t}, (J_{s,t}h)^2\Big)_{L^2_x} + 3c_{t-s,\infty}\|J_{s,t}h\|^2_{L^2_x}. \end{aligned} \tag{4.31}$$

---

[6]and the equation does *not* require any renormalization

We start by estimating $\left|\left(\rotatebox{90}{$\multimap$}_{s,t}, v_{s,t}(J_{s,t}h)^2\right)_{L^2_x}\right|$. To this end, we apply [**111**, Proposition A.8] to get

$$\left|\left(\rotatebox{90}{$\multimap$}_{s,t}, v_{s,t}(J_{s,t}h)^2\right)_{L^2_x}\right| \lesssim \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\left\|v_{s,t}(J_{s,t}h)^2\right\|_{B^{\alpha}_{1,1}}$$

and then use Proposition A.9 in [**111**] such that we end up with

$$\left\|v_{s,t}(J_{s,t}h)^2\right\|_{B^{\alpha}_{1,1}} \lesssim \left\|v_{s,t}(J_{s,t}h)^2\right\|^{1-\alpha}_{L^1_x}\left\|\nabla\left(v_{s,t}(J_{s,t}h)^2\right)\right\|^{\alpha}_{L^1_x} + \left\|v_{s,t}(J_{s,t}h)^2\right\|_{L^1_x}.$$

Moreover, the Cauchy–Schwarz inequality and the chain rule yield

$$\begin{aligned}
&\left\|v_{s,t}(J_{s,t}h)^2\right\|^{1-\alpha}_{L^1_x}\left\|\nabla\left(v_{s,t}(J_{s,t}h)^2\right)\right\|^{\alpha}_{L^1_x} + \left\|v_{s,t}(J_{s,t}h)^2\right\|_{L^1_x} \\
&\lesssim \|J_{s,t}h\|^{1-\alpha}_{L^2_x}\|v_{s,t}(J_{s,t}h)\|^{1-\alpha}_{L^2_x}\left\|(J_{s,t}h)^2\nabla v_{s,t} + 2v_{s,t}(J_{s,t}h)\nabla J_{s,t}h\right\|^{\alpha}_{L^1_x} \\
&+ \|J_{s,t}h\|_{L^2_x}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}.
\end{aligned}$$

The Cauchy–Schwarz inequality again implies

$$\begin{aligned}
\left\|(J_{s,t}h)^2\nabla v_{s,t} + 2v_{s,t}(J_{s,t}h)\nabla J_{s,t}h\right\|^{\alpha}_{L^1_x} \lesssim& \|\nabla v_{s,t}\|^{\alpha}_{L^\infty_x}\|J_{s,t}h\|^{2\alpha}_{L^2_x} \\
&+ 2^{\alpha}\|v_{s,t}(J_{s,t}h)\|^{\alpha}_{L^2_x}\|\nabla J_{s,t}h\|^{\alpha}_{L^2_x}.
\end{aligned}$$

Hence we have shown that

$$\begin{aligned}
\left|\left(\rotatebox{90}{$\multimap$}_{s,t}, v_{s,t}(J_{s,t}h)^2\right)_{L^2_x}\right| &\lesssim \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|\nabla v_{s,t}\|^{\alpha}_{L^\infty_x}\|v_{s,t}(J_{s,t}h)\|^{1-\alpha}_{L^2_x}\|J_{s,t}h\|^{1+\alpha}_{L^2_x} \\
&+ \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|J_{s,t}h\|^{1-\alpha}_{L^2_x}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}\|\nabla J_{s,t}h\|^{\alpha}_{L^2_x} \\
&+ \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|J_{s,t}h\|_{L^2_x}\|v_{s,t}(J_{s,t}h)\|_{L^2_x} \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

Then we have by Young's inequality for some $\lambda > 0$ to be chosen later

$$I_3 = \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|J_{s,t}h\|_{L^2_x}\|v_{s,t}(J_{s,t}h)\|_{L^2_x} \leq \frac{1}{2\lambda}\left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|^2_{-\alpha}\|J_{s,t}h\|^2_{L^2_x} + \frac{\lambda}{2}\|v_{s,t}(J_{s,t}h)\|^2_{L^2_x}$$

and

$$\begin{aligned}
I_1 &= \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|\nabla v_{s,t}\|^{\alpha}_{L^\infty_x}\|v_{s,t}(J_{s,t}h)\|^{1-\alpha}_{L^2_x}\|J_{s,t}h\|^{1+\alpha}_{L^2_x} \\
&\leq \frac{1+\alpha}{2\lambda^{\frac{1+\alpha}{2}}}\left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|^{\frac{2}{1+\alpha}}_{-\alpha}\|\nabla v_{s,t}\|^{\frac{2\alpha}{1+\alpha}}_{L^\infty_x}\|J_{s,t}h\|^2_{L^2_x} + \frac{(1-\alpha)\lambda^{\frac{2}{1-\alpha}}}{2}\|v_{s,t}(J_{s,t}h)\|^2_{L^2_x}
\end{aligned}$$

as well as

$$I_2 = \left\|\rotatebox{90}{$\multimap$}_{s,t}\right\|_{-\alpha}\|J_{s,t}h\|^{1-\alpha}_{L^2_x}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}\|\nabla J_{s,t}h\|^{\alpha}_{L^2_x}$$

$$
\begin{aligned}
&\le \frac{1}{2\lambda}\Big\|\RSI_{s,t}\Big\|_{-\alpha}^{2}\|J_{s,t}h\|_{L^2_x}^{2(1-\alpha)}\|\nabla J_{s,t}h\|_{L^2_x}^{2\alpha}+\frac{\lambda}{2}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}^{2}\\
&\le \frac{1-\alpha}{2^{\frac{1}{1-\alpha}}\lambda^{\frac{2}{1-\alpha}}}\Big\|\RSI_{s,t}\Big\|_{-\alpha}^{\frac{2}{1-\alpha}}\|J_{s,t}h\|_{L^2_x}^{2}+\alpha\lambda^{\frac{1}{\alpha}}\|\nabla J_{s,t}h\|_{L^2_x}^{2}+\frac{\lambda}{2}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}^{2}.
\end{aligned}
$$

For the second term on the right hand side of (4.31) we proceed similarly. First of all, Proposition A.8 in [**111**] yields

$$
\left|\Big(\RSV_{s,t},(J_{s,t}h)^2\Big)_{L^2_x}\right|\le\Big\|\RSV_{s,t}\Big\|_{-\alpha}\Big(\|J_{s,t}h\|_{L^2_x}^{2(1-\alpha)}\|2(J_{s,t}h)\nabla J_{s,t}h\|_{L^1_x}^{\alpha}+\|J_{s,t}h\|_{L^2_x}^{2}\Big)
$$

and hence the Cauchy–Schwarz inequality combined with Young's inequality with the same $\lambda>0$ as before yields

$$
\begin{aligned}
\left|\Big(\RSV_{s,t},(J_{s,t}h)^2\Big)_{L^2_x}\right| \le& 2^{\alpha}\frac{2-\alpha}{2\lambda^{\frac{2}{2-\alpha}}}\Big\|\RSV_{s,t}\Big\|_{-\alpha}^{\frac{2}{2-\alpha}}\|J_{s,t}h\|_{L^2_x}^{2}+\frac{\alpha\lambda^{\frac{2}{\alpha}}}{2}\|\nabla J_{s,t}h\|_{L^2_x}^{2}\\
&+\Big\|\RSV_{s,t}\Big\|_{-\alpha}\|J_{s,t}h\|_{L^2_x}^{2}.
\end{aligned}
$$

Then we set

$$
\begin{aligned}
g(s,t):=&\frac{1}{2\lambda}\Big\|\RSI_{s,t}\Big\|_{-\alpha}^{2}+\frac{1+\alpha}{2\lambda^{\frac{1+\alpha}{2}}}\Big\|\RSI_{s,t}\Big\|_{-\alpha}^{\frac{2}{1+\alpha}}\|\nabla v_{s,t}\|_{L^\infty_x}^{\frac{2\alpha}{1+\alpha}}+\frac{1-\alpha}{2^{\frac{1}{1-\alpha}}\lambda^{\frac{2}{1-\alpha}}}\Big\|\RSI_{s,t}\Big\|_{-\alpha}^{\frac{2}{1-\alpha}}\\
&+\Big\|\RSV_{s,t}\Big\|_{-\alpha}+2^{\alpha}\frac{2-\alpha}{2\lambda^{\frac{2}{2-\alpha}}}\Big\|\RSV_{s,t}\Big\|_{-\alpha}^{\frac{2}{2-\alpha}}+3c_{t-s,\infty}.
\end{aligned}
$$

By choosing λ small enough, we can absorb some of the terms into $\|\nabla J_{s,t}h\|_{L^2_x}^2$ respectively $\|v_{s,t}(J_{s,t}h)\|_{L^2_x}^2$ into the right hand side and we end up with the estimate

(4.32)
$$
\frac{1}{2}\partial_t\|J_{s,t}h\|_{L^2_x}^{2}+m\|J_{s,t}h\|_{L^2_x}^{2}+\frac{1}{2}\|\nabla J_{s,t}h\|_{L^2_x}^{2}+\frac{1}{2}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}^{2}\le g(s,t)\|J_{s,t}h\|_{L^2_x}^{2}.
$$

Then the chain rule combined with (4.32) yields

$$
\begin{aligned}
&\partial_t\Big(e^{2mt-2\int_0^t g(s,r)\,\mathrm{d}r}\|J_{s,t}h\|_{L^2_x}^{2}\Big)+e^{2mt-2\int_0^t g(s,r)\,\mathrm{d}r}\|\nabla J_{s,t}h\|_{L^2_x}^{2}\\
&+e^{2mt-2\int_0^t g(s,r)\,\mathrm{d}r}\|v_{s,t}(J_{s,t}h)\|_{L^2_x}^{2}\le 0. \qquad (4.33)
\end{aligned}
$$

Integrating (4.33) from t' to t we end up with

$$
\begin{aligned}
&\|J_{s,t}h\|_{L^2_x}^{2}+\int_{t'}^{t}e^{-2m(t-r)+2\int_r^t g(s,r')\,\mathrm{d}r'}\|\nabla J_{s,r}h\|_{L^2_x}^{2}\,\mathrm{d}r\\
&+\int_{t'}^{t}e^{-2m(t-r)+2\int_r^t g(s,r')\,\mathrm{d}r'}\|v_{s,t}(J_{s,r}h)\|_{L^2_x}^{2}\,\mathrm{d}r\\
&\le e^{-2m(t-t')+2\int_{t'}^t g(s,r)\,\mathrm{d}r}\|J_{s,t'}h\|_{L^2_x}^{2}.
\end{aligned}
$$

