

LONG-TIME BEHAVIOR OF A FINITE VOLUME DISCRETIZATION FOR A FOURTH ORDER DIFFUSION EQUATION

JAN MAAS AND DANIEL MATTHES

ABSTRACT. We consider a non-standard finite-volume discretization of a strongly non-linear fourth order diffusion equation on the d -dimensional cube, for arbitrary $d \geq 1$. The scheme preserves two important structural properties of the equation: the first is the interpretation as a gradient flow in a mass transportation metric, and the second is an intimate relation to a linear Fokker-Planck equation. Thanks to these structural properties, the scheme possesses two discrete Lyapunov functionals. These functionals approximate the entropy and the Fisher information, respectively, and their dissipation rates converge to the optimal ones in the discrete-to-continuous limit. Using the dissipation, we derive estimates on the long-time asymptotics of the discrete solutions. Finally, we present results from numerical experiments which indicate that our discretization is able to capture significant features of the complex original dynamics, even with a rather coarse spatial resolution.

1. INTRODUCTION

1.1. **The QDD equation.** In this note, we introduce and analyze a particular spatial discretization of the following non-linear parabolic equation of fourth order,

$$(1) \quad \partial_t u = -\nabla \cdot \left(u \nabla \left[\frac{\Delta u}{u} + \Delta \log u \right] \right) + \nabla \cdot (u \nabla W), \quad u = u(t; x) > 0, \quad t > 0, \quad x \in \Omega := [0, 1]^d,$$

subject to variational boundary conditions, see (5) below. The potential $W : \Omega \rightarrow \mathbb{R}$ is assumed to satisfy certain structural conditions (4) and (15); a possible choice is $W(x) = \lambda^*/2|x - \bar{x}|^2$ for arbitrary $\lambda^* > 0$ and $\bar{x} \in \mathbb{R}^d$

Equation (1), which is referred to as *Quantum drift diffusion (QDD) equation* or as *Derrida-Lebowitz-Speer-Spohn (DLSS) equation* in the literature, appears, e.g., in semi-conductor modelling [10, 27] and in the analysis of interface motion in spin systems [12, 13]. Depending on the context, the non-linear term is written in one of several equivalent forms:

$$\nabla \cdot \left(u \nabla \left[\frac{\Delta u}{u} + \Delta \log u \right] \right) = 2 \nabla \cdot \left(u \nabla \left(\frac{\Delta \sqrt{u}}{\sqrt{u}} \right) \right) = \nabla^2 : (u \nabla^2 \log u).$$

Existence and qualitative properties of (weak) solutions to (1) have been intensively analyzed in the past two decades [3, 21, 28, 30, 31, 35]. For instance, it has been proven — see [21] for the most comprehensive result — that the initial boundary value problem for (1)&(5) possesses a non-negative and mass preserving global weak solution $u : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ for all non-negative initial conditions $u_0 \in L^1(\Omega)$ of finite entropy. By scaling invariance, we may assume without loss of generality in the following that the solution u is a time-dependent probability density.

Date: February 10, 2016.

2010 *Mathematics Subject Classification.* Primary: 65M08 Secondary: 35K35 53C21 60J25.

This research was supported by the DFG Collaborative Research Centers TRR 109, “Discretization in Geometry and Dynamics” and 1060 “The Mathematics of Emergent Effects”.

Several (semi-)discrete approximations of (1) have been studied, both analytically and numerically. The schemes presented in [4, 6, 15, 29, 36] inherit *some* structural properties of (1), like monotonicity of certain quantities. All of these schemes have in common that they provide non-negative (semi-)discrete solutions.

Here we continue in the spirit of [36], where a discretization was performed on grounds of (1)'s gradient flow structure with respect to the L^2 -Wasserstein metric, which leads to a scheme that simultaneously preserves two essential Lyapunov functionals. From these Lyapunov functionals, estimates on the fully discrete solutions were derived and have been used to analyze their long-time asymptotics [41] and the discrete-to-continuous limit [36].

However, here we do not use the Lagrangian structure behind (1) — which was essential in [36, 41] — but define a scheme on grounds of a finite-volume discretization. Our ansatz is motivated by a particular structure-preserving discretization of linear Fokker-Planck equations, which has been introduced simultaneously in [8, 33, 39]. Using this “Eulerian approach”, we overcome the limitation of [36, 41] to $d = 1$ space dimension. The similarities with [36, 41] are that we rely on the gradient flow formulation of (1), and that we design the discretization in such a way that enforces monotonicity of two Lyapunov functionals. We remark that the general idea to preserve simultaneous monotonicity of several functionals in the discretization has been used for other equations before, like in the context of the formally similar thin film equations, see [24, 25, 44].

1.2. Structural properties and long-time asymptotics. Most of the qualitative results for (1) are based on two fundamental structural properties: the first is its gradient flow structure with respect to the L^2 -Wasserstein metric [21], and the second is an intimate relation to a certain Fokker-Planck equation [11]. That Fokker-Planck equation has the form

$$(2) \quad \partial_s v_s = \Delta v_s + \nabla \cdot (v_s \nabla V) \quad \text{in } \Omega, \quad \partial_\nu (v_s / \pi) = 0 \quad \text{on } \partial\Omega,$$

where $\pi : \Omega \rightarrow \mathbb{R}_+$ is given by

$$(3) \quad \pi(x) = \frac{1}{Z} e^{-V(x)} \quad \text{with} \quad Z = \int_{\Omega} e^{-V(x')} dx'$$

and defines the unique stationary probability density π for (2). To establish the connection between (1) and (2), we shall assume henceforth that the respective potentials V and W are related via

$$(4) \quad W = |\nabla V|^2 - 2\Delta V,$$

and that V is λ -convex with some positive λ , i.e., $\nabla^2 V \geq \lambda > 0$. Notice that $V(x) = \frac{\lambda}{2}|x - \bar{x}|^2$ is an admissible choice, and leads to $W(x) = \lambda^2|x - \bar{x}|^2 - 2d\lambda$.

A direct computation shows that π is a stationary solution to (1) as well, provided the boundary conditions are chosen appropriately:

$$(5) \quad \partial_\nu (u/\pi) = \partial_\nu \left(\frac{\Delta u}{u} + \Delta \log u + W \right) = 0 \quad \text{on } \partial\Omega.$$

Another formal computation reveals that (1) and (2) have two Lyapunov functionals in common, namely the relative logarithmic entropy \mathcal{H} and the relative Fisher information \mathcal{I} , given by

$$(6) \quad \mathcal{H}(w) = \int_{\Omega} w \log(w/\pi) dx \quad \text{and} \quad \mathcal{I}(w) = \int_{\Omega} w |\nabla \log(w/\pi)|^2 dx.$$

In fact, both (2) and (1) are gradient flows — for \mathcal{H} and for \mathcal{I} , respectively — in the L^2 -Wasserstein metric. That is, formally, we can write

$$(7) \quad \partial_s v = -\mathbf{K}_v D_v \mathcal{H} \quad \text{and} \quad \partial_t u = -\mathbf{K}_u D_u \mathcal{I},$$

respectively, where \mathbf{K} is the Onsager operator (inverse metric tensor) of the Wasserstein metric,

$$\mathbf{K}_u \xi = -\nabla \cdot (u \nabla \xi).$$

The final but most important connection between (2) and (1) is the following relation between the respective potentials of the two gradient flows:

$$(8) \quad \mathcal{I}(v) = D_v \mathcal{H}[\mathbf{K}_v D_v \mathcal{H}].$$

That is, the potential of the gradient flow (1) is the dissipation of the entropy \mathcal{H} along its own gradient flow. Despite the fact that the representation (8) of \mathcal{I} is classical, implications on the dynamics of the fourth order equation (1) have been drawn only recently in [11], see also [7, 35, 38].

It turns out [11, 35] that the rate of convergence to equilibrium of (1) is intimately related to the corresponding rate of (2). We summarize the relevant estimates. Thanks to the λ -convexity of V , it follows that both \mathcal{H} and \mathcal{I} decay with exponential rate λ along solutions v to (2),

$$(9) \quad \mathcal{H}(v_s) \leq \mathcal{H}(v_{s'}) e^{-2\lambda(s-s')} \quad \text{and} \quad \mathcal{I}(v_s) \leq \mathcal{I}(v_{s'}) e^{-2\lambda(s-s')} \quad \text{for all } s \geq s' \geq 0,$$

and that the Fisher information can be estimated just in terms of the initial value of the entropy,

$$(10) \quad \mathcal{I}(v_s) \leq \mathcal{H}(v_0) s^{-1} \quad \text{for all } s > 0.$$

With V and W related by (4), the following analogous estimate can be shown for solutions to the QDD equation (1):

$$(11) \quad \mathcal{H}(u_t) \leq \mathcal{H}(u_{t'}) e^{-(2\lambda)^2(t-t')} \quad \text{and} \quad \mathcal{I}(u_t) \leq \mathcal{I}(u_{t'}) e^{-(2\lambda)^2(t-t')} \quad \text{for all } t \geq t' \geq 0,$$

$$(12) \quad \mathcal{I}(u_t) \leq \mathcal{H}(u_0) (2\lambda t)^{-1} \quad \text{for all } t > 0.$$

We review the derivation of (9)–(12) in Section 2.4.

1.3. Discretization and main result. The leading principle for our spatial discretization of (1) is that the semi-discrete solutions to that scheme inherit the estimates in (11) and (12). We discretize (1) and (2) simultaneously in order to preserve their close relation. For the discretization of (2), we follow an approach based on the entropy gradient flow structure for Markov chains developed in [8, 33, 39], which has been subsequently applied in [14, 17, 22, 40].

Our finite volume discretization is performed with respect to a regular cubic lattice: fix a box length $h = 1/N$ with $N \in \mathbb{N}$ and consider piecewise constant probability densities u^h on the equi-distant subdivision of Ω in N^d sub-cubes of side length h . Now, we replace (7) by

$$(13) \quad \partial_s v^h = -\mathbf{K}_v^h D_v \mathcal{H}^h \quad \text{and} \quad \partial_t u^h = -\mathbf{K}_u^h D_u \mathcal{I}^h,$$

respectively, where the discretized entropy \mathcal{H}^h is given (up to an additive constant $\gamma^h > 0$ defined in (47)) by the restriction of \mathcal{H} , and the discretized Fisher information \mathcal{I}^h is obtained by the relation (8), i.e.,

$$(14) \quad \mathcal{H}^h(u^h) = \mathcal{H}(u^h) - \gamma^h, \quad \mathcal{I}^h(u^h) = D_{u^h} \mathcal{H}^h[\mathbf{K}_{u^h} D_{u^h} \mathcal{H}^h].$$

The discretized Onsager operator \mathbf{K}^h — which implicitly determines a metric on the piecewise constant density functions — is designed such that the gradient flow of \mathcal{H}^h is the forward equation for a continuous time Markov chain. The appropriate and rather non-obvious choice for \mathbf{K}^h , see (48), was independently found in [33] and in [39].

For the main result that we formulate below we need an additional hypothesis on the potential V , namely that

$$(15) \quad V(x) = V^{[1]}(x_1) + \dots + V^{[d]}(x_d)$$

for suitable functions $V^{[k]} : [0, 1] \rightarrow \mathbb{R}$, which, by definition of π in (3), is equivalent to the following factorization of the steady state:

$$(16) \quad \pi(x) = \pi^{[1]}(x_1) \cdots \pi^{[d]}(x_d), \quad \text{where} \quad \pi^{[k]}(x) = \frac{1}{Z^{[k]}} e^{-V^{[k]}(x)},$$

with suitable normalization constants $Z^{[k]} > 0$ such that the $\pi^{[1]}, \dots, \pi^{[d]}$ are probability densities on $[0, 1]$. Under the discretization, π is replaced by a particular piecewise constant approximation π^h , which is the unique minimizer of \mathcal{H}^h , see Lemma 1. The approximation π^h still factors in the same form as above, see (46).

Theorem 1. *Assume that a pair of potentials V, W satisfying the relation (4) and the technical hypothesis (15) is given. Assume further that V is λ -convex with some $\lambda > 0$.*

For a given discretization parameter $h > 0$, define discretized entropy and Fisher information as in (14), and a discrete Onsager operator as in (48). Then any solution u^h of the discrete gradient flow

$$(17) \quad \partial_t u^h = -\mathbf{K}_{u^h}^h D_{u^h} \mathcal{I}^h$$

satisfies the following analogues of (11) and (12),

$$(18) \quad \mathcal{H}^h(u_t^h) \leq \mathcal{H}^h(u_{t'}^h) e^{-(2\lambda^h)^2(t-t')} \quad \text{and} \quad \mathcal{I}^h(u_t^h) \leq \mathcal{I}^h(u_{t'}^h) e^{-(2\lambda^h)^2(t-t')} \quad \text{for all } t \geq t' \geq 0,$$

$$(19) \quad \mathcal{I}^h(u_t^h) \leq \mathcal{H}^h(u_0) (2\lambda^h t)^{-1} \quad \text{for all } t > 0.$$

Consequently, u_t^h approaches the equilibrium π^h exponentially fast,

$$(20) \quad \|u_t^h - \pi^h\|_{L^1(\Omega)} \leq \sqrt{2\mathcal{H}^h(u_0^h)} e^{-2(\lambda^h)^2 t}.$$

Above, $\lambda^h = \lambda + O(h^2)$ as $h \downarrow 0$.

We conjecture that the result remains true on more general domains than $\Omega = [0, 1]^d$, and even for (at least slightly) more flexible discretizations than perfect square lattices. These restrictions come with our particular method of proof, which combines the one-dimensional estimates from [40] with a tensorisation result from [16], and thus requires a product structure of the state space.