□

Proof of Corollary 4.10. The estimate (4.32) yields for $s=0$

$$\partial_t\Big(e^{2mt}\|J_{0,t}h\|_{L^2_x}^2\Big)+e^{2mt}\Big(\|\nabla J_{0,t}\|_{L^2_x}^2+\|v_{0,t}J_{0,t}h\|_{L^2_x}^2\Big)\le 2e^{2mt}g(0,t)\|J_{0,t}h\|_{L^2_x}^2.$$

Then we integrate from t' to t to obtain

$$\begin{aligned}\|J_{0,t}h\|_{L^2_x}^2&+\int_{t'}^t e^{-2m(t-s)}\Big(\|\nabla J_{0,s}\|_{L^2_x}^2+\|v_{0,s}J_{0,s}h\|_{L^2_x}^2\Big)\,\mathrm{d}s\\ &\le e^{-2m(t-t')}\|J_{0,t'}h\|_{L^2_x}^2+2\int_{t'}^t e^{-2m(t-s)}g(0,s)\|J_{0,s}h\|_{L^2_x}^2\,\mathrm{d}s.\end{aligned}$$

Applying (4.26) to $\|J_{0,t'}h\|_{L^2_x}^2$ respectively to $\|J_{0,s}h\|_{L^2_x}^2$ yields the assertion. □

Proof of Proposition 4.11. First of all, Duhamel's formula yields

$$\begin{aligned}J_{0,t}h&=S_{\frac t2}J_{0,\frac t2}h-3\int_{\frac t2}^t S_{t-s}\Big\{\Big(v_{0,s}^2+2v_{0,s}\uparrow_{0,s}+\vee_{0,s}-c_{s,\infty}\Big)J_{0,s}h\Big\}\,\mathrm{d}s\\ &=I_1+I_2+I_3+I_4+I_5.\end{aligned}$$

Then we estimate I_1 according to

$$\Big\|S_{\frac t2}J_{0,\frac t2}h\Big\|_{\mathcal{B}^\kappa_{2,2}}\overset{(4.15)}{\lesssim}(t\wedge1)^{-\frac\kappa2}e^{-m\frac t2}\Big\|J_{0,\frac t2}h\Big\|_{L^2_x}\overset{(4.26)}{\lesssim}(t\wedge1)^{-\frac\kappa2}e^{-mt+C\theta^\gamma N(\frac t2)}\|h\|_{L^2_x}.$$

For I_2 we further estimate

$$\begin{aligned}\int_{\frac t2}^t\Big\|S_{t-s}\Big(v_{0,s}^2J_{0,s}h\Big)\Big\|_{\mathcal{B}^\kappa_{2,2}}\,\mathrm{d}s&\overset{(4.15)}{\lesssim}\int_{\frac t2}^t((t-s)\wedge1)^{-\frac\kappa2}e^{-m(t-s)}\Big\|v_{0,s}^2J_{0,s}h\Big\|_{L^2_x}\,\mathrm{d}s\\ &\lesssim\int_{\frac t2}^t((t-s)\wedge1)^{-\frac\kappa2}\|v_{0,s}\|_{L^\infty_x}e^{-m(t-s)}\|v_{0,s}J_{0,s}h\|_{L^2_x}\,\mathrm{d}s\\ &\lesssim\left(\int_{\frac t2}^t((t-s)\wedge1)^{-\kappa}\|v_{0,s}\|_{L^\infty_x}^2\,\mathrm{d}s\right)^{\frac12}\left(\int_{\frac t2}^t e^{-2m(t-s)}\|v_{0,s}J_{0,s}h\|_{L^2_x}^2\,\mathrm{d}s\right)^{\frac12}\\ &\overset{(4.29)}{\lesssim}\left(\int_{\frac t2}^t((t-s)\wedge1)^{-\kappa}\|v_{0,s}\|_{L^\infty_x}^2\,\mathrm{d}s\right)^{\frac12}e^{-mt}\left(e^{2c\theta^\gamma N(\frac t2)}+\int_{\frac t2}^t e^{2c\theta^\gamma N(s)}g(0,s)\,\mathrm{d}s\right)^{\frac12}\|h\|_{L^2_x},\end{aligned}$$

where we used again Hölder's inequality in the third step.
Estimating I_3 yields

$$\begin{aligned}&\int_{\frac t2}^t\Big\|S_{t-s}\Big(v_{0,s}\uparrow_{0,s}J_{0,s}h\Big)\Big\|_{\mathcal{B}^\kappa_{2,2}}\,\mathrm{d}s\\ &\overset{(4.15)}{\lesssim}\int_{\frac t2}^t((t-s)\wedge1)^{-\frac{\kappa+\alpha}{2}}e^{-m(t-s)}\Big\|v_{0,s}\uparrow_{0,s}J_{0,s}h\Big\|_{\mathcal{B}^{-\alpha}_{2,2}}\,\mathrm{d}s\\ &\overset{(4.16),(4.17),(4.14)}{\lesssim}\int_{\frac t2}^t((t-s)\wedge1)^{-\frac{\kappa+\alpha}{2}}\|\uparrow_{0,s}\|_{-\alpha}\|v_{0,s}\|_{2\alpha}e^{-m(t-s)}\|J_{0,s}h\|_{\mathcal{B}^1_{2,2}}\,\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&\lesssim \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \|\overset{\bullet}{\mathsf{I}}_{0,s}\|_{-\alpha}^2 \|v_{0,s}\|_{2\alpha}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \left(\int_{\frac{t}{2}}^{t} e^{-2m(t-s)} \|J_{0,s}h\|_{\mathcal{B}^1_{2,2}}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\overset{(4.29)}{\lesssim} \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \|\overset{\bullet}{\mathsf{I}}_{0,s}\|_{-\alpha}^2 \|v_{0,s}\|_{2\alpha}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\qquad \times e^{-mt} \left(e^{2c\theta^\gamma N(\frac{t}{2})} + \int_{\frac{t}{2}}^{t} e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s\right)^{\frac{1}{2}} \|h\|_{L^2_x} \\
&\lesssim \sup_{0\le s\le t} \left\|\overset{\bullet}{\mathsf{I}}_{0,s}\right\|_{-\alpha} \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \|v_{0,s}\|_{2\alpha}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\qquad \times e^{-mt} \left(e^{2c\theta^\gamma N(\frac{t}{2})} + \int_{\frac{t}{2}}^{t} e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s\right)^{\frac{1}{2}} \|h\|_{L^2_x},
\end{aligned}
$$

using Hölder's inequality in the third step.
The term I_4 is estimated via

$$
\begin{aligned}
&\int_{\frac{t}{2}}^{t} \left\|S_{t-s}\left(\overset{\bullet\bullet}{\mathsf{V}}_{0,s} J_{0,s}h\right)\right\|_{\mathcal{B}^{\kappa}_{2,2}} \mathrm{d}s \overset{(4.15)}{\lesssim} \int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\frac{\kappa+\alpha}{2}} e^{-m(t-s)} \left\|\overset{\bullet\bullet}{\mathsf{V}}_{0,s} J_{0,s}h\right\|_{\mathcal{B}^{-\alpha}_{2,2}} \mathrm{d}s \\
&\overset{(4.17),(4.14)}{\lesssim} \int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\frac{\kappa+\alpha}{2}} \left\|\overset{\bullet\bullet}{\mathsf{V}}_{0,s}\right\|_{-\alpha} e^{-m(t-s)} \|J_{0,s}h\|_{\mathcal{B}^1_{2,2}} \,\mathrm{d}s \\
&\lesssim \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \left\|\overset{\bullet\bullet}{\mathsf{V}}_{0,s}\right\|_{-\alpha}^2 \mathrm{d}s\right)^{\frac{1}{2}} \left(\int_{\frac{t}{2}}^{t} e^{-2m(t-s)} \|J_{0,s}h\|_{\mathcal{B}^1_{2,2}}^2 \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\overset{(4.29)}{\lesssim} \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \left\|\overset{\bullet\bullet}{\mathsf{V}}_{0,s}\right\|_{-\alpha}^2 \mathrm{d}s\right)^{\frac{1}{2}} \\
&\times e^{-mt} \left(e^{2c\theta^\gamma N(\frac{t}{2})} + \int_{\frac{t}{2}}^{t} e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s\right)^{\frac{1}{2}} \|h\|_{L^2_x} \\
&\lesssim \sup_{0\le s\le t} \left\|\overset{\bullet\bullet}{\mathsf{V}}_{0,s}\right\|_{-\alpha} \left(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha} \,\mathrm{d}s\right)^{\frac{1}{2}} \\
&\times e^{-mt} \left(e^{2c\theta^\gamma N(\frac{t}{2})} + \int_{\frac{t}{2}}^{t} e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s\right)^{\frac{1}{2}} \|h\|_{L^2_x},
\end{aligned}
$$

where again we have used Hölder's inequality in the third step.
Finally, we estimate I_5

$$
\int_{\frac{t}{2}}^{t} \|S_{t-s} c_{s,\infty} J_{0,s}h\|_{\mathcal{B}^{\kappa}_{2,2}} \,\mathrm{d}s \overset{(4.15)}{\lesssim} \int_{\frac{t}{2}}^{t} s^{-\frac{\beta}{2}} e^{-m(t-s)} \|J_{0,s}h\|_{B^1_{2,2}} \,\mathrm{d}s
$$

$$\lesssim \left(\int_{\frac{t}{2}}^t s^{-\beta}\,\mathrm{d}s\right)^{\frac{1}{2}}\left(\int_{\frac{t}{2}}^t e^{-2m(t-s)}\|J_{0,s}h\|_{L^2_x}^2\,\mathrm{d}s\right)^{\frac{1}{2}}$$

$$\overset{(4.29)}{\lesssim} \left(\int_{\frac{t}{2}}^t s^{-\beta}\,\mathrm{d}s\right)^{\frac{1}{2}} e^{-mt}\left(e^{2c\theta^\gamma N(\frac{t}{2})} + \int_{\frac{t}{2}}^t e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s\right)^{\frac{1}{2}}\|h\|_{L^2_x},$$

where we have used Hölder's inequality in the second step.

Then we use monotonicity of $t \mapsto N(t)$ to infer

$$\int_{\frac{t}{2}}^t e^{2c\theta^\gamma N(s)} g(0,s)\,\mathrm{d}s \le e^{2c\theta^\gamma N(t)} \int_{\frac{t}{2}}^t g(0,s)\,\mathrm{d}s$$

which all in all yields

$$\begin{aligned}\|J_{0,t}h\|_{\mathcal{B}^\kappa_{2,2}} &\lesssim e^{-mt+C\theta^\gamma N(t)}\left(1+\int_{\frac{t}{2}}^t g(0,s)\,\mathrm{d}s\right)\|h\|_{L^2_x}\\ &\quad\times\left((t\wedge 1)^{-\frac{\kappa}{2}} + \left(\int_{\frac{t}{2}}^t ((t-s)\wedge 1)^{-\kappa}\|v_{0,s}\|_{L^\infty_x}^2\,\mathrm{d}s\right)^{\frac{1}{2}}\right.\\ &\quad + \sup_{0\le s\le t}\left\|\dot{\mid}_{0,s}\right\|_{-\alpha}\left(\int_{\frac{t}{2}}^t ((t-s)\wedge 1)^{-\kappa-\alpha}\|v_{0,s}\|_{2\alpha}^2\,\mathrm{d}s\right)^{\frac{1}{2}}\\ &\quad\left. + \sup_{0\le s\le t}\left\|\ddot{\vee}_{0,s}\right\|_{-\alpha}\left(\int_{\frac{t}{2}}^t ((t-s)\wedge 1)^{-\kappa-\alpha}\,\mathrm{d}s\right)^{\frac{1}{2}} + \left(\int_{\frac{t}{2}}^t s^{-\beta}\,\mathrm{d}s\right)^{\frac{1}{2}}\right).\end{aligned}$$

Using the definition of g we see that

$$\begin{aligned}\int_{\frac{t}{2}}^t g(0,s)\,\mathrm{d}s &\lesssim t\sup_{0\le s\le t}\left\|\dot{\mid}_{0,s}\right\|_{-\alpha} + \sup_{0\le s\le t}\left\|\dot{\mid}_{0,s}\right\|_{-\alpha}^{\frac{2}{1+\alpha}}\int_{\frac{t}{2}}^t \|\nabla v_{0,s}\|_{L^\infty_x}^{\frac{2\alpha}{1+\alpha}}\,\mathrm{d}s + t\sup_{0\le s\le t}\left\|\dot{\mid}_{0,s}\right\|_{-\alpha}^{\frac{2}{1-\alpha}}\\ &\quad + t\sup_{0\le s\le t}\left\|\ddot{\vee}_{0,s}\right\|_{-\alpha}^{\frac{2}{2-\alpha}} + t^{1-\beta}\end{aligned}$$

and for any $p \ge 1$ we can estimate

$$\int_{\frac{t}{2}}^t \mathbb{E}\left[\|\nabla v_{0,s}\|_{L^\infty_x}^{\frac{2\alpha p}{1+\alpha}}\right]^{\frac{1}{p}}\mathrm{d}s \overset{(4.21)}{\lesssim} \int_{\frac{t}{2}}^t s^{-\frac{(1+\beta)2\alpha}{1+\alpha}}\,\mathrm{d}s \lesssim t^{1-\frac{(1+\beta)2\alpha}{1+\alpha}}.$$

Moreover, by Proposition 4.5 for every $p < \infty$ there exists $r > 0$ such that

$$\mathbb{E}\left[\sup_{0\le s\le t}\left\|\dot{\mid}_{0,s}\right\|_{-\alpha}^p\right]^{\frac{1}{p}} \lesssim (1+t)^r, \tag{4.34}$$

$$\mathbb{E}\left[\sup_{0\le s\le t}\left\|\ddot{\vee}_{0,s}\right\|_{-\alpha}^p\right]^{\frac{1}{p}} \lesssim (1+t)^r. \tag{4.35}$$

Any positive power of t can brutally be bounded by $C_\sigma e^{\sigma t}$ for $\sigma > 0$, thus we have for any $p \geq 1$

$$\mathbb{E}\Bigg[\bigg(\int_{\frac{t}{2}}^{t} g(0,s)\,\mathrm{d}s\bigg)^p\Bigg]^{\frac{1}{p}} \lesssim e^{\sigma t}$$

where we implicitly used Hölder's inequality in expectation.
Also, again for any $p \geq 1$ we have

$$\bigg(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa}\,\mathbb{E}\Big[\|v_{0,s}\|_{L^\infty_x}^p\Big]^{\frac{2}{p}}\,\mathrm{d}s\bigg)^{\frac{1}{2}} \overset{(4.20)}{\lesssim} \bigg(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa} s^{-1-\alpha}\,\mathrm{d}s\bigg)^{\frac{1}{2}} \lesssim (t\wedge 1)^{-\frac{\kappa+\alpha}{2}}$$

and similarly

$$\bigg(\int_{\frac{t}{2}}^{t} ((t-s)\wedge 1)^{-\kappa-\alpha}\,\mathbb{E}[\|v_{0,s}\|_{2\alpha}^p]^{\frac{2}{p}}\,\mathrm{d}s\bigg)^{\frac{1}{2}} \overset{(4.20)}{\lesssim} (t\wedge 1)^{-\frac{\kappa+5\alpha}{2}}.$$

Dividing by $\|h\|_{L^2_x}$, taking the supremum and using (4.34), (4.35) and (4.8), we conclude that

$$\mathbb{E}\Big[\|J_{0,t}\|_{L^2_x\to H^\kappa_x}^p\Big]^{\frac{1}{p}} \lesssim_{\alpha,\kappa,L} (t\wedge 1)^{-\frac{\kappa+5\alpha}{2}} e^{-\left(m-10\sigma-\frac{2\ln 2}{\theta}\right)t},$$

where we have used Hölder's inequality in probability repeatedly. ☐

Appendix

4.A. Estimate on the renormalization constant

Proposition 4.13. *The following estimate holds for any $\beta \in (0,1)$ and $t > 0$,*

$$c_{t,\infty} = 2\int_t^\infty \mathrm{d}s\, H_{2s}(0) \lesssim_\beta t^{-\frac{\beta}{2}}.$$

Proof. By a simple computation in Fourier space we have that

$$c_{t,\infty} = \sum_{k\in\mathbb{Z}^2} \frac{e^{-t(m+|k|^2)}}{m+|k|^2}.$$

Noticing that $e^{-t(m+|k|^2)} \lesssim_\gamma \frac{t^{-\frac{\beta}{2}}}{(m+|k|^2)^{\frac{\beta}{2}}}$ for any $\beta \in (0,1)$ we get the assertion since the sum $\sum_{k\neq 0} \frac{1}{|k|^{2+\beta}}$ is finite. □

4.B. Besov-norm estimates

Lemma 4.14 ([**112**, p. 308, (A.2)])**.** *Let $\alpha \leq \beta$ and $p,q \geq 1$, then we have*

$$\|f\|_{\mathcal{B}^\alpha_{p,q}} \leq \|f\|_{\mathcal{B}^\beta_{p,q}}.$$

Lemma 4.15 ([**112**, p. 309, Proposition A.5])**.** *Let $\alpha \leq \beta$ and $p,q \geq 1$, then it holds that*

$$\|S_t f\|_{\mathcal{B}^\beta_{p,q}} \lesssim e^{-mt}(t\wedge 1)^{\frac{\alpha-\beta}{2}} \|f\|_{\mathcal{B}^\alpha_{p,q}}$$

where S_t denotes the semigroup generated by $\Delta - m$ for $m \geq 0$.

Lemma 4.16 ([**111**, p. 309, Proposition A.6])**.** *Let $\alpha \geq 0$ and $p,q \geq 1$, then*

$$\|fg\|_{\mathcal{B}^\alpha_{p,q}} \lesssim \|f\|_{\mathcal{B}^\alpha_{p_1,q_1}} \|g\|_{\mathcal{B}^\alpha_{p_2,q_2}}$$

where $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$ as well as $\frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}$.

Lemma 4.17 ([**111**, p. 309, Proposition A.7])**.** *Let $\alpha < 0$ and $\beta > 0$ such that $\alpha+\beta > 0$ and $p,q \geq 1$, then*

$$\|fg\|_{\mathcal{B}^\alpha_{p,q}} \lesssim \|f\|_{\mathcal{B}^\alpha_{p_1,q_1}} \|g\|_{\mathcal{B}^\beta_{p_2,q_2}}$$

where $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2}$ and $\frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}$.

4.C. Stochastic estimates

In this section we provide an alternative argument for the stochastic estimates in Proposition 4.5 using the spectral gap inequality (4.36) for the noise ξ in the spirit of [**67**, Section 5] and [**81**].

Let F be cylindrical in ξ, i.e. there is $n \in \mathbb{N}$, $\overline{F} \in C_c^\infty(\mathbb{R}^n, \mathbb{R})$ and $h_1, \dots, h_n \in L^2_{t,x}$ such that $F(\xi) = \overline{F}(\xi(h_1), \dots, \xi(h_n))$. Since ξ is Gaussian it satisfies the following spectral gap inequality (cf. [**36**, p. 652, Proposition 4.1])

$$\mathbb{E}\Big[|F(\xi) - \mathbb{E}[F(\xi)]|^2\Big] \le \mathbb{E}\left[\left\|\frac{\partial}{\partial\xi}F(\xi)\right\|^2_{L^2_{t,x}}\right] \tag{4.36}$$

where $\frac{\partial}{\partial\xi}$ denotes the Malliavin derivative with respect to the noise ξ. This in turn can be used to construct the singular products as follows.