1.4. Geodesic convexity vs. convex entropy decay. At the basis for proving our main Theorem 1 on the long-time asymptotics of the discretized DLSS equation (17) is the thorough understanding of the long-time asymptotics of the discretized Fokker-Planck equation (13). The latter is the gradient flow of the geodesically λ -convex discrete entropy functional \mathcal{H}^h on a (finite-dimensional) metric space, and thus possesses a variety of nice properties. We exploit that some of these properties are inherited by the discretized DLSS flow — which is a gradient flow in the same metric, but of the *non-convex* discrete Fisher information functional \mathcal{I}^h .

More precisely, our proof of Theorem 1 does not require to use the full power of λ -convexity. Instead, we work with a weaker property that we call *convex decay inequality*, see (CDI) in Section 2.2. In a nutshell, the difference is that we do not require the Hessian of the functional to be larger or equal to λ in *every* direction, but only in the direction of the functional's own gradient, at each given point.

This weaker form of convexity has been used in numerous places and in various disguises for the derivation of rates of convergence to equilibrium, typically in connection with the Bakry-Émery method, see e.g. [34] and references therein. Recently, an adaptation of this convexity concept to Markov chains has been developed in [5]. There are examples — see Remark 6 — where the optimal constant improves (slightly) upon relaxation from geodesic convexity to convex decay.

A key technical ingredient in the proof of our main result is the tensorization property of the convex decay inequality. This result is given in Section 3.4, and might be of independent interest.

Finally, we emphasize that there are various different but related techniques for proving mesh-uniform exponential convergence to equilibrium in discretized Fokker-Planck and more general diffusion equations. See, for instance, the recent work on reaction-diffusion systems in [20, 23]: in these papers, long-time asymptotics have been derived for the relative free energy functional associated with finite-volume discretizations on grounds of log-Sobolev type estimates. These estimates appear more flexible and easier to generalize than the ones obtained in the framework of convex gradient flows [8, 33, 39]. However, in the course of establishing Theorem 1, it is of crucial importance to work with the convex entropy decay inequality (which compares first- and second-order derivatives of the entropy) instead of the (weaker) log-Sobolev inequality (which compares zeroth- and first-order derivatives).

1.5. Plan of the paper. In Section 2 below, we review the basic results from the general theory of gradient flows and the Bakry-Émery method which are relevant for the study of our equation (1) and our discretization. Sections 3 and 4 are devoted to discretizations. In Section 3, we analyze the properties of a finite-volume discretization for the linear Fokker-Planck equation (2) in the spirit of [33, 39]. In Section 4, we define a “compatible” discretization of the QDD equation and prove the main result Theorem 1. We conclude by discretizing in time as well, and performing a series of numerical experiments in dimension $d = 2$ to illustrate the (non-)optimality of the theoretical decay estimates.

2. ESTIMATES FOR λ -CONVEX GRADIENT FLOWS

In this section, we shall mainly collect and rephrase classical and recent results about the large-time behavior of gradient flows. Throughout this section, we assume smoothness of all appearing analytical structures. These smoothness assumptions are justified in the analysis of the discretizations in Sections 3&4 below, provided that one restricts to strictly positive probability densities. The application to solutions of the original evolution equation (1), however, are purely formal and only serve as a motivation.

2.1. λ -convex gradient flows. Let a smooth Riemannian manifold \mathfrak{M} with metric d be given. For simplicity we assume that \mathfrak{M} is an open subset of a (finite-dimensional) affine space \mathcal{X} . At each point $u \in \mathfrak{M}$, there is a one-to-one correspondence between the scalar product $\langle \cdot, \cdot \rangle_u$ on $T_u\mathfrak{M}$ and the Onsager operator $\mathbf{K}_u : T_u^*\mathfrak{M} \rightarrow T_u\mathfrak{M}$, which is the uniquely determined linear isomorphism with

$$\langle \mathbf{K}_u p, \xi \rangle_u = p[\xi] \quad \text{for all } \xi \in T_u\mathfrak{M} \text{ and } p \in T_u^*\mathfrak{M}.$$

Note that \mathbf{K}_u is symmetric, in the sense that

$$p_1[\mathbf{K}_u p_2] = \langle \mathbf{K}_u p_1, \mathbf{K}_u p_2 \rangle_u = \langle \mathbf{K}_u p_2, \mathbf{K}_u p_1 \rangle_u = p_2[\mathbf{K}_u p_1] \quad \text{for all } p_1, p_2 \in T_u^*\mathfrak{M}.$$

In the application discussed here, the Onsager operator (and not the scalar product) will be the given quantity. In fact, in our application, the Onsager operator extends continuously to the boundary of \mathfrak{M} in \mathcal{X} , while the Riemannian metric degenerates at the boundary.

The gradient flow of a given smooth potential $\Phi : \mathfrak{M} \rightarrow \mathbb{R}$ is then defined as (solution to) the differential equation

$$(21) \quad \dot{u} = F_\Phi(u) := -\mathbf{K}_u D_u \Phi.$$

By smoothness of Φ , local solutions $u : [0, T) \rightarrow \mathfrak{M}$ to (21) exist for any initial condition $u_0 \in \mathfrak{M}$, and the only possible obstruction to global existence is that u leaves \mathfrak{M} at time $T > 0$.

A central notion in the theory is that of λ -convexity of Φ (with $\lambda \in \mathbb{R}$), which means that $\text{Hess } \Phi \geq \lambda$, where the Hessian is to be understood in the Riemannian structure of the \mathfrak{M} , and the inequality holds in the sense of quadratic forms:

$$\text{Hess}_u \Phi[\xi, \xi] \geq \lambda \|\xi\|_u^2 \quad \text{for all } u \in \mathfrak{M} \text{ and } \xi \in T_u \mathfrak{M}.$$

An elegant ‘‘Eulerian’’ approach for proving λ -convexity has been developed in [42, 9]. This approach has been implemented in [16] on the manifold of probability measures over a finite state space. A useful characterization of λ -convexity, that does not involve the metric but only the Onsager operator, has been formulated in [32]: at each $u \in \mathfrak{M}$, define the bi-linear form \mathbf{M}_u on $T_u^* \mathfrak{M}$ via

$$\mathbf{M}_u[p, p] = -p[D_u F_\Phi[\mathbf{K}_u p]] + \frac{1}{2} p[D_u \mathbf{K}[F_\Phi(u)]p].$$

Then the functional Φ is λ -convex if and only if the tensor \mathbf{M} satisfies the estimate

$$(22) \quad \mathbf{M} \geq \lambda \mathbf{K}$$

in the sense that $\mathbf{M}_u[p, p] \geq \lambda p[\mathbf{K}_u p]$ for all $u \in \mathfrak{M}$ and $p \in T_u^* \mathfrak{M}$. Here the differential D_u is to be interpreted using the linear structure of the ambient space \mathcal{X} :

$$D_u F_\Phi[\xi] = \lim_{\varepsilon \rightarrow 0} \frac{F_\Phi(u + \varepsilon \xi) - F_\Phi(u)}{\varepsilon} \in T_u \mathfrak{M},$$

and analogously for $D_u \mathbf{K}$.

Remark 1. *In a smooth Riemannian setting, λ -convexity of Φ implies λ -contractivity for its gradient flow, i.e.,*

$$d(u_t, u'_t) \leq e^{-\lambda t} d(u_0, u'_0) \quad \text{for all } t > 0 \text{ and arbitrary solutions } u, u' \text{ to (21).}$$

2.2. Estimates on the flow. In the following discussion, we will *not* require the full strength of the λ -convexity assumption $\mathbf{M} \geq \lambda \mathbf{K}$ from (22). Instead, in our calculations we will only apply (22) to the argument $D_v \Phi$. The resulting *convex decay inequality*

$$(CDI) \quad \mathbf{M}[D_v \Phi, D_v \Phi] \geq \lambda D_v \Phi[\mathbf{K}_v D_v \Phi]$$

is weaker than (22). Since

$$-\frac{d}{ds} \Phi(v_s) = D_v \Phi[\mathbf{K}_v D_v \Phi] \quad \text{and} \quad \frac{1}{2} \frac{d^2}{ds^2} \Phi(v_s) = \mathbf{M}[D_v \Phi, D_v \Phi]$$

— as will be shown in the proof below — (CDI) provides a relation between the first and the second derivative of Φ along its gradient flow. The inequality (CDI) lies at the heart of the Bakry-Émery approach to functional inequalities [2]. In a Markov chain setting, the inequality (CDI) has been studied in [5], see also [18, 19].

Our general hypothesis in the remainder of this section is the following:

Assumption 1. *The inequality (CDI) holds for some $\lambda > 0$. Moreover, Φ has a unique global minimizer $\bar{u} \in \mathfrak{M}$, and, without loss of generality, $\Phi(u) \geq \Phi(\bar{u}) = 0$ for all $u \in \mathfrak{M}$.*

Clearly, this assumption is satisfied when Φ is λ -convex, in which case \bar{u} is the only critical point of Φ .

It will be convenient to introduce the auto-dissipation $|\partial\Phi|^2 : \mathfrak{M} \rightarrow \mathbb{R}$ of Φ , which is defined by

$$(23) \quad |\partial\Phi|^2(u) = D_u\Phi[\mathbf{K}_u D_u\Phi].$$

It follows from our Assumption 1 that $|\partial\Phi|^2(\bar{u}) = 0$.

Proposition 1 (Gradient flow estimates for Φ). *Suppose that Assumption 1 holds. Then, along any solution $(v_s)_{s \geq 0}$ of the gradient flow (21), we have, for arbitrary $s \geq s' \geq 0$,*

$$(24) \quad \Phi(v_s) \leq \Phi(v_{s'})e^{-2\lambda(s-s')},$$

$$(25) \quad |\partial\Phi|^2(v_s) \leq |\partial\Phi|^2(v_{s'})e^{-2\lambda(s-s')},$$

and further, for arbitrary $s > 0$,

$$(26) \quad |\partial\Phi|^2(v_s) \leq \Phi(v_0) \frac{2\lambda}{e^{2\lambda s} - 1} \leq \Phi(v_0) s^{-1}.$$

Moreover, the following functional inequality holds for arbitrary $v \in \mathfrak{M}$:

$$(27) \quad 2\lambda\Phi(v) \leq |\partial\Phi|^2(v).$$

These results are classical. We sketch a proof here, which derives the estimates directly from the hypothesis (CDI) by elementary calculations. The idea is to use the method of iterated gradients from [2].

Proof. We start by proving (25). To this end, we estimate the decay in time of

$$(28) \quad |\partial\Phi|^2(v) = D_v\Phi[\mathbf{K}_v D_v\Phi] = \sum_{i,j} \mathbf{K}_{ij}(v) \partial_i\Phi(v) \partial_j\Phi(v),$$

which is

$$(29) \quad J(v_s) := -\frac{d}{ds} |\partial\Phi|^2(v_s).$$

From the last representation in (28), we obtain, writing $\Phi_i = \partial_i\Phi$ and $\Phi_{ij} = \partial_i\partial_j\Phi$ for brevity,

$$J(v) = -\sum_{i,j,k} \left(2\mathbf{K}_{ij}(v)\Phi_{ik}(v) + \partial_k\mathbf{K}_{ij}(v)\Phi_i(v) \right) \Phi_j(v) F_k^\Phi(v).$$

On the other hand, the definition of \mathbf{M}_v yields

$$\begin{aligned} \mathbf{M}_v[D_v\Phi, D_v\Phi] &= D_v\Phi[D_v F_\Phi[F_\Phi(v)]] + \frac{1}{2} D_v\Phi[D_v \mathbf{K}[F_\Phi(v)] D_v\Phi] \\ &= \sum_{i,j} \Phi_i(v) \partial_j F_i^\Phi(v) F_j^\Phi(v) + \frac{1}{2} \sum_{i,j,k} \Phi_i(v) \partial_k \mathbf{K}_{ij}(v) F_k^\Phi(v) \Phi_j(v) \\ &= -\sum_{i,j,k} \Phi_i(v) \left(\partial_j \mathbf{K}_{ik}(v) \Phi_k(v) + \mathbf{K}_{ik}(v) \Phi_{jk}(v) \right) F_j^\Phi(v) - \frac{1}{2} \Phi_i(v) \partial_k \mathbf{K}_{ij}(v) F_k^\Phi(v) \Phi_j(v) \\ &= -\sum_{i,j,k} \Phi_i(v) \left(\frac{1}{2} \partial_j \mathbf{K}_{ik}(v) \Phi_k(v) + \mathbf{K}_{ik}(v) \Phi_{jk}(v) \right) F_j^\Phi(v), \end{aligned}$$

where the last identity follows by relabeling the indices. We thus obtain the crucial identity

$$(30) \quad J(v) = 2\mathbf{M}_v[\mathbf{D}_v\Phi, \mathbf{D}_v\Phi].$$

Applying the assumption (CDI), we infer that

$$(31) \quad J(v) \geq 2\lambda\mathbf{D}_v\Phi[\mathbf{K}_v\mathbf{D}_v\Phi] = 2\lambda|\partial\Phi|^2(v).$$

Now apply Gronwall's lemma to the resulting inequality

$$-\frac{d}{ds}|\partial\Phi|^2(v_s) \geq 2\lambda|\partial\Phi|^2(v_s)$$

to obtain (25). Next, we verify the functional inequality (27). First, observe that

$$(32) \quad -\frac{d}{ds}\Phi(v_s) = \mathbf{D}_{v_s}\Phi[-F_\Phi(v_s)] = |\partial\Phi|^2(v_s).$$