Proof of Proposition 4.5. For simplicity we assume that the noise ξ is smooth. By the spectral gap inequality (4.36) we know that for nice enough functionals $\Pi[\xi]$ and $p \ge 2$ there holds

$$\mathbb{E}^{\frac{1}{p}}|\Pi[\xi] - \mathbb{E}\Pi[\xi]|^p \lesssim_p \mathbb{E}^{\frac{1}{p}}\left\|\frac{\partial}{\partial\xi}\Pi[\xi]\right\|^p_{L^2_{t,x}}.$$

By duality, an estimate of the form

$$\left|\mathbb{E}\frac{\partial}{\partial\xi}\Pi[\xi](\delta\xi)\right| \le C\,\mathbb{E}^{\frac{1}{q}}\|\delta\xi\|^q_{L^2_{t,x}} \tag{4.37}$$

for any $\delta\xi : \Omega \to L^2_{t,x}$[7], where $q \in [1,2]$ is the dual exponent of p, implies

$$\mathbb{E}^{\frac{1}{p}}\left\|\frac{\partial}{\partial\xi}\Pi[\xi]\right\|^p_{L^2_{t,x}} \le C.$$

For $t > 0$ and $x \in \mathbb{T}^2$ we consider $\Pi_t(x) \in \Big\{\mathsf{I}_{0,t}(x), \mathsf{V}_{0,t}(x), \mathsf{W}_{0,t}(x)\Big\}$, where

$$\mathsf{V}_{0,t}(x) := \mathsf{I}^2_{0,t}(x) - c_{0,t}, \quad \mathsf{W}_{0,t}(x) := \mathsf{I}^3_{0,t}(x) - 3c_{0,t}\mathsf{I}_{0,t}(x),$$

for $c_{0,t} = \mathbb{E}\mathsf{I}_{0,t}(0)^2$. We treat $\Pi_t(x) \equiv \Pi_t[z](x)$ as a functional of ξ and aim to prove the following stochastic estimates (replacing x by 0 using stationarity) which are uniform in m,

$$\mathbb{E}^{\frac{1}{p}}|\Pi_{t\lambda}(0)|^p \lesssim \lambda^{-|\Pi|\alpha}\sqrt{t}^{|\Pi|\alpha}, \tag{4.38}$$

$$\mathbb{E}^{\frac{1}{p}}|(\Pi_{t+r} - \Pi_t)_\lambda(0)|^p \lesssim \lambda^{-(|\Pi|+1)\alpha}\sqrt{r}^\alpha\sqrt{t+r}^{|\Pi|\alpha} \tag{4.39}$$

for every $\alpha \in (0, \frac{1}{|\Pi|+1})$, where $|\Pi| = 1, 2, 3$ for $\Pi = \mathsf{I}, \mathsf{V}, \mathsf{W}$ respectively and $(\cdot)_\lambda$ denotes convolution with a suitable semigroup ψ_λ. By a Kolmogorov-type continuity criterion,

[7] where Ω denotes the underline probability space

see for [**91**, Lemma 10], we then obtain (4.12). It is important to stress the uniformity of our estimates in m which allows us to ensure that m_* in Theorem 4.1 does not depend on m[8]. This will be obvious in what follows except (4.52) where one should pay attention on how the power on $\sqrt{r}$ is chosen.

For $\delta\xi \in L^q_\omega L^2_{t,x}$ we let $\delta\mathbf{I}_{0,t}(x) := \frac{\partial}{\partial\xi}\mathbf{I}_{0,t}(\delta\xi) = \int_0^t \mathrm{d}s H_{t-s} * \delta\xi(s,x)$ and consider $\delta\Pi_t \in \{\delta\mathbf{I}_{0,t}, \delta\mathbf{I}_{0,t}\mathbf{I}_{0,t}, \delta\mathbf{I}_{0,t}\mathbf{V}_{0,t}\}$. As in [**81**], in order to prove (4.38) and (4.39) we appeal to duality and derive the following estimates for the Malliavin derivative of Π_t,

$$\mathbb{E}^{\frac{1}{q'}}|\delta\Pi_{t\lambda}(0)|^{q'} \lesssim \lambda^{-|\Pi|\alpha}\sqrt{t}^{|\Pi|\alpha}\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\xi|^q\Big\|_{L^2_{t,x}}, \tag{4.40}$$

$$\mathbb{E}^{\frac{1}{q'}}|(\delta\Pi_{t+r} - \delta\Pi_t)_\lambda(0)|^{q'} \lesssim \lambda^{-(|\Pi|+1)\alpha}\sqrt{r}^{\alpha}\sqrt{t+r}^{|\Pi|\alpha}\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\xi|^q\Big\|_{L^2_{t,x}}, \tag{4.41}$$

for all $q' < q < 2$. Note that in (4.40) and (4.41) we ask for an estimate of the $L^{q'}_\omega$-norm by the $L^2_{t,x}L^q_\omega$-norm which is stronger than the $L^q_\omega L^2_{t,x}$-norm for $q < 2$, therefore implying the dual estimate (4.37). As in [**81**] estimating the $L^{q'}_\omega$-norm for all $q' < q < 2$ allows us to proceed inductively, namely, in order to derive the dual estimate for $\delta\mathbf{V}_{0,t}$ we need the stronger estimate on $\delta\mathbf{I}_{0,t}$ and similarly for $\delta\mathbf{W}_{0,t}$.

To this end, we denote by $\overline{w}$ the $L^2_{t,x}L^q_\omega$-norm on the r.h.s. of (4.40) and (4.41) and introduce another scaling parameter Λ, coming from $(\cdot)_\Lambda$. We estimate commutators of the form

$$([\delta\Pi, (\cdot)_\lambda]\Pi_\lambda)_\Lambda(0) = \int \mathrm{d}x\,\psi_\Lambda(-x)\int \mathrm{d}y\,\psi_\lambda(y)(\delta\Pi(x-y) - \delta\Pi(x))\Pi_\lambda(x-y).$$

Using the Cauchy–Schwarz inequality in the x-variable we have[9]

$$\begin{aligned}
&\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\Big[\delta\mathbf{I}_{0,t}, (\cdot)_\lambda\Big]\Pi_\lambda\Big)_\Lambda(0)\Big|^{q'} \\
&= \mathbb{E}^{\frac{1}{q'}}\Big|\int \mathrm{d}x\psi_\Lambda(-x)\int \mathrm{d}y\psi_\lambda(y)(\delta\Pi(x-y) - \delta\Pi(x))\Pi_\lambda(x-y)\Big|^{q'} \\
&\leq \int \mathrm{d}x|\psi_\lambda(y)|\|\psi_\Lambda\|_{L^2}\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi(x-y) - \delta\Pi(x)|^q\Big\|_{L^2_x}\mathbb{E}^{\frac{1}{p}}|\Pi_\lambda(0)|^p.
\end{aligned} \tag{4.42}$$

For (4.40) we let $\delta\Pi = \delta\mathbf{I}_t$ and $\Pi \in \{\mathbf{I}_{0,t}, \mathbf{V}_{0,t}\}$. Using the interpolation inequality Lemma 4.18 and the Cauchy-Schwarz inequality in the s-variable we see that

$$\begin{aligned}
\Big\|\mathbb{E}^{\frac{1}{q}}\Big|\delta\mathbf{I}_{0,t}(x-y) - \delta\mathbf{I}_{0,t}(x)\Big|^q\Big\|_{L^2_x} &\leq \int_0^t \mathrm{d}s\int \mathrm{d}z\,|H_{t-s}(z-y) - H_{t-s}(z)|\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\xi|^q\Big\|_{L^2_x} \\
&\leq |y|^{1-\alpha}\Big(\int_0^t \mathrm{d}s\,e^{-2m(t-s)}(t-s)^{-1+\alpha}\Big)^{\frac{1}{2}}\overline{w}
\end{aligned}$$

[8] or equivalently θ in Proposition 4.5 does not depend on m

[9] Here $p \geq 2$ satisfies $\frac{1}{q'} = \frac{1}{q} + \frac{1}{p}$.

$$\leq |y|^{1-\alpha}\sqrt{t}^{\alpha}\overline{w}, \tag{4.43}$$

for all $\alpha \in (0,1)$ uniformly in m. Combining (4.42) and (4.43) yields

$$\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\Big[\delta\overset{\bullet}{\mathsf{I}}_{0,t},(\cdot)_\lambda\Big]\Pi_\lambda\Big)_\Lambda(0)\Big|^{q'} \lesssim \Lambda^{-1}\lambda^{1-\alpha}\sqrt{t}^{\alpha}\mathbb{E}^{\frac{1}{p}}|\Pi_\lambda(0)|^p\overline{w}. \tag{4.44}$$

Using (4.38) and the dyadic summation identity

$$([\delta\Pi,(\cdot)_\lambda]\Pi)_\Lambda = \sum_{\substack{k\geq 1\\ \lambda'=\frac{\lambda}{2^k}}}([\delta\Pi,(\cdot)_{\lambda'}](\Pi)_{\lambda'}])_{\Lambda+\lambda-2\lambda'},$$

we obtain via (4.44)

$$\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\Big[\delta\overset{\bullet}{\mathsf{I}}_{0,t},(\cdot)_\lambda\Big]\Pi\Big)_\Lambda(0)\Big|^{q'} \lesssim \Lambda^{-1}\lambda^{1-|\Pi|\alpha}\sqrt{t}^{|\Pi|\alpha}\overline{w}.$$

A simple post-processing of the last estimate choosing $\Lambda \sim \lambda$ gives

$$\mathbb{E}^{\frac{1}{q'}}|(\delta\overset{\bullet}{\mathsf{I}}_{0,t}\Pi)_\lambda(0)|^{q'} \lesssim \lambda^{-|\Pi|\alpha}\sqrt{t}^{|\Pi|\alpha}\overline{w}, \tag{4.45}$$

therefore yielding (4.40).

For (4.41) we write $\delta\overset{\bullet}{\mathsf{I}}_{0,t+r}\Pi_{t+r} - \delta\overset{\bullet}{\mathsf{I}}_{0,t}\Pi_t = \delta\overset{\bullet}{\mathsf{I}}_{0,t+r}(\Pi_{t+r}-\Pi_t) + \Pi_t\Big(\delta\overset{\bullet}{\mathsf{I}}_{0,t+r} - \delta\overset{\bullet}{\mathsf{I}}_{0,t}\Big)$ and use (4.42) for the pairs $\delta\Pi = \delta\overset{\bullet}{\mathsf{I}}_{0,t+r}$, $\Pi = \Pi_{t+r} - \Pi_t$ and $\delta\Pi = \delta\overset{\bullet}{\mathsf{I}}_{0,t+r} - \delta\overset{\bullet}{\mathsf{I}}_t$, $\Pi = \Pi_t$. For the first pair we apply (4.44) to get

$$\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\Big[\delta\overset{\bullet}{\mathsf{I}}_{0,t+r},(\cdot)_\lambda\Big](\Pi_{t+r}-\Pi_t)_\lambda\Big)_\Lambda(0)\Big|^{q'} \lesssim \Lambda^{-1}\lambda^{1-\alpha}\sqrt{t+r}^{\alpha}\mathbb{E}^{\frac{1}{p}}|(\Pi_{t+r}-\Pi_t)_\lambda(0)|^p\overline{w}.$$

Plugging in (4.39) for $\Pi_t \in \{\overset{\bullet}{\mathsf{I}}_{0,t}, \overset{\bullet\,\bullet}{\vee}_{0,t}\}$ and proceeding as for (4.45) yields

$$\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\delta\overset{\bullet}{\mathsf{I}}_{0,t+r}(\Pi_{t+r}-\Pi_t)\Big)_\lambda(0)\Big|^{q'} \lesssim \lambda^{-(|\Pi|+2)\alpha}\sqrt{r}^{\alpha}\sqrt{t+r}^{(|\Pi|+1)\alpha}\overline{w}. \tag{4.46}$$