For any fixed $s > s' > 0$, this allows to conclude that

$$\Phi(v_{s'}) - \Phi(v_s) = \int_{s'}^s |\partial\Phi|^2(v_r) dr \leq |\partial\Phi|^2(v_{s'}) \int_{s'}^s e^{-2\lambda(r-s')} dr,$$

which implies that

$$2\lambda(\Phi(v_{s'}) - \Phi(v_s)) \leq (1 - e^{-2\lambda(s-s')})|\partial\Phi|^2(v_{s'}).$$

In the limit $s \rightarrow \infty$, we have $\Phi(u_s) \rightarrow \Phi(\bar{u}) = 0$, and thus we end up with

$$2\lambda\Phi(u_{s'}) \leq |\partial\Phi|^2(u_{s'}),$$

which verifies (27). To prove (24), simply combine (32) with (27) and apply Gronwall's lemma again. Finally, the estimate (26) is a consequence of the following calculation, using that $s \mapsto |\partial\Phi|^2(v_s)$ is a monotone function thanks to (25):

$$\begin{aligned} \frac{e^{2\lambda s} - 1}{2\lambda} |\partial\Phi|^2(v_s) &= \int_0^s e^{2\lambda(s-s')} ds' |\partial\Phi|^2(v_s) \leq \int_0^s |\partial\Phi|^2(v_{s'}) ds' \\ &= - \int_0^s \frac{d}{dr} \Big|_{r=s'} \Phi(v_r) ds' = \Phi(v_0) - \Phi(v_s). \end{aligned}$$

By non-negativity of Φ , we arrive at (26). \square

2.3. Estimates on the flow of the dissipation functional. We continue to assume that (CDI) holds with some $\lambda > 0$. We also assume the normalization $\Phi(\bar{u}) = 0$, with \bar{u} being the global minimizer. Below, we study another gradient flow, namely the one generated by the dissipation $\Psi = |\partial\Phi|^2$,

$$(33) \quad \dot{u} = F_\Psi(u) = -\mathbf{K}_u\mathbf{D}_u\Psi.$$

In general, no information is available on the convexity of the flow induced by F_Ψ . Still, the following analogue of Proposition 1 holds, thanks to the intimate relation of Ψ to the λ -convex functional Φ .

Proposition 2 (Gradient flow estimates for Ψ). *Suppose that Assumption 1 holds. Then, along any solution $(u_t)_{t>0}$ of the auxiliary gradient flow (33), we have, for arbitrary $t \geq t' \geq 0$,*

$$(34) \quad \Phi(u_t) \leq \Phi(u_{t'})e^{-(2\lambda)^2(t-t')},$$

$$(35) \quad |\partial\Phi|^2(u_t) \leq |\partial\Phi|^2(u_{t'})e^{-(2\lambda)^2(t-t')},$$

and further, for arbitrary $t > 0$,

$$(36) \quad |\partial\Phi|^2(u_t) \leq \Phi(u_0) \frac{2\lambda}{e^{(2\lambda)^2 t} - 1} \leq \Phi(u_0) (2\lambda t)^{-1}.$$

This result has been proven in the setting of metric spaces in [35, Section 3]. As for Proposition 1, we sketch a proof here which only uses the inequality (CDI) and some elementary calculations.

Proof. We start by estimating the decay of $\Phi(u_t)$ in time, i.e.,

$$J(u_t) := -\frac{d}{dt}\Phi(u_t).$$

Observe that, thanks to the symmetry of the Onsager operator,

$$J(u) = D_u\Phi[-F_\Psi(u)] = D_u\Phi[\mathbf{K}_u D_u\Psi] = D_u\Psi[\mathbf{K}_u D_u\Phi] = D_u\Psi[-F_\Phi(u)],$$

thus the J defined above coincides with the J defined in (29). From the inequality (31) in combination with the inequality (27), it follows that

$$(37) \quad J(u) \geq 2\lambda\Psi(u) \geq (2\lambda)^2\Phi(u).$$

Another application of Gronwall's lemma yields (34). In preparation for the proof of (35), observe that the Cauchy-Schwarz inequality for the scalar product $\langle \cdot, \cdot \rangle_u$ translates into the following inequality for the Onsager operator \mathbf{K}_u :

$$(p[\mathbf{K}_u q])^2 \leq p[\mathbf{K}_u p] q[\mathbf{K}_u q] \quad \text{for all } p, q \in T_u^*\mathfrak{M}.$$

In combination with the estimate (31), we obtain

$$(2\lambda)^2 (D_u\Phi[\mathbf{K}_u D_u\Phi])^2 \leq J(u)^2 = (D_u\Psi[\mathbf{K}_u D_u\Phi])^2 \leq D_u\Psi[\mathbf{K}_u D_u\Psi] D_u\Phi[\mathbf{K}_u D_u\Phi].$$

Division by $D_u\Phi[\mathbf{K}_u D_u\Phi]$ leads to

$$(38) \quad I(u) := D_u\Psi[\mathbf{K}_u D_u\Psi] \geq (2\lambda)^2 D_u\Phi[\mathbf{K}_u D_u\Phi] = (2\lambda)^2 \Psi(u).$$

Since $I(u_t) = -\frac{d}{dt}\Psi(u_t)$, we obtain (35) by yet another application of Gronwall's lemma. For the proof of (36), we use the inequality (35) and the first inequality from (37). We thus obtain

$$\begin{aligned} \frac{e^{(2\lambda)^2 t} - 1}{(2\lambda)^2} |\partial\Phi|^2(u_t) &= \int_0^t e^{(2\lambda)^2(t-t')} |\partial\Phi|^2(u_t) dt' \leq \int_0^t |\partial\Phi|^2(u_{t'}) dt' \\ &\leq \frac{1}{2\lambda} \int_0^t J(u_{t'}) dt' = -\frac{1}{2\lambda} \int_0^t \frac{d}{dr} \Big|_{r=t'} \Phi(u_r) dt' = \frac{\Phi(u_0) - \Phi(u_t)}{2\lambda}, \end{aligned}$$

from which the first inequality in (36) follows since Φ is non-negative. The second inequality is elementary. \square

2.4. Application: asymptotics for the Fokker-Planck and QDD equation. To conclude our short review on gradient flows with (CDI), we show how the the estimates (9)–(12) on the long-time asymptotics for solutions to (2) and (1), respectively, can be obtained from Propositions 1 and 2 above, at least formally. For the *rigorous* derivation of the stated long-time asymptotics by variational methods, we refer the reader to [1] and to [35].

We consider the set $\mathcal{P}_+(\Omega)$ of strictly positive probability densities $u : \Omega \rightarrow \mathbb{R}_+$, endowed with the L^2 -Wasserstein metric, as Riemannian manifold \mathfrak{M} . Tangent and cotangent vectors at $u \in \mathfrak{M}$ are identified with functions $\xi, p \in L^2(\Omega)$ of vanishing mean, their pairing being given by

$$p[\xi] = \int_{\Omega} p(x)\xi(x) dx.$$

The definition of the scalar product on the tangent spaces is intricate (it requires the solution of an auxiliary elliptic problem), but the associated Onsager operator \mathbf{K}_u has an explicit form:

$$(39) \quad \mathbf{K}_u p = -\nabla \cdot (u \nabla p).$$

In this framework, the Fokker-Planck equation (2) can be written as the gradient flow of the entropy \mathcal{H} from (6):

$$(40) \quad \partial_s v_s = \Delta_\pi v_s \quad \text{with} \quad \Delta_\pi v = -\mathbf{K}_v D_v \mathcal{H} = \nabla \cdot (v \nabla \log(v/\pi)) = \Delta v + \nabla \cdot (v \nabla V).$$

This representation has been the starting point for the existence proof in the celebrated work [26].

Next, by the results of McCann [37], the λ -convexity of the potential V implies λ -convexity of this gradient flow; see also [9] for an alternative proof of this fact using the formalism developed above. Proposition 1 immediately yields the convergence properties stated in (9) as well as the regularization estimate (12).

We proceed to analyze (1). To begin with, let us rewrite — by integration by parts — the Fisher information \mathcal{I} from (6) with the help of Δ_π introduced in (40):

$$(41) \quad \mathcal{I}(w) = - \int_{\Omega} \log(w/\pi) \nabla \cdot (w \nabla (w/\pi)) dx = - \int_{\Omega} \log(w/\pi) \Delta_\pi w dx.$$

From this representation, it is immediate to deduce the relation (8) between entropy and Fisher information, i.e., that

$$|\partial \mathcal{H}|^2(w) = -D_w \mathcal{H}[\Delta_\pi w] = \mathcal{I}(w).$$

Next, we use (41) to compute the first variation of \mathcal{I} :

$$(42) \quad D_u \mathcal{I}[\xi] = - \int_{\Omega} [(\xi/u) \Delta_\pi u + \log(u/\pi) \Delta_\pi \xi] dx = - \int_{\Omega} \left[\frac{\Delta_\pi u}{u} + \Delta_\pi^* \log(u/\pi) \right] \xi dx,$$

where Δ_π^* is the $L^2(dx)$ -adjoint of Δ_π , that is

$$\Delta_\pi^* \log(u/\pi) = \Delta_\pi^* \log u + \Delta_\pi^* V = \Delta \log u - \frac{\nabla V \cdot \nabla u}{u} + \Delta V - |\nabla V|^2.$$

We thus obtain

$$D_u \mathcal{I}[\xi] = - \int_{\Omega} \left[\frac{\Delta u}{u} + \Delta \log u + 2\Delta V - |\nabla V|^2 \right] \xi dx.$$

From this and the relation (4) between V and W , it is obvious that (1) can be written as the gradient flow of \mathcal{I} :

$$\partial_t u_t = F_{\mathcal{I}}(u_t) \quad \text{with} \quad F_{\mathcal{I}}(u) = -\mathbf{K}_u D_u \mathcal{I} = -\nabla \cdot \left(u \nabla \left[\frac{\Delta u}{u} + \Delta \log u - W \right] \right).$$

Remark 2. For later reference, we point out that in view of (42), the equation (1) can be equivalently written in the form

$$(43) \quad \partial_t u = \mathbf{K}_u \left(\frac{\Delta_\pi u}{u} + \Delta_\pi^* \log(u/\pi) \right).$$

This is the representation which naturally appears after discretization, see (77) below.

In combination, this means that Proposition 2 applies to solutions u_t of (1). The respective estimates (34) and (35) turn into (11), and (36) becomes (12).

3. DISCRETIZATION OF THE FOKKER-PLANCK EQUATION

3.1. Finite volume discretization. For given $N \in \mathbb{N}$, define the length parameter $h := 1/N$, and introduce the d -dimensional cubic lattice of side length N ,

$$J^h := \{1, \dots, N\}^d \subset \mathbb{Z}^d.$$

Multi-indices in J^h are denoted by \mathbf{i} and \mathbf{j} , and we write $\mathbf{i} \leftrightarrow \mathbf{j}$ if \mathbf{i} and \mathbf{j} are neighbors, i.e., $|\mathbf{i} - \mathbf{j}| = 1$. Intuitively, each $\mathbf{j} \in J^h$ labels a subcube

$$\omega_{\mathbf{j}} := (h(j_1 - 1), h j_1) \times \dots \times (h(j_d - 1), h j_d) \subset \Omega$$

of side length h in Ω ; for later reference, we introduce for each $\mathbf{j} \in J^h$ the center $x_{\mathbf{j}}^h$ of the \mathbf{j} th cube,

$$x_{\mathbf{j}}^h := (h(j_1 - 1/2), \dots, h(j_d - 1/2)).$$

Further, each vector $U \in \mathbb{R}^{J^h}$ is associated to a function $u^h \in L^\infty(\Omega)$ that is piecewise constant on each $\omega_{\mathbf{j}}$:

$$u^h(x) = U_{\mathbf{j}} \quad \text{for all } x \in \omega_{\mathbf{j}}.$$

In this spirit, we refer to

$$\mathcal{P}_+(J^h) := \left\{ U \in \mathbb{R}_+^{J^h}; h^d \sum_{\mathbf{j} \in J^h} U_{\mathbf{j}} = 1 \right\}$$

as the space of positive probability densities on J^h ; indeed, for each $U \in \mathcal{P}_+(J^h)$,

$$\int_{\Omega} u^h(x) dx = h^d \sum_{\mathbf{j} \in J^h} U_{\mathbf{j}} = 1.$$

Both vectors $\Xi \in \mathbb{T}_U \mathcal{P}_+(J^h)$ and cotangent vectors $P \in \mathbb{T}_U^* \mathcal{P}_+(J^h)$ are identified with elements in \mathbb{R}^{J^h} of vanishing mean,

$$h^d \sum_{\mathbf{i}} P_{\mathbf{i}} = 0, \quad h^d \sum_{\mathbf{j}} \Xi_{\mathbf{j}} = 0,$$

and their pairing is given by

$$P[\Xi] = h^d \sum_{\mathbf{j} \in J^h} P_{\mathbf{j}} \Xi_{\mathbf{j}}.$$

Next, we introduce a discrete approximation $\Pi^h \in \mathcal{P}_+(J^h)$ of the steady state π from (3). First, define vectors $V^{[1],h}, \dots, V^{[d],h} \in \mathbb{R}^N$ by

$$(44) \quad V_j^{[k],h} = \frac{1}{h} \int_{h(j-1)}^{hj} V^{[k]}(r) dr,$$

and accordingly $\Pi^{[1],h}, \dots, \Pi^{[d],h} \in \mathcal{P}_+(\{1, \dots, N\})$ by

$$(45) \quad \Pi_j^{[k],h} = \frac{1}{Z^{[k],h}} \exp(-V_j^{[k],h}),$$

with the appropriate choice of the normalization constant $Z^{[k],h} > 0$. Now, $\Pi^h \in \mathcal{P}_+(J^h)$ itself is defined such that it inherits the product structure (16):

$$(46) \quad \Pi_{\mathbf{j}}^h = \Pi_{j_1}^{[1],h} \dots \Pi_{j_d}^{[d],h}.$$

Since V is smooth, the respective piecewise constant densities π^h converge to π uniformly on Ω as $h \downarrow 0$.