For the second pair, abbreviating $\delta\Pi_{t,t+r} := \delta\overset{\bullet}{\mathsf{I}}_{0,t+r} - \delta\overset{\bullet}{\mathsf{I}}_{0,t}$, (4.42) implies

$$\begin{aligned}&\mathbb{E}^{\frac{1}{q'}}\Big|\Big(\Big[\delta\overset{\bullet}{\mathsf{I}}_{0,t+r} - \delta\overset{\bullet}{\mathsf{I}}_{0,t},(\cdot)_\lambda\Big]\Pi_{t\lambda}\Big)_\Lambda(0)\Big|^{q'}\\ &\leq \int \mathrm{d}x|\psi_\lambda(y)|\|\psi_\Lambda\|_{L^2}\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi_{t,t+r}(x-y) - \delta\Pi_{t,t+r}(x)|^q\Big\|_{L^2_x}\mathbb{E}^{\frac{1}{p}}|\Pi_{t\lambda}(0)|^p.\end{aligned} \tag{4.47}$$

We use the following estimate

$$\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi_{t,t+r}(x-y) - \delta\Pi_{t,t+r}(x)|^q\Big\|_{L^2_x} \lesssim |y|^{1-2\alpha}\sqrt{r}^{\alpha}\sqrt{t+r}^{\alpha}\overline{w} \tag{4.48}$$

for every $\alpha \in (0,\frac{1}{2})$, which itself is an interpolation[10] of the two estimates

$$\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi_{t,t+r}(x-y) - \delta\Pi_{t,t+r}(x)|^q\Big\|_{L^2_x} \lesssim |y|^{1-\beta}\sqrt{t+r}^{\beta}\overline{w}, \tag{4.49}$$

$$\Big\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi_{t,t+r}(x-y) - \delta\Pi_{t,t+r}(x)|^q\Big\|_{L^2_x} \lesssim \sqrt{r}^{1-\beta}\sqrt{t+r}^{\beta}\overline{w}, \tag{4.50}$$

---

[10] using $\beta = \alpha$ and $\frac{\alpha}{1-\alpha} + \frac{1-2\alpha}{1-\alpha}$

for every $\beta \in (0,1)$. Estimate (4.49) follows along the same lines as (4.43) using the triangle inequality. For (4.50) using again the triangle inequality, translation invariance and the semigroup property in the form

$$\delta\dot{|}_{0,t+r}(x) = \int \mathrm{d}z\, e^{-mr} \widetilde{H}_r(z)\delta\dot{|}_{0,t}(x-z) + \underbrace{\int_t^{t+r} \mathrm{d}s\, H_{t-s} * \delta\xi(s,x)}_{=:\delta\dot{|}_{t,t+r}},$$

where $\widetilde{H}_r$ stands for the massless heat kernel, we observe

$$\begin{aligned}\left\|\mathbb{E}^{\frac{1}{q}}|\delta\Pi_{t,t+r}(x-y) - \delta\Pi_{t,t+r}(x)|^q\right\|_{L^2_x} &\leq 2\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t+r} - \delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} \\ &\lesssim \int \mathrm{d}z\, H_r(z)\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t}(\cdot - z) - \delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} \\ &\quad + |e^{-mr}-1|\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} + \left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{t,t+r}\right|^q\right\|_{L^2_x} \\ &=: I_1 + I_2 + I_3.\end{aligned}$$

For I_1 using (4.43) we obtain

$$\int \mathrm{d}z\, H_r(z)\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t}(\cdot - z) - \delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} \lesssim \sqrt{r}^{1-\beta}\sqrt{t}^{\beta}\overline{w} \lesssim \sqrt{r}^{1-\beta}\sqrt{t+r}^{\beta}\overline{w}.$$

To estimate I_2 we use Young's inequality for convolution, the Cauchy–Schwarz inequality in the s-variable and the Hölder's inequality again in the s-variable to treat the integral of the exponential yielding

$$\begin{aligned}\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} &\leq \int_0^t \mathrm{d}s\, \|H_{t-s}\|_{L^1_x}\left\|\mathbb{E}^{\frac{1}{q}}|\delta\xi|^q\right\|_{L^2_x} \\ &\leq \left(\int_0^t \mathrm{d}s\, e^{-2m(t-s)}\right)^{\frac{1}{2}}\left(\int_0^t \mathrm{d}s\left\|\mathbb{E}^{\frac{1}{q}}|\delta\xi|^q\right\|^2_{L^2_x}\right)^{\frac{1}{2}} \lesssim \frac{1}{\sqrt{m}^{1-\beta}}\sqrt{t}^{\beta}\overline{w}\end{aligned} \tag{4.51}$$

for every $\beta \in [0,1)$. This in turn implies the estimate

$$|e^{-mr}-1|\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{0,t}\right|^q\right\|_{L^2_x} \lesssim |e^{-mr}-1|\frac{1}{\sqrt{m}^{1-\beta}}\sqrt{t}^{\beta}\overline{w} \lesssim \sqrt{r}^{1-\beta}\sqrt{t+r}^{\beta}\overline{w}, \tag{4.52}$$

where the implicit constant is uniform in m. To estimate I_3 we use (4.51) for $\beta = 0$ and a change of variables in s which leads to

$$\left\|\mathbb{E}^{\frac{1}{q}}\left|\delta\dot{|}_{t,t+r}\right|^q\right\|_{L^2_x} \lesssim \sqrt{r}\,\overline{w} \lesssim \sqrt{r}^{1-\beta}\sqrt{t+r}^{\beta}\overline{w}.$$

In total, (4.47) and (4.48) imply the estimate

$$\mathbb{E}^{\frac{1}{q'}}\left|\left(\left[\delta\dot{|}_{0,t+r} - \delta\dot{|}_{0,t}, (\cdot)_\lambda\right]\Pi_{t\lambda}\right)_\Lambda(0)\right|^{q'} \lesssim \Lambda^{-1}\lambda^{1-2\alpha}\sqrt{r}^{\alpha}\sqrt{t+r}^{\alpha}\mathbb{E}^{\frac{1}{p}}|\Pi_{t\lambda}(0)|^p\overline{w}.$$

Plugging in (4.38) for $\Pi_t \in \{\dot{|}_{0,t}, \ddot{\vee}_{0,t}\}$ and proceeding as in (4.46) gives

$$\mathbb{E}^{\frac{1}{q'}}\left|\left((\delta\dot{|}_{0,t+r} - \delta\dot{|}_{0,t})\Pi_t\right)_\lambda(0)\right|^{q'} \lesssim \lambda^{-(|\Pi|+2)\alpha}\sqrt{r}^{\alpha}\sqrt{t+r}^{(|\Pi|+1)\alpha}\overline{w}. \tag{4.53}$$

Combining (4.46) and (4.53) implies (4.41). □

Lemma 4.18. *For all $\alpha \in (0,1)$ the following estimate holds*

$$\int \mathrm{d}z\, |H_{t-s}(z-y) - H_{t-s}(z)| \lesssim e^{-m(t-s)} |y|^{\alpha} \sqrt{t-s}^{-\alpha}.$$

Proof. Interpolating the two estimates

$$\int \mathrm{d}z\, |H_{t-s}(z-y) - H_{t-s}(z)| \le 2\|H_{t-s}\|_{L^1_x} = 2e^{-m(t-s)}$$

and

$$\int \mathrm{d}z\, |H_{t-s}(z-y) - H_{t-s}(z)| \le \|\nabla H_{t-s}\|_{L^1_x} |y| \le e^{-m(t-s)} \sqrt{t-s}^{-1} |y|$$

yields the assertion. □

4.D. Estimates on the remainder

Lemma 4.19. *Let $\alpha > 0$ be sufficiently small. For every $p < \infty$*

$$\sup_{t \le 1} t^{\frac{1}{2}} \|v_{0,t}\|_{L^p_x} \le C,$$

where C depends polynomially on $\sup_{t\le 1} \|\<k>_{0,t}\|_{C^{-\alpha}}$ for $k = 1,2,3$ and is uniform in the initial condition f. In particular, C has finite moments of every order.

Proof. Follows from [**111**, Proposition 3.7]. The constant $c_{t,\infty}$ in Proposition 4.13 can be absorbed into the terms $\<2>_{0,t}$ and $\<3>_{0,t}$ which together with Proposition 4.13 yield

$$\begin{aligned}
&\sup_{t \le 1} t^{\alpha'} \|\<2>_{0,t} - c_{t,\infty}\|_{-\alpha} \lesssim \sup_{t \le 1} \|\<2>_{0,t}\|_{C^{-\alpha}}, \\
&\sup_{t \le 1} t^{\alpha'} \|\<3>_{0,t} - 3c_{t,\infty} \<1>_{0,t}\|_{-\alpha} \lesssim \max_{k=1,3} \sup_{t \le 1} \|\<k>_{0,t}\|_{C^{-\alpha}},
\end{aligned}$$

for any $\alpha' > 0$, allowing us to apply [**111**, Proposition 3.7]. □

Lemma 4.20. *Let $\alpha > 0$ be sufficiently small. Then for every $\kappa > 0$ sufficiently small the following estimate holds*

$$\sup_{t \le 1} t^{\frac{1}{2}+\kappa} \|v_{0,t}\|_{\kappa} \le C,$$

where $C \equiv$ depends polynomially on $\sup_{t\le 1} \|\<k>_{0,t}\|_{-\alpha}$ for $k = 1,2,3$ and is uniform in the initial condition f.

Proof. The statement follows essentially from the proof of Lemma 5.1 for $s = \frac{t}{2}$ in [**112**] working with $\<1>_{0,t}$, $\<2>_{0,t} - c_{t,\infty}$, $\<3>_{0,t} - 3c_{t,\infty} \<1>_{0,t}$ as we explained in the proof of Lemma 4.19. The terms I_6 and I_7 in the notation of [**112**, proof of Lemma 5.1] can be ignored. □

Lemma 4.21. *Let $\alpha > 0$ be sufficiently small. Then for any $\varepsilon > 0$ the following estimate holds*

$$\sup_{t\leq 1} t^{1+\varepsilon}\|\nabla v_{0,t}\|_{L^\infty_x} \leq C,$$

where C depends polynomially on $\sup_{t\leq 1}\|\Psi^{\langle k\rangle}_{0,t}\|_{-\alpha}$ for $k = 1, 2, 3$ and is uniform in the initial condition f.

Proof. To ease the notation we set $\eta := \max_{k=1,2,3}\sup_{t\leq 1}\|\Psi^{\langle k\rangle}_{0,t}\|_{-\alpha}$. By Duhamel's formula, we have

$$\begin{aligned}\|\nabla v_{0,t}\|_{L^\infty_x} \leq \Big\|\nabla H_{\frac{t}{2}} * v_{0,\frac{t}{2}}\Big\|_{L^\infty_x} + \sum_{k=0}^{3}\int_{\frac{t}{2}}^{t}\Big\|\nabla H_{t-r} * \Big(\Psi^{\langle k\rangle}_{0,r} v_{0,r}^{3-k}\Big)\Big\|_{L^\infty_x}\,\mathrm{d}r \\ + 3\int_{\frac{t}{2}}^{t} c_{r,\infty}\Big\|\nabla H_{t-r} * \Big(\Psi^{\langle 1\rangle}_{0,r} + v_{0,r}\Big)\Big\|_{L^\infty_x}\,\mathrm{d}r.\end{aligned}$$

Note in the following that $B^{\varepsilon}_{\infty,\infty}(\mathbb{T}^2) = C^{\varepsilon} \hookrightarrow L^\infty_x$ continuously for any $\varepsilon > 0$. Then, first of all, by Young's inequality and (4.19), we have

$$\Big\|\nabla H_{\frac{t}{2}} * v_{0,\frac{t}{2}}\Big\|_{L^\infty_x} \leq \Big\|\nabla H_{\frac{t}{2}}\Big\|_{L^{p'}_x}\Big\|v_{0,\frac{t}{2}}\Big\|_{L^p_x} \lesssim t^{-\frac{1}{2}-\frac{1}{p}} t^{-\frac{1}{2}} = t^{-1-\varepsilon}$$

for p large enough. In the same vain, using (4.19) and p large enough yields

$$\begin{aligned}\int_{\frac{t}{2}}^{t}\Big\|\nabla H_{t-r} * v_{0,r}^3\Big\|_{L^\infty_x}\,\mathrm{d}r \leq \int_{\frac{t}{2}}^{t}\|\nabla H_{t-r}\|_{L^{p'}_x}\Big\|v_{0,r}^3\Big\|_{L^p_x}\,\mathrm{d}r = \int_{\frac{t}{2}}^{t}\|\nabla H_{t-r}\|_{L^{p'}_x}\|v_{0,r}\|^3_{L^{3p}_x}\,\mathrm{d}r \\ \leq \int_{\frac{t}{2}}^{t}(t-r)^{-\frac{1}{2}-\frac{1}{p}} r^{-\frac{3}{2}}\,\mathrm{d}r \lesssim t^{-1-\varepsilon}.\end{aligned}$$

Moreover, using the semigroup property of the heat kernel and Young's inequality again we note

$$\begin{aligned}\int_{\frac{t}{2}}^{t}\Big\|\nabla H_{t-r} * \Psi^{\langle 3\rangle}_{0,r}\Big\|_{L^\infty_x}\,\mathrm{d}r &= \int_{\frac{t}{2}}^{t}\Big\|\nabla H_{\frac{t-r}{2}} * H_{\frac{t-r}{2}} * \Psi^{\langle 3\rangle}_{0,r}\Big\|_{L^\infty_x}\,\mathrm{d}r \\ &\leq \int_{\frac{t}{2}}^{t}\Big\|\nabla H_{\frac{t-r}{2}}\Big\|_{L^1_x}\Big\|H_{\frac{t-r}{2}} * \Psi^{\langle 3\rangle}_{0,r}\Big\|_{L^\infty_x}\,\mathrm{d}r.\end{aligned}$$

Moreover, by Lemma 4.15 we conclude

$$\int_{\frac{t}{2}}^{t}\Big\|\nabla H_{t-r} * \Psi^{\langle 3\rangle}_{0,r}\Big\|_{L^\infty_x}\,\mathrm{d}r \lesssim \int_{\frac{t}{2}}^{t}(t-r)^{-\frac{1}{2}}(t-r)^{-\alpha-\varepsilon}\,\mathrm{d}r \lesssim \eta t^{\frac{1}{2}-\alpha-\varepsilon}.$$

Similarly, using (4.20) and Lemma 4.15 we end up with

$$\begin{aligned}\int_{\frac{t}{2}}^{t}\Big\|\nabla H_{t-r} * \Big(v_{0,r}^2\Psi^{\langle 1\rangle}_{0,r}\Big)\Big\|_{L^\infty_x}\,\mathrm{d}r &\leq \int_{\frac{t}{2}}^{t}(t-r)^{-\frac{1}{2}-\alpha-\varepsilon}\Big\|v_{0,r}^2\Psi^{\langle 1\rangle}_{0,r}\Big\|_{-\alpha}\,\mathrm{d}r \\ &\leq \int_{\frac{t}{2}}^{t}(t-r)^{-\frac{1}{2}-\alpha-\varepsilon}\Big\|v_{0,r}^2\Big\|_{2\alpha}\Big\|\Psi^{\langle 1\rangle}_{0,r}\Big\|_{-\alpha}\,\mathrm{d}r\end{aligned}$$

$$\lesssim \eta \int_{\frac{t}{2}}^{t} (t-r)^{-\frac{1}{2}-\alpha-\varepsilon} r^{-1-2\alpha}\,\mathrm{d}r \lesssim \eta t^{-\frac{1}{2}-3\alpha-\varepsilon}$$

and in the same vain

$$\begin{aligned}\int_{\frac{t}{2}}^{t} \left\| \nabla H_{t-r} * \left(v_{0,r} \curlyvee_{0,r} \right) \right\|_{L_x^\infty} \mathrm{d}r &\lesssim \int_{\frac{t}{2}}^{t} (t-r)^{-\frac{1}{2}-\alpha-\varepsilon} \left\| v_{0,r} \curlyvee_{0,r} \right\|_{-\alpha} \mathrm{d}r \\ &\lesssim \eta \int_{\frac{t}{2}}^{t} (t-r)^{-\frac{1}{2}-\alpha-\varepsilon} r^{-\frac{1}{2}-\alpha}\,\mathrm{d}r \lesssim \eta t^{-2\alpha-\varepsilon}.\end{aligned}$$

Finally, using Lemma 4.15, (4.20) and (4.13) we get

$$\begin{aligned}\int_{\frac{t}{2}}^{t} c_{r,\infty} \left\| \uparrow_{0,r} + v_{0,r} \right\|_{L_x^\infty} \mathrm{d}r &\lesssim \int_{\frac{t}{2}}^{t} (t-r)^{-\frac{1}{2}-\alpha-\varepsilon} r^{-\gamma} \left\| v_{0,r} + \uparrow_{0,r} \right\|_{-\alpha} \mathrm{d}r \\ &\lesssim \eta t^{-2\alpha-2\varepsilon}.\end{aligned}$$

□

4.E. Proof of the Bakry–Émery identity

In [**105**] it was proved that

$$\mathcal{E}(F,F) := \int_{\mathcal{S}'(\mathbb{T}^2)} \|DF\|_{L_x^2}^2 \,\mathrm{d}\nu$$

where $F \in \mathcal{F}C_b^\infty$ is closable (see also [**3**]) and the closure gives rise to a quasi-regular Dirichlet form (cf. [**84**]), hence to a generator $\mathcal{L}$ with domain $D(\mathcal{L}) \subset D(\mathcal{E})$ such that

$$\mathcal{E}(F,F) = -\int_{\mathcal{S}'(\mathbb{T}^2)} F\mathcal{L}F\,\mathrm{d}\nu.$$

We denote by $\{\widetilde{P}_t\}_{t\geq 0}$ the associated semi-group. Then, by [**105**, Theorem 3.13], we infer that $P_tF = \widetilde{P}_tF$ ν-almost surely for all $F \in \mathcal{F}C_b^\infty$ and hence by continuity in time they are indistinguishable (see also [**71**, p. 67]). Moreover, by [**105**, Theorem 3.7] (and the discussion thereafter) $K := C^\infty(\mathbb{T}^2) \subset L^2(\mathbb{T}^2)$ is a a dense and linear subspace consisting of ν-admissible elements. Hence assumptions (C.1), (C.2) and (C.3) of [**3**, Section 4] are fulfilled. Moreover, $f \mapsto P_tF(f)$ is quasi-continuous for any $F \in \mathcal{F}C_b^\infty$. Now we can prove Proposition 4.12.

Proof of Proposition 4.12. Following [**72**, Proof of Theorem 1.1] we prove the ν-a.s. identity

$$\frac{\mathrm{d}}{\mathrm{d}s} P_{t-s}(P_sF)^2 = -2P_{t-s}\Big(\|DP_sF\|_{L_x^2}^2 \Big).$$

and use the same notation. Let $0 \leq r_1, r_2 \leq t$ and define $H(r_1,r_2) := P_{t-r_1}(P_{r_2}F)^2$.

By [**3**, p. 364, Theorem 4.3] and since $P_{r_2}F \in D(\mathcal{E})$ it holds that

$$P_{r_2}F(u_r^f) - P_{r_2}F(f) = \int_0^r \mathcal{L}(P_{r_2}F)(u_s^f)\,\mathrm{d}s + M_r$$

where M is a continuous martingale.
Moreover, by [**3**, p. 365, Proposition 4.5] the quadratic variation of M is given by

$$\langle M\rangle_r = \int_0^r \left\| DP_{r_2}F(u_s^f)\right\|_{L_x^2}^2 \mathrm{d}s$$

Then by Itô's formula [**103**, p. 222, Theorem 3.3] we compute

$$\begin{aligned}(P_{r_2}F)^2(u_r^f) = (P_{r_2}F)^2(f) + 2\int_0^r P_{r_2}F(u_s^f)\,\mathrm{d}M_s + 2\int_0^r P_{r_2}F(u_s^f)\mathcal{L}(P_{r_2}F)(u_s^f)\,\mathrm{d}s \\ + 2\int_0^r \|DP_{r_2}F(u_s^f)\|_{L_x^2}^2\,\mathrm{d}s\end{aligned}$$

and hence

$$\begin{aligned}P_{t-r_1}(P_{r_2}F)^2(f) = (P_{r_2}F)^2(f) + 2\int_0^{t-r_1} P_s(P_{r_2}F\mathcal{L}P_{r_2}F)(f)\,\mathrm{d}s \\ + 2\int_0^{t-r_1} P_s\|DP_{r_2}F\|_{L_x^2}^2(f)\,\mathrm{d}s.\end{aligned}$$

Then we see that on the one hand

$$\frac{\partial}{\partial r_1}P_{t-r_1}(P_{r_2}F)^2(f) = -2P_{t-r_1}(P_{r_2}F\mathcal{L}P_{r_2}F)(f) - 2P_{t-r_1}\|DP_{r_2}F\|_{L_x^2}^2(f)$$

and on the other hand we have

$$\frac{\partial}{\partial r_2}P_{t-r_1}(P_{r_2}F)^2(f) = 2P_{t-r_1}(P_{r_2}F\mathcal{L}P_{r_2}F)(f).$$