Lemma 1. *The piecewise constant representation $\pi^h \in \mathcal{P}_+(\Omega)$ with respective values Π_j^h on the cubes ω_j is the unique minimizer of \mathcal{H} on the subspace of piecewise constant densities in $\mathcal{P}_+(\Omega)$. Moreover,*

$$(47) \quad \gamma^h := \mathcal{H}(\pi^h) = \log \frac{Z^{[1]} \dots Z^{[d]}}{Z^{[1],h} \dots Z^{[d],h}}.$$

This lemma justifies the definition of the discretized entropy \mathcal{H}^h in (14).

Remark 3. *It is easily seen that $Z^{[k],h} \searrow Z^{[k]}$ for each $k = 1, \dots, d$ as $h \searrow 0$. Hence $\gamma^h \searrow 0$.*

Proof. If u^h is piecewise constant on the boxes ω_j with respective values U_j , then

$$\begin{aligned} \mathcal{H}(u^h) &= \int_{\Omega} u^h \log(u^h/\pi) \, dx = \sum_{j \in J^h} \int_{\omega_j} U_j (\log U_j - \log \pi) \, dx \\ &= h^d \sum_{j \in J^h} U_j \left(\log U_j + \log Z + \frac{1}{h^d} \int_{\omega_j} V(x) \, dx \right) \\ &= h^d \sum_{j \in J^h} U_j \log(U_j/\Pi_j^h) + \log \frac{Z^{[1]} \dots Z^{[d]}}{Z^{[1],h} \dots Z^{[d],h}}. \end{aligned}$$

For the last line, we have used the property (15) of V , which yields that

$$\frac{1}{h^d} \int_{\omega_j} V(x) \, dx = \frac{1}{h^d} \int_{\omega_j} (V^{[1]}(x_1) + \dots + V^{[d]}(x_d)) \, dx = V_{j_1}^{[1],h} + \dots + V_{j_d}^{[d],h},$$

the property $Z = Z^{[1]} \dots Z^{[d]}$, and the definition of Π^h in (45)&(46) above. Since both $U^h, \Pi^h \in \mathcal{P}_+(J^h)$, we may further write

$$\mathcal{H}(u^h) = h^d \sum_{j \in J^h} \Pi_j^h \left(1 + (U_j/\Pi_j^h) [\log(U_j/\Pi_j^h) - 1] \right) + \log \frac{Z^{[1]} \dots Z^{[d]}}{Z^{[1],h} \dots Z^{[d],h}}.$$

Using that $r \mapsto r(\log r - 1) + 1$ is strictly convex with minimum zero attained at $r = 1$, Jensen's inequality implies that

$$\mathcal{H}(u^h) \geq \log \frac{Z^{[1]} \dots Z^{[d]}}{Z^{[1],h} \dots Z^{[d],h}},$$

with equality if and only if $u^h = \pi^h$. □

3.2. Discretized Fokker-Planck equation. We implicitly introduce a metric on $\mathcal{P}_+(J^h)$ by means of the Onsager operator $\mathbf{K}_U^h : \mathbb{T}_U^* \mathcal{P}_+(J^h) \rightarrow \mathbb{T}_U \mathcal{P}_+(J^h)$ with

$$(48) \quad P[\mathbf{K}_U^h Q] = h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \Lambda_{\mathbf{ij}}(U) \left(\frac{P_{\mathbf{i}} - P_{\mathbf{j}}}{h} \right) \left(\frac{Q_{\mathbf{i}} - Q_{\mathbf{j}}}{h} \right),$$

for all $P, Q \in \mathbb{T}_U^* \mathcal{P}_+(J^h)$, where the sum runs over all pairs of neighboring indices $\mathbf{i} \leftrightarrow \mathbf{j}$, i.e., over all edges of unit length in J^h , and $\Lambda_{\mathbf{ij}}(U)$ is an abbreviation of

$$\Lambda_{\mathbf{ij}}(U) = \Lambda \left(\frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h}, \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \right),$$

with the logarithmic mean $\Lambda : \mathbb{R}_+ \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$, given by

$$(49) \quad \Lambda(a, b) = \int_0^1 a^{1-x} b^x dx = \begin{cases} \frac{a-b}{\log a - \log b} & \text{if } a \neq b, \\ a & \text{if } a = b. \end{cases}$$

It has been shown in [33] that \mathbf{K}^h induces a distance on $\mathcal{P}_+(J^h)$, which extends to the closure $\mathcal{P}(J^h)$ of merely non-negative probability densities. The resulting metric space is geodesic and complete.

Remark 4. *The definition (48) of the discrete Onsager operator \mathbf{K}^h above is consistent with that of the Onsager operator \mathbf{K} for the L^2 -Wasserstein metric on $\mathcal{P}_+(\Omega)$ from (39) in the following sense. Let a smooth density $u \in \mathcal{P}_+(\Omega)$ with positive lower bound and two smooth functions $p, q : \Omega \rightarrow \mathbb{R}$ be given. For $h \downarrow 0$, let $U^h \in \mathcal{P}_+(J^h)$ and $P^h, Q^h \in \mathbb{T}_{U^h}^* \mathcal{P}_+(J^h)$ be approximations of u and p, q with the asymptotics*

$$(50) \quad U_j^h = u(x_j^h) + \mathcal{O}(h), \quad P_j^h = p(x_j^h) + \mathcal{O}(h^2), \quad Q_j^h = q(x_j^h) + \mathcal{O}(h^2).$$

Then:

$$(51) \quad P^h [\mathbf{K}_{U^h}^h Q^h] = p[\mathbf{K}_u q] + \mathcal{O}(h).$$

This follows by a straightforward asymptotic expansion in h which is detailed in Appendix 5.1.

As announced in (14), the entropy functional \mathcal{H}^h on $\mathcal{P}_+(J^h)$ is defined by restriction of the original entropy \mathcal{H} ,

$$\mathcal{H}^h(U) = h^d \sum_{j \in J_N} U_j \log(U_j / \Pi_j^h) - \gamma^h = \mathcal{H}(u^h) - \gamma^h,$$

where γ^h defined in (47) is such that the convex functional $\mathcal{H}^h(U)$ is non-negative for all $U \in \mathcal{P}_+(J^h)$, and vanishes precisely for $U = \Pi^h$ given in (46). Accordingly, introduce the discretization of the Fokker-Planck operator Δ_π on $\mathcal{P}_+(J^h)$ by

$$(52) \quad \mathbb{M}^h U := -\mathbf{K}_U^h D_U \mathcal{H}^h.$$

The representation as a linear operator is justified by the following.

Lemma 2. *The discrete Fokker-Planck operator \mathbb{M}^h is linear on the simplex $\mathcal{P}_+(J^h)$:*

$$(\mathbb{M}^h U)_i = \sum_{j \in J^h} \mathbb{M}_{ij}^h U_j \quad \text{for all } U \in \mathcal{P}_+(J^h),$$

with the matrix elements of $\mathbb{M}^h \in \mathbb{R}^{J^h \times J^h}$ being given by

$$(53) \quad \mathbb{M}_{ij}^h = \begin{cases} h^{-2} \sqrt{\Pi_i^h / \Pi_j^h} & \text{if } i \leftrightarrow j, \\ -h^{-2} \sum_{i': i' \leftrightarrow j} \sqrt{\Pi_{i'}^h / \Pi_j^h} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, the adjoint operator $\mathcal{L}^h = (\mathbb{M}^h)^*$ given by $(\mathcal{L}^h \psi)_i = \sum_{j \in J^h} \mathbb{M}_{ji}^h \psi_j$ is the generator of an irreducible and reversible Markov chain on J^h with invariant distribution Π^h .

Proof. First observe that, at each $U \in \mathcal{P}_+(J^h)$,

$$D_U \mathcal{H}^h[\Xi] = h^d \sum_{j \in J^h} (1 + \log(U_j / \Pi_j^h)) \Xi_j \quad \text{for all } \Xi \in \mathbb{T}_U \mathcal{P}_+(J^h).$$

Thus, by definition of Λ , we have for each $P \in \mathbb{T}_U^* \mathcal{P}_+(J^h)$:

$$\begin{aligned} P[-\mathbf{K}_U^h D_U \mathcal{H}^h] &= -h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \Lambda_{\mathbf{ij}}(U) \left(\frac{\log(U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h) - \log(U_{\mathbf{j}}/\Pi_{\mathbf{j}}^h)}{h} \right) \left(\frac{P_{\mathbf{i}} - P_{\mathbf{j}}}{h} \right) \\ &= -h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left(\frac{U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h - U_{\mathbf{j}}/\Pi_{\mathbf{j}}^h}{h} \right) \left(\frac{P_{\mathbf{i}} - P_{\mathbf{j}}}{h} \right) \\ &= h^d \sum_{\mathbf{i}} P_{\mathbf{i}} \left[\sum_{\mathbf{j}: \mathbf{i} \leftrightarrow \mathbf{j}} h^{-2} \left(\sqrt{\frac{\Pi_{\mathbf{i}}^h}{\Pi_{\mathbf{j}}^h}} U_{\mathbf{j}} - \sqrt{\frac{\Pi_{\mathbf{j}}^h}{\Pi_{\mathbf{i}}^h}} U_{\mathbf{i}} \right) \right]. \end{aligned}$$

This shows the linearity of the operator in (52), and yields the representation (53).

\mathbb{M}^h being the adjoint generator of a Markov chain means that all of its off-diagonal entries are non-negative, and that the column sums vanish. Both properties are immediately verified by inspection of (53). Irreducibility means that for any two indices $\mathbf{i}_*, \mathbf{i}^*$, one finds a chain $(\mathbf{i}_m)_{m=0, \dots, M}$ of indices $\mathbf{i}_m \in J^h$ with $\mathbf{i}_0 = \mathbf{i}_*$ and $\mathbf{i}_M = \mathbf{i}^*$ such that $\mathbb{M}_{\mathbf{i}_m \mathbf{i}_{m-1}}^h > 0$ for all $m = 1, \dots, M$. Since $\mathbb{M}_{\mathbf{ij}}^h > 0$ whenever $\mathbf{i} \leftrightarrow \mathbf{j}$, one may take for $(\mathbf{i}_m)_{m=0, \dots, M}$ any chain with $\mathbf{i}_{m-1} \leftrightarrow \mathbf{i}_m$ connecting \mathbf{i}_* with \mathbf{i}^* . For reversibility, we need to verify the detailed balance condition

$$(54) \quad \mathbb{M}_{\mathbf{ij}}^h \Pi_{\mathbf{j}}^h = \mathbb{M}_{\mathbf{ji}}^h \Pi_{\mathbf{i}}^h \quad \text{for all } \mathbf{i}, \mathbf{j} \in J^h.$$

This again is an immediate consequence of the representation (53). Note that (54) together with the Markov property implies $\mathbb{M}^h \Pi^h = \Pi^h$, i.e., Π^h is indeed an (in fact: the unique) invariant distribution. \square

Remark 5. *Also here, one can verify consistency between discrete and continuous objects: let $f: \Omega \rightarrow \mathbb{R}$ be a smooth function, and define the discrete approximation F^h by $F_{\mathbf{j}}^h = f(x_{\mathbf{j}}^h)$. Then*

$$(55) \quad (\mathbb{M}^h F^h)_{\mathbf{i}} = \Delta_{\pi} f(x_{\mathbf{i}}^h) + \mathcal{O}(h^2).$$

The respective calculation is given in Appendix 5.2. Also, a rigorous convergence analysis for the flow induced by \mathbb{M}^h to the Fokker-Planck flow has been carried out in [14], see also [22].

3.3. λ -contractivity of the Fokker-Planck flow. The goal of this section is to prove:

Proposition 3. *The flow $F_{\mathcal{H}^h}$ from (52) satisfies (CDI) with*

$$(56) \quad \lambda^h = \frac{2}{h^2} \left(1 - \exp \left(-\frac{h^2}{2} \lambda \right) \right) = \lambda + \mathcal{O}(h^2).$$

This result appears to be novel for dimensions $d > 1$, but its proof is obtained by combination of two results from the literature. The key observation is that, in view of the factorization property (46), the space $\mathcal{P}_+(J^h)$ carries a natural tensorial structure that is compatible with the evolution (52) of the spatially discrete Fokker-Planck equation. More precisely:

Lemma 3. *For each pair of indices $\mathbf{i}, \mathbf{j} \in J^h$,*

$$(57) \quad \mathbb{M}_{\mathbf{ij}}^h = \mathbb{M}_{i_1 j_1}^{[1],h} \delta_{i_2 j_2} \delta_{i_3 j_3} \cdots \delta_{i_d j_d} + \delta_{i_1 j_1} \mathbb{M}_{i_2 j_2}^{[2],h} \delta_{i_3 j_3} \cdots \delta_{i_d j_d} + \cdots + \delta_{i_1 j_1} \delta_{i_2 j_2} \delta_{i_3 j_3} \cdots \mathbb{M}_{i_d j_d}^{[d],h}.$$

Proof. Eventually, we will apply [5, Theorem 3.1], which deals precisely with matrices $\mathbb{M}^{h,[k]}$ and operators $\mathbf{K}^{[k],h}$ of the forms (58) and (62), respectively. But first, we establish the following auxiliary estimate

$$(64) \quad \sqrt{\Pi_{i+1}^{[k],h} \Pi_{i-1}^{[k],h}} \leq \left(1 - \frac{h^2}{2} \lambda^h\right) \Pi_i^{[k],h}.$$

Indeed, by λ -convexity of $V^{[k]}$, we have that

$$\frac{1}{2}(V^{[k]}(x_k + h) + V^{[k]}(x_k - h)) \geq V^{[k]}(x_k) + \frac{\lambda}{2} h^2.$$

Integration of this inequality from $x_k = (i-1)h$ to $x_k = ih$ yields

$$\frac{1}{2}(V_{i+1}^{[k],h} + V_{i-1}^{[k],h}) \geq V_i^{[k],h} + \frac{\lambda}{2} h^2,$$

which further implies that

$$\sqrt{\exp(-V_{i+1}^{[k],h}) \exp(-V_{i-1}^{[k],h})} \leq \exp\left(-\frac{h^2}{2} \lambda\right) \exp(-V_i^{[k],h}).$$

Recalling the definition (45) of $\Pi^{[k],h}$, and the definition (56) of λ^h , the estimate (64) follows.