Continuity follows in the same vain as in [**72**, Proof of Theorem 1.1]. Finally, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}P_{t-s}(P_sF)^2 &= \frac{\partial}{\partial r_1}P_{t-r_1}(P_{r_2}F)^2(f)\Big|_{r_1=r_2=s} + \frac{\partial}{\partial r_2}P_{t-r_1}(P_{r_2}F)^2(f)\Big|_{r_1=r_2=s} \\ &= -2P_{t-r_1}\|DP_{r_2}F\|_{L_x^2}^2(f).\end{aligned}$$

Integrating from 0 to t proves the claim. □

4.F. Differentiability with respect to the initial data

We set

$$G(f,v)(t) := S(t)f + \int_0^t S(t-s)F(v_s)\,\mathrm{d}s - v_t$$

where $F(v_t) := -\Big(v_t^3 + 3v_t^2\<1>_t + 3v_t\<2>_t + \<3>_t - c_{t,\infty}(v_t + \<1>_t)\Big)$. By [**111**, Theorem 3.9] there exist fixed parameters $\gamma, \beta > 0$ such that for any $f^* \in C^{-\alpha_0}$ and $T > 0$ we can find a unique solution v^* to (4.13) satisfying $G(f^*, v^*)(t) = 0$, for every $0 \le t \le T$, and

$\sup_{0\le t\le T}(t\wedge 1)^\gamma \|v_t^*\|_\beta < \infty$. We define

$$X := \Big\{f\in C^{-\alpha_0} : \|f\|_{-\alpha_0}\le R\Big\},\quad Y := \Big\{v:[0,T]\to C^\beta : \sup_{0\le t\le T\wedge T^*} t^\gamma\|v_t\|_\beta \le 1\Big\}$$

for some T^* to be chosen below. Then again by [**111**, Theorem 3.9] we know that $G(f^*,v^*)(t) = 0$, for every $0\le t\le T\wedge T^*$. It is easy to check that G is Frechét-differentiable and we have

$$G_v(f^*,v^*)\delta v(t) = \int_0^t S(t-s)\big(F'(v_s)\delta v_s\big)\,\mathrm{d}s - \delta v_t =: (K-Id)\delta v_t$$

where $F'(v_t)\delta v_t := -3\Big(v_t^2 + 2v_t\dot{|}_t + \ddot{\vee}_t - c_{t,\infty}\Big)\delta v_t$. A simple calculation shows that

$$\|K\delta v_t\|_\beta \lesssim \int_0^t (t-s)^{-\frac{\alpha+\beta}{2}} s^{-\gamma}\|\delta v_s\|_\beta\,\mathrm{d}s \lesssim (T\wedge T^*)^{1-\frac{\alpha+\beta}{2}-\gamma}\sup_{0\le t\le T^*\wedge T'} t^\gamma\|\delta v_t\|_\beta.$$

Choosing T^* small enough such that the r.h.s. above is strictly smaller than 1 we get by the Neumann-series criterion that $G_v(f^*,v^*) : Y\to Y$ is a bijection. Hence by [**120**, Theorem 4.E] we get that $f\mapsto v^f$ is differentiable and its derivative in h is a mild solution to (4.20) on $(0,T\wedge T*]$. Concatenating this argument to cover the whole time interval $(0,T]$ proves the assertion.

Bibliography

[1] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: a new micro-macro passage. *Comm. Math. Phys.*, 307(3):791–815, 2011.

[2] S. Adams, N. Dirr, M. Peletier, and J. Zimmer. Large deviations and gradient flows. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 371(2005):20120341, 17, 2013.

[3] S. Albeverio and M. Röckner. Stochastic differential equations in infinite dimensions: solutions via Dirichlet forms. *Probab. Theory Related Fields*, 89(3):347–386, 1991.

[4] R. Almgren. Singularity formation in Hele-Shaw bubbles. *Phys. Fluids*, 8(2):344–352, 1996.

[5] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[6] H. Bahouri, J. Chemin, and R. Danchin. *Fourier analysis and nonlinear partial differential equations*, volume 343 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2011.

[7] D. Bakry. Functional inequalities for Markov semigroups. In *Probability measures on groups: recent directions and trends*, pages 91–147. Tata Inst. Fund. Res., Mumbai, 2006.

[8] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

[9] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.

[10] S. Bartels. *Numerical methods for nonlinear partial differential equations*, volume 47 of *Springer Series in Computational Mathematics*. Springer, Cham, 2015.

[11] R. Bauerschmidt and T. Bodineau. Log-Sobolev inequality for the continuum sine-Gordon model. *Comm. Pure Appl. Math.*, 74(10):2064–2113, 2021.

[12] R. Bauerschmidt and B. Dagallier. Log-sobolev inequality for the φ^4_2 and φ^4_3 measures, 2022.

[13] J. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[14] E. Beretta, M. Bertsch, and R. Dal Passo. Nonnegative solutions of a fourth-order nonlinear degenerate parabolic equation. *Arch. Rational Mech. Anal.*, 129(2):175–200, 1995.

[15] F. Bernis. Finite speed of propagation and continuity of the interface for thin viscous flows. *Adv. Differential Equations*, 1(3):337–368, 1996.

[16] F. Bernis and A. Friedman. Higher order nonlinear degenerate parabolic equations. *J. Differential Equations*, 83(1):179–206, 1990.

[17] A. L. Bertozzi and M. Pugh. The lubrication approximation for thin viscous films: regularity and long-time behavior of weak solutions. *Comm. Pure Appl. Math.*, 49(2):85–123, 1996.

[18] M. Bertsch, R. Dal Passo, H. Garcke, and G. Grün. The thin viscous flow equation in higher space dimensions. *Adv. Differential Equations*, 3(3):417–440, 1998.

[19] Y. Bruned, F. Gabriel, M. Hairer, and L. Zambotti. Geometric stochastic heat equations. *J. Amer. Math. Soc.*, 35(1):1–80, 2021.

[20] J. Carrillo, S. Lisini, G. Savaré, and D. Slepčev. Nonlinear mobility continuity equations and generalized displacement convexity. *J. Funct. Anal.*, 258(4):1273–1309, 2010.

[21] T. Cass, C. Litterer, and T. Lyons. Integrability and tail estimates for Gaussian rough differential equations. *Ann. Probab.*, 41(4):3026–3050, 2013.

[22] P. Cattiaux and A. Guillin. Semi log-concave Markov diffusions. In *Séminaire de Probabilités XLVI*, volume 2123 of *Lecture Notes in Math.*, pages 231–292. Springer, Cham, 2014.

[23] F. Cornalba. A priori positivity of solutions to a non-conservative stochastic thin-film equation. *arXiv preprint arXiv:1811.07826*, 2018.

[24] F. Cornalba and J. Fischer. The dean-kawasaki equation and the structure of density fluctuations in systems of diffusing particles. *arXiv preprint arXiv:2109.06500*, 2021.

[25] G. Da Prato and A. Debussche. Strong solutions to the stochastic quantization equations. *Ann. Probab.*, 31(4):1900–1916, 2003.

[26] G. Da Prato and A. Debussche. Gradient estimates and maximal dissipativity for the Kolmogorov operator in Φ^4_2. *Electron. Commun. Probab.*, 25:Paper No. 9, 16, 2020.

[27] R. Dal Passo, H. Garcke, and G. Grün. On a fourth-order degenerate parabolic equation: global entropy estimates, existence, and qualitative behavior of solutions. *SIAM J. Math. Anal.*, 29(2):321–342, 1998.

[28] R. Dal Passo, L. Giacomelli, and G. Grün. A waiting time phenomenon for thin film equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 30(2):437–463, 2001.

[29] K. Dareiotis, B. Gess, M. Gnann, and G. Grün. Non-negative martingale solutions to the stochastic thin-film equation with nonlinear gradient noise. *arXiv e-prints*, page arXiv:2012.04356, December 2020.

[30] B. Davidovitch, E. Moro, and H. Stone. Spreading of viscous fluid drops on a solid substrate assisted by thermal fluctuations. *Physical review letters*, 95(24):244505, 2005.

[31] D. Dawson and J. Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[32] D. Dean. Langevin equation for the density of a system of interacting Langevin processes. *J. Phys. A*, 29(24):L613–L617, 1996.

[33] J. Deuschel and G. Giacomin. Entropic repulsion for massless fields. *Stochastic Process. Appl.*, 89(2):333–354, 2000.

[34] N. Dirr, M. Stamatakis, and J. Zimmer. Entropic and gradient flow formulations for nonlinear diffusion. *Journal of Mathematical Physics*, 57(8):081505, 2016.

[35] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009.

[36] M. Duerinckx and F. Otto. Higher-order pathwise theory of fluctuations in stochastic homogenization. *Stoch. Partial Differ. Equ. Anal. Comput.*, 8(3):625–692, 2020.

[37] M. Durán-Olivencia, R. Gvalani, S. Kalliadasis, and G. Pavliotis. Instability, rupture and fluctuations in thin liquid films: theory and computations. *J. Stat. Phys.*, 174(3):579–604, 2019.

[38] M. Erbar, K. Kuwada, and K.T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.

[39] L. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[40] B. Fehrman and B. Gess. Large deviations for conservative stochastic pde and non-equilibrium fluctuations. *arXiv preprint arXiv:1910.11860*, 2019.

[41] B. Fehrman and B. Gess. Well-posedness of the dean-kawasaki and the nonlinear dawson-watanabe equation with correlated noise. *arXiv preprint arXiv:2108.08858*, 2021.

[42] J. Fischer. Optimal lower bounds on asymptotic support propagation rates for the thin-film equation. *J. Differential Equations*, 255(10):3127–3149, 2013.

[43] J. Fischer. Upper bounds on waiting times for the thin-film equation: the case of weak slippage. *Arch. Ration. Mech. Anal.*, 211(3):771–818, 2014.

[44] J. Fischer and G. Grün. Existence of positive solutions to stochastic thin-film equations. *SIAM J. Math. Anal.*, 50(1):411–455, 2018.

[45] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984.

[46] P. Friz and M. Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[47] B. Gess and M. Gnann. The stochastic thin-film equation: existence of nonnegative martingale solutions. *Stochastic Process. Appl.*, 130(12):7260–7302, 2020.

[48] B. Gess, R. Gvalani, F. Kunick, and F. Otto. Thermodynamically consistent and positivity-preserving discretization of the thin-film equation with thermal noise. *arXiv preprint arXiv:2109.06083*, 2021.

[49] L. Giacomelli, H. Knüpfer, and F. Otto. Smooth zero-contact-angle solutions to a thin-film equation around the steady state. *J. Differential Equations*, 245(6):1454–1506, 2008.

[50] L. Giacomelli and F. Otto. Rigorous lubrication approximation. *Interfaces Free Bound.*, 5(4):483–529, 2003.

[51] A. Gloria and F. Otto. An optimal variance estimate in stochastic homogenization of discrete elliptic equations. *Ann. Probab.*, 39(3):779–856, 2011.

[52] G. Grün. Degenerate parabolic differential equations of fourth order and a plasticity model with non-local hardening. *Z. Anal. Anwendungen*, 14(3):541–574, 1995.

[53] G. Grün, K. Mecke, and M. Rauscher. Thin-film flow influenced by thermal noise. *J. Stat. Phys.*, 122(6):1261–1291, 2006.

[54] G. Grün and M. Rumpf. Nonnegativity preserving convergent schemes for the thin film equation. *Numer. Math.*, 87(1):113–152, 2000.

[55] M. Gubinelli. Controlling rough paths. *J. Funct. Anal.*, 216(1):86–140, 2004.

[56] M. Gubinelli and M. Hofmanová. Global solutions to elliptic and parabolic Φ^4 models in Euclidean space. *Comm. Math. Phys.*, 368(3):1201–1266, 2019.

[57] M. Gubinelli and M. Hofmanová. A PDE construction of the Euclidean ϕ^4_3 quantum field theory. *Comm. Math. Phys.*, 384(1):1–75, 2021.

[58] M. Gubinelli, P. Imkeller, and N. Perkowski. Paracontrolled distributions and singular PDEs. *Forum Math. Pi*, 3:e6, 75, 2015.

[59] M. Hairer. A theory of regularity structures. *Invent. Math.*, 198(2):269–504, 2014.

[60] M. Hairer, K. Lê, and T. Rosati. The allen-cahn equation with generic initial datum. *arXiv preprint arXiv:2201.08426*, 2022.

[61] M. Hairer and J. Maas. A spatial version of the Itô-Stratonovich correction. *Ann. Probab.*, 40(4):1675–1714, 2012.

[62] M. Hairer, J. Maas, and H. Weber. Approximating rough stochastic PDEs. *Comm. Pure Appl. Math.*, 67(5):776–870, 2014.

[63] M. Hairer and K. Matetski. Discretisations of rough stochastic PDEs. *Ann. Probab.*, 46(3):1651–1709, 2018.

[64] M. Hairer and J. Mattingly. The strong Feller property for singular stochastic PDEs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1314–1340, 2018.

[65] M. Hairer and P. Schönbauer. The support of singular stochastic partial differential equations. *Forum Math. Pi*, 10:Paper No. e1, 127, 2022.

[66] M. Hairer and H. Weber. Large deviations for white-noise driven, nonlinear stochastic PDEs in two and three dimensions. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(1):55–92, 2015.

[67] R. Ignat, F. Otto, T. Ried, and P. Tsatsoulis. Variational methods for a singular spde yielding the universality of the magnetization ripple, 2020.

[68] N. Ikeda and S. Watanabe. *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second edition, 1989.

[69] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[70] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1988.

[71] H. Kawabi. The parabolic Harnack inequality for the time dependent Ginzburg-Landau type SPDE and its application. *Potential Anal.*, 22(1):61–84, 2005.

[72] H. Kawabi. A simple proof of log-Sobolev inequalities on a path space with Gibbs measures. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 9(2):321–329, 2006.

[73] H. Knüpfer and N. Masmoudi. Darcy's flow with prescribed contact angle: well-posedness and lubrication approximation. *Arch. Ration. Mech. Anal.*, 218(2):589–646, 2015.

[74] R. Kohn, F. Otto, Maria G. Reznikoff, and E. Vanden-Eijnden. Action minimization and sharp-interface limits for the stochastic Allen-Cahn equation. *Comm. Pure Appl. Math.*, 60(3):393–438, 2007.

[75] V. Konarovskyi, T. Lehmann, and M. von Renesse. Dean-Kawasaki dynamics: ill-posedness vs. triviality. *Electron. Commun. Probab.*, 24:Paper No. 8, 9, 2019.

[76] N. V. Krylov. *Lectures on elliptic and parabolic equations in Hölder spaces*, volume 12 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1996.

[77] F. Kunick. A first order description of a nonlinear spde in the spirit of rough paths. *arXiv preprint arXiv:2202.01127*, 2022.

[78] F. Kunick and P. Tsatsoulis. Gradient-type estimates for the dynamic φ_2^4-model. *arXiv preprint arXiv:2202.11036*, 2022.

[79] G. M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific Publishing Co. Inc., River Edge, 1996.

[80] E. M. Lifshitz and L. P. Pitaevskii. *Statistical Physics. Theory of the Condensed State*. 1980.

[81] P. Linares, F. Otto, M. Tempelmayr, and P. Tsatsoulis. A diagram-free approach to the stochastic estimates in regularity structures, 2021.

[82] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[83] T. Lyons. Differential equations driven by rough signals. *Rev. Mat. Iberoamericana*, 14(2):215–310, 1998.

[84] Z. M. Ma and M. Röckner. *Introduction to the theory of (nonsymmetric) Dirichlet forms*. Universitext. Springer-Verlag, Berlin, 1992.

[85] K. Matetski, J. Quastel, and D. Remenik. The KPZ fixed point. *Acta Math.*, 227(1):115–203, 2021.

[86] S. Metzger and G. Grün. Existence of nonnegative solutions to stochastic thin-film equations in two space dimensions, 2021.

[87] A. Mielke, M. A. Peletier, and D. R. M. Renger. On the relation between gradient flows and the large-deviation principle, with applications to Markov chains and diffusion. *Potential Anal.*, 41(4):1293–1327, 2014.

[88] L. Modica and S. Mortola. Un esempio di Γ^--convergenza. *Boll. Un. Mat. Ital. B (5)*, 14(1):285–299, 1977.

[89] A. Moinat and H. Weber. Space-time localisation for the dynamic Φ^4_3 model. *Comm. Pure Appl. Math.*, 73(12):2519–2555, 2020.

[90] J.-C. Mourrat and H. Weber. The dynamic Φ^4_3 model comes down from infinity. *Comm. Math. Phys.*, 356(3):673–753, 2017.

[91] J.-C. Mourrat and H. Weber. Global well-posedness of the dynamic Φ^4 model in the plane. *Ann. Probab.*, 45(4):2398–2476, 2017.

[92] E. Nelson. The free Markoff field. *J. Functional Analysis*, 12:211–227, 1973.

[93] B. Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[94] F. Otto. Lubrication approximation with prescribed nonzero contact angle. *Comm. Partial Differential Equations*, 23(11-12):2077–2164, 1998.

[95] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[96] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[97] F. Otto and H. Weber. Quasi-linear spdes in divergence form. *Stochastics and Partial Differential Equations: Analysis and Computations*, 7(1):64–85, 2019.

[98] F. Otto and H. Weber. Quasilinear spdes via rough paths. *Archive for Rational Mechanics and Analysis*, 232(2):873–950, 2019.

[99] G. Parisi and Y. S. Wu. Perturbation theory without gauge fixing. *Sci. Sinica*, 24(4):483–496, 1981.

[100] G. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics*. Springer, New York, 2014. Diffusion processes, the Fokker-Planck and Langevin equations.

[101] M. Peletier. Energies, gradient flows, and large deviations: a modelling point of view, 2012.

[102] N. Perkowski. Paracontrolled distributions and singular distributions. *unpublished lecture notes taught at the Hausdorff center in Bonn*, 2017.

[103] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[104] M. Röckner, R. Zhu, and X. Zhu. Ergodicity for the stochastic quantization problems on the 2D-torus. *Comm. Math. Phys.*, 352(3):1061–1090, 2017.

[105] M. Röckner, R. Zhu, and X. Zhu. Restricted Markov uniqueness for the stochastic quantization of $P(\Phi)_2$ and its applications. *J. Funct. Anal.*, 272(10):4263–4303, 2017.

[106] M. Sauerbrey. Martingale solutions to the stochastic thin-film equation in two dimensions, 2021.

[107] S. Sheffield. Gaussian free fields for mathematicians. *Probab. Theory Related Fields*, 139(3-4):521–541, 2007.

[108] L. Simon. Schauder estimates by scaling. *Calc. Var. Partial Differential Equations*, 5(5):391–407, 1997.

[109] K. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[110] K. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.

[111] P. Tsatsoulis and H. Weber. Spectral gap for the stochastic quantization equation on the 2-dimensional torus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1204–1249, 2018.

[112] P. Tsatsoulis and H. Weber. Exponential loss of memory for the 2-dimensional Allen-Cahn equation with small noise. *Probab. Theory Related Fields*, 177(1-2):257–322, 2020.

[113] K. Twardowska and A. Nowak. On the relation between the Itô and Stratonovich integrals in Hilbert spaces. *Ann. Math. Sil.*, (18):49–63, 2004.

[114] S. Varadhan. *Large deviations and applications*, volume 46 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984.

[115] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[116] M. von Renesse and K. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[117] M. von Renesse and K. Sturm. Entropic measure and Wasserstein diffusion. *Ann. Probab.*, 37(3):1114–1191, 2009.

[118] L. Zambotti. A reflected stochastic heat equation as symmetric dynamics with respect to the 3-d Bessel bridge. *J. Funct. Anal.*, 180(1):195–209, 2001.

[119] L. Zambotti. A conservative evolution of the Brownian excursion. *Electron. J. Probab.*, 13:no. 37, 1096–1119, 2008.

[120] E. Zeidler. *Applied functional analysis*, volume 109 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1995. Main principles and their applications.

[121] L. Zhornitskaya and A. L. Bertozzi. Positivity-preserving numerical schemes for lubrication-type equations. *SIAM J. Numer. Anal.*, 37(2):523–555, 2000.

[122] H. Öttinger. Beyond equilibrium thermodynamics. 2005.

Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, den 25. März 2022

. . . .
(Florian Johannes Kunick)

Daten zum Autor

Name:	Florian Johannes Kunick
Geburtsdatum:	03.04.1994 in Memmingen
10/2013 - 10/2016	Bachelor Studium in Mathematik am Karlsruher Institut für Technologie
10/2016 - 09/2018	Master Studium in Mathematik am Karlsruher Institut für Technologie
seit 10/2018	Doktorand am Max-Planck-Institut für Mathematik in den Naturwissenschaften betreut von Prof. Dr. Benjamin Gess und Prof. Dr. Felix Otto