An immediate consequence of (64) is the validity of the monotonicity hypotheses

$$(65) \quad \alpha_j^{[k]} \leq \alpha_{j-1}^{[k]} \quad \text{and} \quad \beta_j^{[k]} \leq \beta_{j+1}^{[k]}.$$

Therefore, [5, Theorem 3.1] is applicable. It provides the (CDI) for $\mathbb{M}^{[k],h}$ with respect to the Onsager operator $\mathbf{K}^{[k],h}$, for each

$$(66) \quad \lambda \leq \lambda^* := \min_{i=2, \dots, N-1} (\alpha_i^{[k]} - \alpha_{i+1}^{[k]}) + (\beta_i^{[k]} - \beta_{i-1}^{[k]}).$$

Now, from the definitions (59) and (60) of $\alpha^{[k]}$ and $\beta^{[k]}$, it follows via (64) that

$$(67) \quad \begin{aligned} \alpha_i^{[k],h} - \alpha_{i+1}^{[k],h} &= h^{-2} \sqrt{\Pi_{i+1}^{[k],h} / \Pi_i^{[k],h}} - h^{-2} \sqrt{\Pi_{i+2}^{[k],h} / \Pi_{i+1}^{[k],h}} \geq h^{-2} \left(\frac{h^2}{2} \lambda^h\right) \sqrt{\Pi_{i+1}^{[k],h} / \Pi_i^{[k],h}} \\ \beta_i^{[k],h} - \beta_{i-1}^{[k],h} &= h^{-2} \sqrt{\Pi_i^{[k],h} / \Pi_{i+1}^{[k],h}} - h^{-2} \sqrt{\Pi_{i-1}^{[k],h} / \Pi_i^{[k],h}} \geq h^{-2} \left(\frac{h^2}{2} \lambda^h\right) \sqrt{\Pi_i^{[k],h} / \Pi_{i+1}^{[k],h}}, \end{aligned}$$

which implies that

$$(68) \quad \lambda^* \geq \lambda^h \min_{i=2, \dots, N-1} \cosh(V_i^{[k],h} - V_{i+1}^{[k],h}) \geq \lambda^h,$$

as desired. \square

Proposition 3 follows immediately by combining Lemma 4 with the tensorisation result from Theorem 2 below.

Remark 6. *In the setting of Proposition 4, it is possible to prove the stronger property of λ -convexity (with a slightly worse constant) with a minor modification of the proof. Instead of using [5, Theorem 3.1] to obtain the inequality (66) as above, one could apply Mielke's criterion from [40, Theorem 5.1] to obtain $\tilde{\lambda}^h$ -convexity, with*

$$(69) \quad \tilde{\lambda}^h := 2 \min_{i=2, \dots, N-1} \sqrt{(\alpha_i^{[k]} - \alpha_{i+1}^{[k]})(\beta_i^{[k]} - \beta_{i-1}^{[k]})}.$$

Note that the arithmetic mean in (66) is replaced by a geometric mean in (69). It is easily seen that $\lambda^h > \tilde{\lambda}^h$, but that the difference $\tilde{\lambda}^h - \lambda^h = O(h^2)$ becomes negligible in the discrete-to-continuous limit $h \downarrow 0$. In view of the tensorisation result from (70) the result remains valid in any dimension with the same constant.

3.4. Tensorisation of convex entropy decay for Markov chains. In this section, we sketch the proof for stability of the inequality (CDI) under tensorization. This result is independent of the discretization and might be of interest on its own right.

We need to fix some notations. First, we recall an alternative representation of a continuous time Markov chain on a finite set \mathcal{I} : the generator $\mathcal{L} : L^\infty(\mathcal{I}) \rightarrow L^\infty(\mathcal{I})$ can be written as

$$(\mathcal{L}\psi)_i = \sum_{\delta \in \mathcal{G}} c_{i,\delta} (\psi_{\delta(i)} - \psi_i) .$$

Here, \mathcal{G} is a set of maps from \mathcal{I} to \mathcal{I} representing the possible jumps, and $c_{i,\delta} \geq 0$ denotes the jump rate from i to $\delta(i)$. For brevity, we shall write $\nabla_\delta \psi_i := \psi_{\delta(i)} - \psi_i$.

Throughout this section we assume that the following reversibility conditions are satisfied:

- for every $\delta \in \mathcal{G}$ there exists a unique $\delta^{-1} \in \mathcal{G}$ satisfying $\delta^{-1}(\delta(i)) = i$ for all i with $c_{i,\delta} > 0$;
- there exists a probability measure $\pi = (\pi_i)_{i \in \mathcal{I}}$ on \mathcal{I} such that $\pi_i > 0$ for all i , and

$$\sum_{i \in \mathcal{I}, \delta \in \mathcal{G}} F(i, \delta) c_{i,\delta} \pi_i = \sum_{i \in \mathcal{I}, \delta \in \mathcal{G}} F(\delta(i), \delta^{-1}) c_{i,\delta} \pi_i$$

for all $F : \mathcal{I} \times \mathcal{G} \rightarrow \mathbb{R}$.

The relative entropy functional $\mathcal{H}_\pi : \mathcal{P}_+(\mathcal{I}) \rightarrow \mathbb{R}$ is given by

$$\mathcal{H}_\pi(u) = \sum_{i \in \mathcal{I}} u_i \log(u_i / \pi_i) .$$

In accordance with the situation described before, we introduce an Onsager operator \mathbf{K} such that $\mathcal{L}^*(u) = \mathbf{K}_u \mathbf{D}_u \mathcal{H}_\pi(u)$:

$$p[\mathbf{K}_u q] = \frac{1}{2} \sum_{i \in \mathcal{I}, \delta \in \mathcal{G}} \pi_i c_{i,\delta} \Lambda_{i,\delta}(u) \nabla_\delta p_i \nabla_\delta q_i , \quad \text{where } \Lambda_{i,\delta}(u) = \Lambda\left(\frac{u_i}{\pi_i}, \frac{u_{\delta(i)}}{\pi_{\delta(i)}}\right).$$

For later reference, let us calculate the Hessian \mathbf{M} : it follows from the definition that

$$\begin{aligned} \mathbf{M}_u[p, p] &= -\frac{1}{2} \sum_{i \in \mathcal{I}} \sum_{\delta, \eta \in \mathcal{G}} \pi_i c_{i,\delta} \Lambda_{i,\delta}(u) \nabla_\delta p_i \left[c_{\delta(i),\eta} \nabla_\eta p_{\delta(i)} - c_{i,\eta} \nabla_\eta p_i \right] \\ &\quad + \frac{1}{4} \sum_{i \in \mathcal{I}} \sum_{\delta, \eta \in \mathcal{G}} \pi_i c_{i,\delta} (\nabla_\delta p_i)^2 \left[c_{i,\eta} \Lambda_{i,\delta}^1(u) \nabla_\eta (u/\pi)_i + c_{\delta(i),\eta} \Lambda_{i,\delta}^2(u) \nabla_\eta (u/\pi)_{\delta(i)} \right] \\ &=: \sum_{i \in \mathcal{I}} \sum_{\delta, \eta \in \mathcal{G}} F_{u,p}(i, \delta, \eta), \end{aligned}$$

where $\Lambda_{i,\delta}^m(u) = \partial_m \Lambda\left(\frac{u_i}{\pi_i}, \frac{u_{\delta(i)}}{\pi_{\delta(i)}}\right)$ for $m = 1, 2$. See [16, 40] for details.

Now consider a collection of Markov chains $(\mathcal{I}^k, \mathcal{L}^k, \pi^k)$ for $k = 1, \dots, N$. The corresponding product chain $(\mathcal{I}^\otimes, \mathcal{L}^\otimes, \pi^\otimes)$ is defined by

$$\mathcal{I}^\otimes = \mathcal{I}^1 \times \dots \times \mathcal{I}^N, \quad (\mathcal{L}^\otimes \psi)_\mathbf{i} = \sum_k \mathcal{L}^k \psi_{\mathbf{i}} \quad \text{for } \psi \in L^\infty(\mathcal{I}), \quad \pi^\otimes = \pi^1 \otimes \dots \otimes \pi^N.$$

Here it is understood that \mathcal{L}^k acts on the k -th coordinate of ψ . The corresponding Onsager operator and Hessian will be denoted by \mathbf{K}^\otimes and by \mathbf{M}^\otimes , respectively. To simplify notations we shall write $\mathcal{H}^k := \mathcal{H}_{\pi^k}$.

It has been shown in [16] that geodesic λ -convexity is preserved under tensorisation:

$$(70) \quad \left[\mathbf{M}^k \geq \lambda_k \mathbf{K}^k \quad \text{for all } k = 1, \dots, N \right] \implies \left[\mathbf{M}^\otimes \geq \left(\min_k \lambda_k \right) \mathbf{K}^\otimes \right].$$

This result is dimension independent, i.e., the bound does not depend on N . The goal of this section is to verify that the corresponding tensorisation property also holds for the convex entropy decay inequality (CDI).

Theorem 2 (Tensorisation of convex entropy decay). *Suppose that the inequality (CDI) holds for each $k = 1, \dots, N$:*

$$\mathbf{M}_u^k(D_u \mathcal{H}^k, D_u \mathcal{H}^k) \geq \lambda_k D_u \mathcal{H}^k [\mathbf{K}_u^k D_u \mathcal{H}^k]$$

for all $u \in \mathcal{P}_+(\mathcal{I}^k)$ and some $\lambda_k \in \mathbb{R}$. Then (CDI) also holds for the product chain:

$$\mathbf{M}_u^\otimes(D_u \mathcal{H}_\pi, D_u \mathcal{H}_\pi) \geq \left(\min_k \lambda_k \right) D_u \mathcal{H}_\pi [\mathbf{K}_u^\otimes D_u \mathcal{H}_\pi]$$

for all $u \in \mathcal{P}_+(\mathcal{I}^\otimes)$.

The proof follows along the lines of the proof of (70) in [16]. For the convenience of the reader we provide some details.

Proof. For each $k = 1, \dots, K$, set $\mathcal{I}_{\bar{k}} = \prod_{\ell \neq k} \mathcal{I}_\ell$, and for $\mathbf{i} \in \mathcal{I}^\otimes$, let $i_{\bar{k}} \in \mathcal{I}_{\bar{k}}$ be the multi-index with the k th entry i_k omitted. For a function $p : \mathcal{I}^\otimes \rightarrow \mathbb{R}$, define its reduction $p^{i_{\bar{k}}} : \mathcal{I}_{\bar{k}} \rightarrow \mathbb{R}$ where all indices except i_k are fixed to $i_{\bar{k}}$, i.e., $p^{i_{\bar{k}}} = p_{\mathbf{i}}$. Likewise, introduce $u^{i_{\bar{k}}} : \mathcal{I}_{\bar{k}} \rightarrow \mathbb{R}_+$. Finally, set $\pi^{\bar{k}} = \bigotimes_{\ell \neq k} \pi^\ell$.

It follows from the definitions that the Onsager matrix for the product system admits a decomposition of the form

$$(71) \quad p[\mathbf{K}_u^\otimes p] = \sum_{k=1}^N \sum_{i_{\bar{k}} \in \mathcal{I}_{\bar{k}}} \pi_{i_{\bar{k}}}^{\bar{k}} p^{i_{\bar{k}}} [\mathbf{K}_u^{i_{\bar{k}}} p^{i_{\bar{k}}}] .$$

Also \mathbf{M}_u^\otimes can be split into terms corresponding to the different components:

$$\mathbf{M}_u^\otimes[p, p] = \sum_{k, \ell=1}^N \mathbf{M}_u^{k, \ell}[p, p], \quad \text{where} \quad \mathbf{M}_u^{k, \ell}[p, p] = \sum_{\mathbf{i} \in \mathcal{I}^\otimes} \sum_{\delta \in \mathcal{G}^k, \eta \in \mathcal{G}^\ell} F_{u, p}(\mathbf{i}, \delta, \eta),$$

and \mathcal{G}^k denotes the set of maps associated with the operator \mathcal{L}^k . It has been shown in [16] that the off-diagonal terms are non-negative (regardless of the λ -convexity properties of the components):

$$(72) \quad \mathbf{M}_u^{k, \ell}[p, p] \geq 0 \quad \text{if } k \neq \ell$$

for all $u \in \mathcal{P}_+(\mathcal{I}^\otimes)$ and $p : \mathcal{I}^\otimes \rightarrow \mathbb{R}$. The on-diagonal terms satisfy

$$(73) \quad \mathbf{M}_u^{k, k}[p, p] = \sum_{i_{\bar{k}} \in \mathcal{I}_{\bar{k}}} \pi_{i_{\bar{k}}}^{\bar{k}} \mathbf{M}_u^{i_{\bar{k}}} [p^{i_{\bar{k}}}, p^{i_{\bar{k}}}] .$$

Now fix $u \in \mathcal{P}_+(\mathcal{I}^\otimes)$. Then

$$(D_u \mathcal{H}_\pi)_{i_{\bar{k}}}^{i_{\bar{k}}} = \log(u_{\mathbf{i}} / \pi_{\mathbf{i}}) = \log(u_{i_{\bar{k}}}^{i_{\bar{k}}} / \pi_{i_{\bar{k}}}^k) - \log(\pi_{i_{\bar{k}}}^{\bar{k}}), \quad (D_{u^{i_{\bar{k}}}} \mathcal{H}^k)_{i_{\bar{k}}} = \log(u_{i_{\bar{k}}}^{i_{\bar{k}}} / \pi_{i_{\bar{k}}}^k),$$

which allows to conclude that

$$(D_u \mathcal{H}_\pi)^{i_{\bar{k}}} = D_{u^{i_{\bar{k}}}} \mathcal{H}^k - \log(\pi_{i_{\bar{k}}}^{\bar{k}}).$$

Hence both derivatives coincide up to a constant, whose value is irrelevant, since $\mathbf{M}_u[p, p]$ depends on p only through the values of its discrete derivatives $\nabla_\delta p$. Putting everything together, one obtains

$$\begin{aligned} \mathbf{M}_u^\otimes [D_u \mathcal{H}_\pi, D_u \mathcal{H}_\pi] &\stackrel{(72)-(73)}{\geq} \sum_{k=1}^N \sum_{i_{\bar{k}} \in \mathcal{I}_{\bar{k}}} \pi_{i_{\bar{k}}}^{\bar{k}} \mathbf{M}_{u^{i_{\bar{k}}}}^k [D_{u^{i_{\bar{k}}}} \mathcal{H}^k, D_{u^{i_{\bar{k}}}} \mathcal{H}^k] \\ &\stackrel{(CDI)}{\geq} \text{for } k \sum_{k=1}^N \lambda_k \sum_{i_{\bar{k}} \in \mathcal{I}_{\bar{k}}} \pi_{i_{\bar{k}}}^{\bar{k}} D_{u^{i_{\bar{k}}}} \mathcal{H}^k [\mathbf{K}_{u^{i_{\bar{k}}}}^k D_{u^{i_{\bar{k}}}} \mathcal{H}^k] \stackrel{(71)}{\geq} \left(\min_k \lambda_k \right) D_u \mathcal{H}_\pi [\mathbf{K}_u^\otimes D_u \mathcal{H}_\pi], \end{aligned}$$

which is the desired result. \square

4. DISCRETIZATION OF THE QDD EQUATION

In this section, we study the gradient flow of the discretized Fisher information \mathcal{I}^h , which is defined by

$$(74) \quad \mathcal{I}^h(U) = D_U \mathcal{H}^h[-\mathbb{M}^h U] = h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \Lambda_{\mathbf{ij}}(U) \left(\frac{\log(U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h) - \log(U_{\mathbf{j}}/\Pi_{\mathbf{j}}^h)}{h} \right)^2.$$

Notice that this definition is in accordance with (14).

4.1. Existence of the gradient flow.

Lemma 5. \mathcal{I}^h is well-defined and non-negative on $\mathcal{P}_+(J^h)$, with $\mathcal{I}^h(U) = 0$ if and only if $U = \Pi^h$. Moreover, \mathcal{I}^h has the alternative representation

$$(75) \quad \mathcal{I}^h(U) = h^{d-2} \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left(\frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} - \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \right) \left(\log \frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} - \log \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \right)$$

for each $U \in \mathcal{P}_+(J^h)$. Finally, all sublevel sets of \mathcal{I}^h are relatively compact in $\mathcal{P}_+(J^h)$.

Remark 7. Since the closure of $\mathcal{P}_+(J^h)$ in \mathbb{R}^{J^h} is the compact simplex

$$\mathcal{P}(J^h) = \left\{ U \in \mathbb{R}_+^{J^h} ; \sum_{\mathbf{j} \in J^h} U_{\mathbf{j}} = 1 \right\},$$

a subset of $\mathcal{P}_+(J^h)$ is relatively compact in $\mathcal{P}_+(J^h)$ if and only if it is a closed subset of $\mathcal{P}(J^h)$. A consequence is that if $A \subset \mathcal{P}_+(J^h)$ is relatively compact, then it has a positive distance $\delta > 0$ to the boundary of $\mathcal{P}(J^h)$, i.e.,

$$(76) \quad \inf_{U \in A} \min_{\mathbf{j} \in J^h} U_{\mathbf{j}} > 0.$$

Proof. Well-definedness and non-negativity are obvious from (74). Since $\Lambda_{\mathbf{ij}}(U) > 0$ for each $U \in \mathcal{P}_+(J^h)$, and since any two indices $\mathbf{i}, \mathbf{j} \in J^h$ can be connected by a sequence of neighbors, $\mathcal{I}^h(U) = 0$ holds if and only if $\log(U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h)$ is a constant independent of \mathbf{i} . That is, $U = \alpha \Pi^h$ for a global constant $\alpha > 0$. Now $U \in \mathcal{P}_+(J^h)$ implies $\alpha = 1$, i.e., $U = \Pi^h$.

The representation (75) follows immediately from the definition (49) of the logarithmic mean, since $a - b = \Lambda(a, b)(\log a - \log b)$.

It remains to prove the compactness of sublevel sets. By continuity of \mathcal{I}^h , any sublevel set $A := (\mathcal{I}^h)^{-1}([0, a])$ is relatively closed in $\mathcal{P}_+(J^h)$. In view of Remark 7 above, it remains to be verified that A is also closed in $\mathcal{P}(J^h)$, i.e., that the closure of A in \mathbb{R}^{J^h} does not intersect $\mathcal{P}(J^h) \setminus \mathcal{P}_+(J^h)$. Towards a contradiction, assume that a sequence $(U^n)_{n \in \mathbb{N}}$ in A is such that $U^n \rightarrow U^* \notin \mathcal{P}_+(J^h)$; we are going to show that $\mathcal{I}^h(U^n) \rightarrow \infty$. By compactness of $\mathcal{P}(J^h)$ in \mathbb{R}^{J^h} , the limit U^* lies in the boundary $\mathcal{P}(J^h) \setminus \mathcal{P}_+(J^h)$. Thus, there is some $\mathbf{i}_* \in J^h$ with $U_{\mathbf{i}_*}^* = 0$. On the other hand, $U^* \in \mathcal{P}(J^h)$ implies that there is some $\mathbf{i}^* \in J^h$ with $U_{\mathbf{i}^*}^* > 0$. Since \mathbf{i}_* and \mathbf{i}^* can be connected by a sequence of neighbors, there must exist $\mathbf{j}_*, \mathbf{j}^* \in J^h$ with $\mathbf{j}_* \leftrightarrow \mathbf{j}^*$ and $U_{\mathbf{j}_*}^* = 0$, $a := U_{\mathbf{j}^*}^* > 0$. Since all the terms in the summation in (75) are non-negative,

$$\begin{aligned} \mathcal{I}^h(U^n) &\geq h^{d-2} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left(\frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h} - \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right) \left(\log \frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h} - \log \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right) \\ &= h^{d-2} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left[\underbrace{\left(\frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h} - \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right) \log \frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h}}_{=(I)} + \underbrace{\left(\frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h} - \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right) \left(-\log \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right)}_{=(II)} \right]. \end{aligned}$$

By the choices made above,

$$(I) \rightarrow \frac{U_{\mathbf{j}^*}^*}{\Pi_{\mathbf{j}^*}^h} \log \frac{U_{\mathbf{j}^*}^*}{\Pi_{\mathbf{j}^*}^h} \geq -e^{-1},$$

while for all sufficiently large n ,

$$(II) \geq \frac{1}{2} \frac{U_{\mathbf{j}^*}^n}{\Pi_{\mathbf{j}^*}^h} \left(-\log \frac{U_{\mathbf{j}_*}^n}{\Pi_{\mathbf{j}_*}^h} \right),$$

which obviously diverges to $+\infty$ as $n \rightarrow \infty$. \square

We calculated the derivative of \mathcal{I}^h :

$$D_U \mathcal{I}^h[\Xi] = D_U^2 \mathcal{H}^h[-\mathbb{M}^h U, \Xi] + D_U \mathcal{H}^h[-\mathbb{M}^h \Xi] = -h^2 \sum_{\mathbf{i}, \mathbf{j}} \Xi_{\mathbf{i}} \left(\frac{1}{U_{\mathbf{i}}} \mathbb{M}_{\mathbf{ij}}^h U_{\mathbf{j}} + \log(U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h) \mathbb{M}_{\mathbf{ij}}^h \right).$$

In other words, with a certain abuse of notation, the gradient flow of \mathcal{I}^h is given by

$$(77) \quad \dot{U} = -\mathbf{K}_U^h D_U \mathcal{I}^h = \mathbf{K}_U^h \left(\frac{\mathbb{M}^h U}{U} + (\mathbb{M}^h)^T \log(U/\Pi^h) \right).$$

The initial value problem for this gradient flow is well-posed.

Lemma 6. *For every initial condition $U_0 \in \mathcal{P}_+(J^h)$, there is a unique differentiable curve $U : \mathbb{R}_{\geq 0} \rightarrow \mathcal{P}_+(J^h)$ satisfying (77) with $U(0) = U_0$.*

Proof. The right-hand side of (77) is obviously smooth in U . By the standard theory of ordinary differential equations, there exists a *maximal local solution* $U : [0, T) \rightarrow \mathcal{P}_+(J^h)$. Here ‘‘maximal’’ means that either $T = \infty$, i.e., the local solution is global, or that there is no limit point in $\mathcal{P}_+(J^h)$ of $U(t)$ for $t \uparrow T$. We are now going to prove that the second alternative is impossible.

Indeed, since $U_0 \in \mathcal{P}_+(J^h)$, we have $a := \mathcal{I}^h(U(0)) < \infty$ by Lemma 5. Now U being a gradient flow implies that $\mathcal{I}^h(U(t)) \leq a$ for all $t \in [0, T]$. That is, the curve U lies in the sublevel set $A := (\mathcal{I}^h)^{-1}([0, a])$, which is compact by Lemma 5. The smooth vector field $U \mapsto -\mathbf{K}_U^h D_U \mathcal{I}^h$ is bounded on A , and consequently, U is uniformly Lipschitz continuous on $[0, T]$. Therefore, $U(t)$ has a limit in A for $t \uparrow T$. \square

Remark 8. *The discrete equation (77) is consistent with the QDD equation (1) in the following sense: let u_t be a smooth and strictly positive solution to (1), and let $p : \Omega \rightarrow \mathbb{R}$ be smooth as well. Define U_t^h and P^h by restriction of u_t and p , respectively, to the grid, i.e., $(U_t^h)_j = u_t(x_j^h)$ and $P_j^h = p(x_j^h)$. Further, define Q_t^h by*

$$Q_t^h = \frac{\mathbb{M}^h U_t^h}{U_t^h} + (\mathbb{M}^h)^T \log(U_t^h / \Pi^h).$$

Thanks to consistency (55) of the operator \mathbb{M}^h , one easily concludes that

$$(Q_t^h)_i = \frac{\Delta_\pi u_t(x_i^h)}{u_t(x_i^h)} + \Delta_\pi^* \log\left(\frac{u_t}{\pi}\right)(x_i^h) + \mathcal{O}(h^2).$$

Combining this with the consistency (51) of the Onsager operator \mathbf{K}^h — note that the hypotheses (50) are satisfied — yields

$$P^h[\mathbf{K}_{U_t^h} Q_t^h] = p \left[\mathbf{K}_{u_t} \left(\frac{\Delta_\pi u_t}{u_t} + \Delta_\pi^* \log(u_t / \pi) \right) \right].$$

The right-hand side is a weak formulation — with test function p — of the nonlinearity on the right-hand side of (43), which is just an equivalent representation of (1).

4.2. Proof of the main theorem. We are finally in the position to prove Theorem 1, i.e., we derive the estimates (18), (19) and (20) for the gradient flow (77) of the discrete Fisher information functional \mathcal{I}^h .

The estimates (18) and (19) follow easily by means of Proposition 2. Indeed, in order to verify that Proposition 2 applies in our situation, it suffices to observe that the discrete entropy functional \mathcal{H}^h satisfies (CDI), which is a consequence of Proposition 3 above, and of the fact that $\mathcal{I}^h = |\partial \mathcal{H}^h|^2$ by definition in (74). For the proof of (20), we combine the first estimate in (11) with the Csiszar-Kullback inequality, see e.g. [43], which specializes in the case at hand to

$$\|u^h - \pi^h\|_{L^1(\Omega)}^2 \leq 2\mathcal{H}^h(u^h).$$

4.3. Discretization in time. We shall now use our spatial discretization as basis for the implementation of a numerical scheme for approximate solution of (1). More precisely, we apply a discretization in time to the ordinary differential equations (77),

$$\frac{d}{dt} U = F^h(U) = K_U^h S_U, \quad \text{with} \quad (S_U)_i = \frac{(\mathbb{M}^h U)_i}{U_i} + \left[(\mathbb{M}^h)^T \log \left(\frac{U_j}{\Pi_j^h} \right)_{j \in J^h} \right]_i.$$

For discretization in time, an implicit Euler scheme is employed: we replace the function $U^h : [0, T] \rightarrow \mathbb{R}_+^{J^h}$ by a sequence $(U_m^{h,\tau})_m$ with the interpretation that $U_m^{h,\tau}$ approximates $U^h(m\tau)$, and solve

$$(78) \quad \frac{U_m^{h,\tau} - U_{m-1}^{h,\tau}}{\tau} = F^h(U_m^{h,\tau})$$

inductively for $m = 1, 2, \dots$

Remark 9. *Since our spatial discretization relies on the gradient flow structure of the equation, it would be very natural to exploit this structure also for the temporal discretization, by implementing De Giorgi's minimizing movement scheme. This scheme automatically inherits (suitable discrete versions of) the essential a priori estimates and thus guarantees the correct asymptotic behavior of the time-discrete solutions, see [1]. However, a numerical implementation of this scheme would require us to solve a minimization problem involving the associated Riemannian distance in each time-step of the scheme. As this distance is only implicitly defined via solution of an another, auxiliary minimization problem, the numerical computation of each time step would be extremely expensive. The implicit Euler scheme is much faster and much easier to implement — and it still preserves the crucial decay estimates (18), as is shown next.*

Proposition 4. *Assume that a sequence $(U_m^{h,\tau})_{m \geq 0}$ satisfies (78). Then the following time-discrete variants of (18) hold:*

(79)

$$\mathcal{H}^h(U_m^{h,\tau}) \leq \mathcal{H}^h(U_{m'}^{h,\tau}) (1 + (2\lambda^h)^2 \tau)^{-(m-m')} \quad \text{and} \quad \mathcal{I}^h(U_m^{h,\tau}) \leq \mathcal{I}^h(U_{m'}^{h,\tau}) (1 + (2\lambda^h)^2 \tau)^{-(m-m')},$$

for all integers $m \geq m' \geq 0$.

The proof of Proposition 4 is a consequence of the following convexity property.

Lemma 7. *Both \mathcal{H}^h and \mathcal{I}^h are convex on $\mathcal{P}_+(J^h)$ in the sense of linear interpolation.*

Remark 10. *We emphasize that convexity with respect to linear interpolation and geodesic convexity with respect to the Onsager operator \mathbf{K}^h are (almost) unrelated notions.*

Proof of Lemma 7. First, recall that $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ with $\phi(s) = s \log s$ has derivatives

$$\phi'(s) = 1 + \log s, \quad \phi''(s) = 1/s.$$

Given $U \in \mathcal{P}_+(J^h)$ and $\Xi \in \mathbb{T}_U \mathcal{P}_+(J^h)$, we have on the one hand that

$$D_U^2 \mathcal{H}^h[\Xi]^2 = h^d \sum_{\mathbf{j}} \partial_{U_{\mathbf{j}}}^2 (U_{\mathbf{j}} \log U_{\mathbf{j}} - U_{\mathbf{j}} \log \Pi_{\mathbf{j}}^h) \Xi_{\mathbf{j}}^2 = h^d \sum_{\mathbf{j}} U_{\mathbf{j}}^{-1} \Xi_{\mathbf{j}}^2 \geq 0,$$

and on the other hand that

$$\begin{aligned} D_U^2 \mathcal{I}^h[\Xi]^2 &= h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left\{ \partial_{U_{\mathbf{i}}}^2 \left(\left(\frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} - \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \right) \log \frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} \right) \Xi_{\mathbf{i}}^2 + \partial_{U_{\mathbf{j}}}^2 \left(\left(\frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} - \frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} \right) \log \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \right) \Xi_{\mathbf{j}}^2 \right. \\ &\quad \left. - 2 \partial_{U_{\mathbf{i}}} \partial_{U_{\mathbf{j}}} \left(\frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} \log \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} + \frac{U_{\mathbf{j}}}{\Pi_{\mathbf{j}}^h} \log \frac{U_{\mathbf{i}}}{\Pi_{\mathbf{i}}^h} \right) \Xi_{\mathbf{i}} \Xi_{\mathbf{j}} \right\} \\ &= h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \left(\frac{\Pi_{\mathbf{i}}^h}{U_{\mathbf{i}}} + \frac{\Pi_{\mathbf{j}}^h}{U_{\mathbf{j}}} \right) \left(\sqrt{\frac{\Pi_{\mathbf{i}}^h U_{\mathbf{j}}}{U_{\mathbf{i}} \Pi_{\mathbf{j}}^h \Pi_{\mathbf{i}}^h}} \Xi_{\mathbf{i}} - \sqrt{\frac{\Pi_{\mathbf{j}}^h U_{\mathbf{i}}}{U_{\mathbf{j}} \Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h}} \Xi_{\mathbf{j}} \right)^2 \geq 0. \end{aligned}$$

Non-negativity of the second derivatives implies convexity. □

Proof of Proposition 4. Apply the derivative of \mathcal{H}^h at $U_m^{h,\tau}$ to (78) to obtain

$$-D_{U_m^{h,\tau}} \mathcal{H}^h \left[\frac{U_m^{h,\tau} - U_{m-1}^{h,\tau}}{\tau} \right] = D_{U_m^{h,\tau}} \mathcal{H}^h \left[\mathbf{K}_{U_m^{h,\tau}} D_{U_m^{h,\tau}} \mathcal{I}^h \right] \geq (2\lambda^h)^2 \mathcal{H}^h(U_m^{h,\tau}),$$

where we have used the estimate (37) and (CDI) with constant λ^h to obtain the inequality. Furthermore, since \mathcal{H}^h is convex by Lemma 7 above,

$$\mathcal{H}^h(U_{m-1}^{h,\tau}) \geq \mathcal{H}^h(U_m^{h,\tau}) - D_{U^{h,\tau}} \mathcal{H}^h \left[U_m^{h,\tau} - U_{m-1}^{h,\tau} \right] \geq (1 + (2\lambda^h)^2 \tau) \mathcal{H}^h(U_m^{h,\tau}).$$

An iteration of this estimate yields the first inequality in (79). The proof of the second inequality is obtained in an analogous way, now applying the derivative of \mathcal{I}^h in place of \mathcal{H}^h to (78), using the estimate (38), and the convexity of \mathcal{I}^h with respect to linear interpolation. \square

4.4. Numerical experiments. In our experiments, we restrict attention to the two-dimensional situation $d = 2$, i.e., $\Omega = [0, 1]^2$ is the unit square. For the potential, we have used $V(x) = \lambda/2|x - \bar{x}|^2$, with $\bar{x} = (1/2, 1/2)$ the center of $[0, 1]^2$, corresponding to $W(x) = \lambda^2|x - \bar{x}|^2 - 4\lambda$. Different choices for the convexity parameter λ are used in the simulations. In each experiment, a spatial resolution of $N = 30$ grid points in each direction has been used. The time step $\tau > 0$ is chosen in dependence of λ ; since we solve the implicit Euler scheme (78) by an undamped Newton iteration in each time step, a sufficiently small τ is necessary for numerical well-posedness of the scheme.

4.4.1. Illustration of qualitative behavior. For illustration of the complex qualitative behavior of solutions u to (1), we report results for a numerical experiment in the unconfined case $W \equiv 0$, i.e., $\lambda = 0$, for the initial datum

$$(80) \quad u_0(x) = \frac{1}{Z} (\cos^{16} \pi x_1 + \cos^{16} \pi x_2) + 10^{-4},$$

where $Z = 0.392\dots$ is such that u_0 integrates to one on $[0, 1]^2$. This is a straight-forward generalization of the one-dimensional example from [3, Figure 1] to two space dimensions. Plots of the discrete initial condition u_0^h and of the evolved discrete densities u_t^h are given in Figure 1 for a resolution of $N = 30$ cells in both coordinate directions. Notice that u_0 has a large plateau where its values are very small (order 10^{-4}) in comparison to the average value (order one). The sharp flanks at the edge of the plateau drive the dynamics and lead to a rather complicated spatial-temporal behavior of the solution. The top right plot inside Figure 1 shows a one-dimensional cross-section of the solution; qualitatively, the behavior is in perfect agreement with the one-dimensional simulations from [3]. A time step $\tau = 10^{-7}$ has been used for the numerical solution in order to resolve the process of creation and destruction of local minima inside the plateau region, which happens on a time scale of 10^{-6} ; figures produced with a finer time resolution $\tau = 10^{-8}$ look identical.

4.4.2. Rates of convergence to equilibrium. The goal of the following series of experiments is the numerical verification of the analytically estimated rates of convergence to equilibrium. We vary the convexity parameter λ and apply the numerical scheme to the very regular initial condition

$$(81) \quad u_0(x) = \frac{4}{3} \sin^2(3\pi x_1) \sin^2(2\pi x_2) + \frac{1}{3}(1 + x_1 + x_2).$$

The behavior of entropy and Fisher information are monitored for about one thousand time iterations. The qualitative change in density is shown in Figure 3. The corresponding results for entropy and Fisher information are collected in Figure 2.

In agreement with the analytical estimates in (79), both quantities decay with a rate of at least $(2\lambda^h)^2$. In fact, in each experiment we measure a minimal decay rate $(2\lambda_*^h)^2$ that is strictly larger than the analytically predicted rate. Generally, the difference $\lambda_*^h - \lambda^h$ is the larger the smaller $\lambda > 0$ is, and becomes negligible for large values $\lambda \gg 10$.

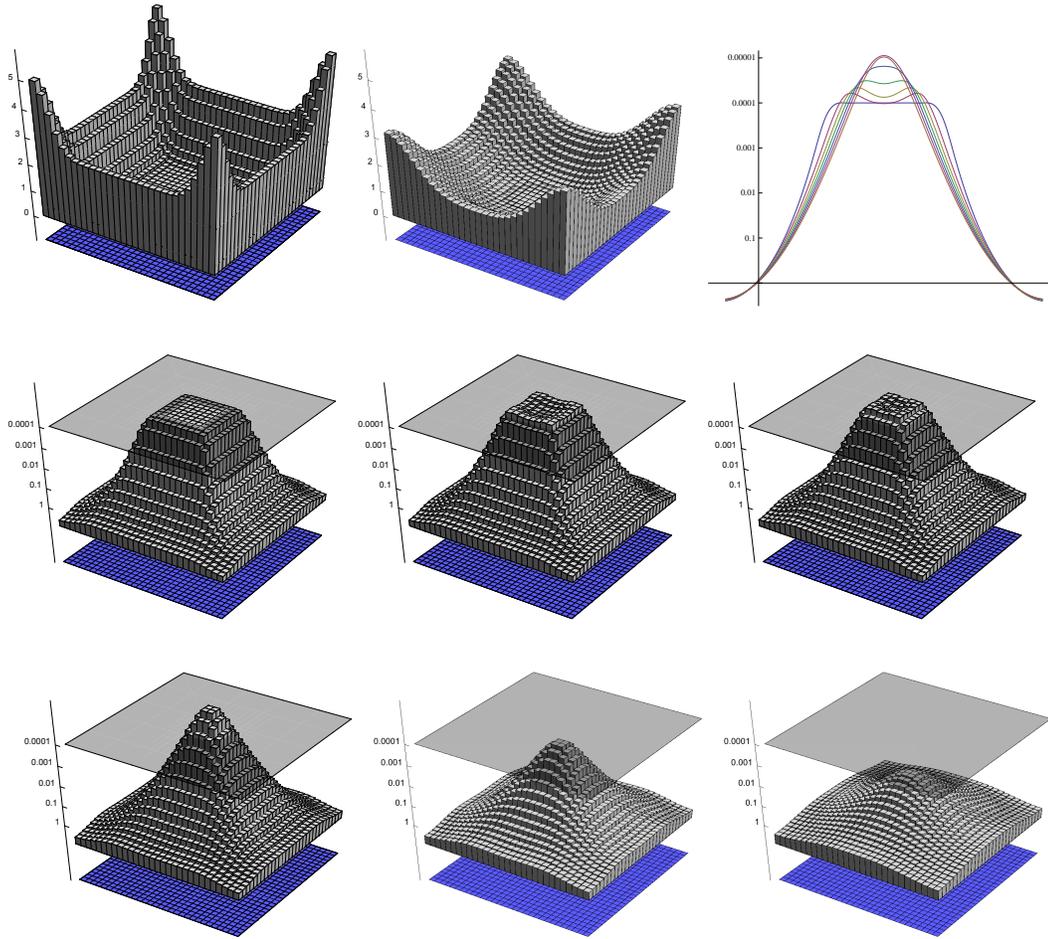


FIGURE 1. Results of an experiment with $\lambda = 0$ and the initial condition from (80), using discretization parameters $N = 30$ and $\tau = 10^{-7}$.

Top row, from left to right: discretized initial condition u_0^h ; corresponding discrete solution u_t^h at $t = 3.0 \times 10^{-5}$; negative logarithmic values on the one-dimensional cross-section $x_2 = 1/2$ at times $t = k \times 10^{-6}$ for $k = 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$.

Middle and bottom row: snapshots of the solution in negative logarithmic plot at times $t = k \times 10^{-6}$ for $k = 0, 0.3, 1.0, 3.0, 10.0, 30.0$. The transparent plane at $u \equiv 10^{-4}$ has been introduced to visualize that the solution does not obey a maximum principle.

This phenomenon is apparently independent of the spatial resolution $h > 0$. Our conjecture is the following. For solutions to the Fokker-Planck equation (2), the estimates (9) are not sharp: the lower bound on the rate of convergence to equilibrium is given by $2\lambda_*$ with some $\lambda_* > \lambda$. This improvement is due to boundary effects: it is negligible if the steady state π is very concentrated

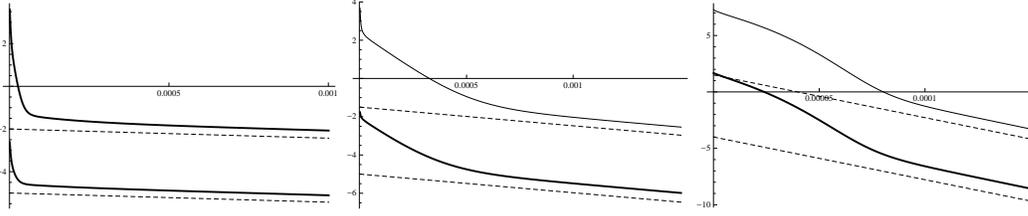


FIGURE 2. Logarithmic plot of entropy (lower bold curve) and Fisher information (upper bold curve) along discrete solutions for the initial condition (81), using $\lambda = 1$, $\lambda = 10$ and $\lambda = 100$ (left to right). The dotted lines correspond to multiples of $\exp(-(2\lambda_*)^2 t)$.

inside Ω (as is the case for $\lambda \gg 10$), but is significant for more equally distributed stationary densities π (for $\lambda < 10$ or less). Thanks to its intimate relation to the Fokker-Planck equation (2), the fourth order equation (1) apparently inherits these improved rates, i.e., one can replace $(2\lambda)^2$ by $(2\lambda_*)^2$ in the estimates (11). These improved estimates pass on to the estimates (18) and (79) on the discretization.

Our conjecture is strongly supported by the outcome of the experiments. In Figure (2), the decay rates of entropy and Fisher information are compared to $(2\lambda_*^h)^2$, where $\lambda_*^h > \lambda^h$ is the smallest non-zero eigenvalue of the associated Markov generator \mathbb{M}^h on J^h . In all of the experiments that have been performed, the numerically measured rate of decay of entropy and Fisher information,

$$\frac{1}{\tau} \left(\log \mathcal{H}^h(U_m^{h,\tau}) - \log \mathcal{H}^h(U_{m+1}^{h,\tau}) \right) \quad \text{and} \quad \frac{1}{\tau} \left(\log \mathcal{I}^h(U_m^{h,\tau}) - \log \mathcal{I}^h(U_{m+1}^{h,\tau}) \right), \quad \text{respectively,}$$

never fall below the value $(2\lambda_*^h)^2$. In fact, the numerically measured rates have always been larger but appear to tend towards $(2\lambda_*^h)^2$ as the system approaches equilibrium. This is in accordance with the observation from [35] that the rates of convergence to equilibrium are minimized in the linearized regime around the steady state.

Remark 11. *It is tempting to turn the above conjecture into a proof, simply using the spectral gap λ_* instead of λ as a lower bound on the modulus of geodesic convexity of \mathcal{H} and performing all the estimates accordingly. However, to our knowledge, there is no result available which allows to estimate the modulus of geodesic convexity of a Markov chain — or the corresponding constant λ in the inequality (CDI) — by its spectral gap from below.*

5. APPENDIX

Here we provide two formal calculations for consistency.

5.1. Consistency for the Onsager operator. We verify (51). By hypothesis (50), the values $U_{\mathbf{i}}^h$ and $U_{\mathbf{j}}^h$ at neighboring sites $\mathbf{i} \leftrightarrow \mathbf{j}$ are $\mathcal{O}(h)$ -close to each other. Hence, the logarithmic and geometric mean of $U_{\mathbf{i}}/\Pi_{\mathbf{i}}^h$ and $U_{\mathbf{j}}/\Pi_{\mathbf{j}}^h$ are also $\mathcal{O}(h)$ -close — in fact, even $\mathcal{O}(h^2)$ -close — to each other. Hence, we have

$$\sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \Lambda_{\mathbf{ij}}(U^h) = \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \sqrt{\frac{U_{\mathbf{i}}^h}{\Pi_{\mathbf{i}}^h} \frac{U_{\mathbf{j}}^h}{\Pi_{\mathbf{j}}^h}} + \mathcal{O}(h) = \sqrt{U_{\mathbf{i}}^h U_{\mathbf{j}}^h} + \mathcal{O}(h) = u(x_{\mathbf{j}}^h) + \mathcal{O}(h)$$

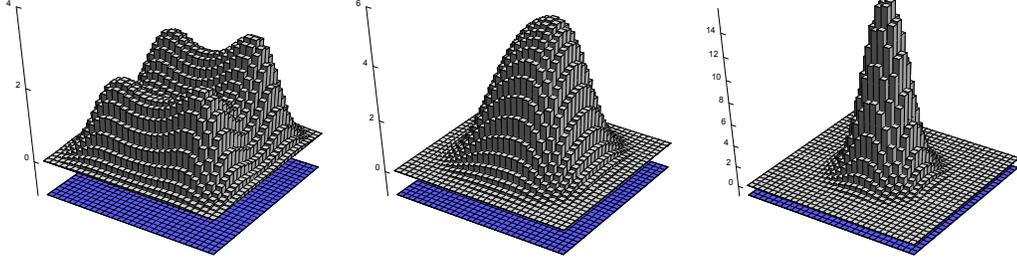


FIGURE 3. Snapshots of the discrete solution for the initial condition (81) using $\lambda = 100$ at, respectively, $t = 10^{-5}$, $t = 3 \cdot 10^{-5}$ and $t = 10^{-4}$ (left to right). Note the changing scale.

inside the definition (48). Further, thanks to the square grid combinatorics of J^h , and to the hypotheses (50),

$$\frac{P_{\mathbf{i}}^h - P_{\mathbf{j}}^h}{h} = (\mathbf{i} - \mathbf{j}) \cdot \nabla p(x_{\mathbf{j}}^h) + \mathcal{O}(h),$$

and similarly for the difference quotients of Q^h . Writing $\mathbf{i} - \mathbf{j} = \pm h \mathbf{e}_k$ and working out the combinatorics, one obtains from the definition of the discretize Onsager operator in (48) the following integral approximation:

$$\begin{aligned} P^h[\mathbf{K}_{U^h}^h Q^h] &= h^d \sum_{\mathbf{i} \leftrightarrow \mathbf{j}} \sqrt{\Pi_{\mathbf{i}}^h \Pi_{\mathbf{j}}^h} \Lambda_{\mathbf{ij}}(U^h) \left(\frac{P_{\mathbf{i}}^h - P_{\mathbf{j}}^h}{h} \right) \left(\frac{Q_{\mathbf{i}}^h - Q_{\mathbf{j}}^h}{h} \right) \\ &= h^d \sum_{\mathbf{j} \in J^h} \left[(u(x_{\mathbf{j}}^h) + \mathcal{O}(h)) \sum_{k=1}^d (\mathbf{e}_k \cdot \nabla p(x_{\mathbf{j}}^h) + \mathcal{O}(h)) (\mathbf{e}_k \cdot \nabla q(x_{\mathbf{j}}^h) + \mathcal{O}(h)) \right] \\ &= \int_{\Omega} u(x) \nabla p(x) \cdot \nabla q(x) dx + \mathcal{O}(h) = p[\mathbf{K}_u q] + \mathcal{O}(h), \end{aligned}$$

by definition of the original Onsager operator from (39).

5.2. Consistency for the Markov operator. We verify (55). From the potential W , define the averaged quantity

$$\mathfrak{W}(x) = \frac{h^{-d}}{2} \int_{[-\frac{h}{2}, +\frac{h}{2}]^d} W(x+a) da,$$

so that, for $\mathbf{i} \leftrightarrow \mathbf{j}$:

$$\sqrt{\frac{\Pi_{\mathbf{i}}^h}{\Pi_{\mathbf{j}}^h}} = \exp(\mathfrak{W}(x_{\mathbf{j}}^h) - \mathfrak{W}(x_{\mathbf{i}}^h)).$$

For a smooth function $f : \Omega \rightarrow \mathbb{R}$, let F^h be its restriction to grid points, $F_j^h = f(x_j^h)$. Then

$$\begin{aligned} (\mathbb{M}^h F^h)_i &= h^{-2} \sum_{j:i \leftrightarrow j} \left(\sqrt{\frac{\Pi_i^h}{\Pi_j^h}} F_j^h - \sqrt{\frac{\Pi_j^h}{\Pi_i^h}} F_i^h \right) \\ &= h^{-2} \sum_{j:i \leftrightarrow j} \left(e^{\mathfrak{W}(x_j^h) - \mathfrak{W}(x_i^h)} f(x_j^h) - e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_j^h)} f(x_i^h) \right). \end{aligned}$$

Assuming that \mathbf{i} is not a point on the boundary, we can use the square-grid combinatorics to rewrite this sum as follows:

$$\begin{aligned} (\mathbb{M}^h F^h)_i &= h^{-2} \sum_{k=1}^d \left(e^{\mathfrak{W}(x_i^h + h\mathbf{e}_k) - \mathfrak{W}(x_i^h)} f(x_i^h + h\mathbf{e}_k) + e^{\mathfrak{W}(x_i^h - h\mathbf{e}_k) - \mathfrak{W}(x_i^h)} f(x_i^h - h\mathbf{e}_k) \right. \\ &\quad \left. - \left[e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h + h\mathbf{e}_k)} + e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h - h\mathbf{e}_k)} \right] f(x_i^h) \right) \end{aligned} \quad (82)$$

$$\begin{aligned} &= \frac{1}{2} \sum_{k=1}^d \left[e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h + h\mathbf{e}_k)} + e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h - h\mathbf{e}_k)} \right] \frac{f(x_i^h + h\mathbf{e}_k) + f(x_i^h - h\mathbf{e}_k) - 2f(x_i^h)}{h^2} \\ &\quad + \frac{h^{-2}}{2} \sum_{k=1}^d \left(\left[e^{\mathfrak{W}(x_i^h + h\mathbf{e}_k) - \mathfrak{W}(x_i^h)} - e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h - h\mathbf{e}_k)} \right] f(x_i^h + h\mathbf{e}_k) \right. \end{aligned} \quad (83)$$

$$\begin{aligned} &\quad \left. + \left[e^{\mathfrak{W}(x_i^h - h\mathbf{e}_k) - \mathfrak{W}(x_i^h)} - e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h + h\mathbf{e}_k)} \right] f(x_i^h - h\mathbf{e}_k) \right) \\ &\quad + h^{-2} \sum_{k=1}^d \left[\sinh(\mathfrak{W}(x_i^h + h\mathbf{e}_k) - \mathfrak{W}(x_i^h)) f(x_i^h + h\mathbf{e}_k) \right. \\ &\quad \left. + \sinh(\mathfrak{W}(x_i^h - h\mathbf{e}_k) - \mathfrak{W}(x_i^h)) f(x_i^h - h\mathbf{e}_k) \right]. \end{aligned} \quad (84)$$

We shall now analyze the Taylor expansions (around x_i^h , in terms of h) of the expressions in (82), (83) and (84). Since W is smooth, so is \mathfrak{W} , and it is easily seen that

$$e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h + h\mathbf{e}_k)} + e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h - h\mathbf{e}_k)} = 2 + \mathcal{O}(h^2).$$

Further, it is one of the properties of the symmetric finite difference discretization of the Laplacian that

$$\frac{f(x_i^h + h\mathbf{e}_k) + f(x_i^h - h\mathbf{e}_k) - 2f(x_i^h)}{h^2} = \partial_{x_k}^2 f(x_i^h) + \mathcal{O}(h^2).$$

So we obtain that

$$(82) = \Delta f(x_i^h) + \mathcal{O}(h^2).$$

For the other two sums (83) and (84), we need an expansion of \mathfrak{W} :

$$\begin{aligned} \mathfrak{W}(x_i^h \pm h\mathbf{e}_k) - \mathfrak{W}(x_i^h) &= \frac{h^{-d}}{2} \int_{[-\frac{h}{2}, +\frac{h}{2}]^d} [W(x_i^h \pm h\mathbf{e}_k + a) - W(x_i^h + a)] da \\ &= \frac{h^{-d}}{2} \int_{[-\frac{h}{2}, +\frac{h}{2}]^d} \left[\pm h \partial_{x_k} W(x_i^h) + h \left(\frac{h}{2} \mathbf{e}_k \pm a \right) \cdot \nabla \partial_{x_k} W(x_i^h) + \mathcal{O}(h^3) \right] da \\ &= \pm \frac{h}{2} \partial_{x_k} W(x_i^h) + \frac{h^2}{4} \partial_{x_k}^2 W(x_i^h) + \mathcal{O}(h^3). \end{aligned}$$

Inside (83), this gives

$$\begin{aligned} e^{\mathfrak{W}(x_i^h + h\mathbf{e}_k) - \mathfrak{W}(x_i^h)} - e^{\mathfrak{W}(x_i^h) - \mathfrak{W}(x_i^h - h\mathbf{e}_k)} &= 2e^{\frac{h}{2}\partial_{x_k} W(x_i^h)} \sinh\left(\frac{h^2}{4}\partial_{x_k}^2 W(x_i^h)\right) + \mathcal{O}(h^3) \\ &= \frac{h^2}{2}\partial_{x_k}^2 W(x_i^h) + \mathcal{O}(h^3), \end{aligned}$$

and similarly for the other term involving exponentials. Therefore,

$$\begin{aligned} (83) &= \frac{1}{2} \sum_{k=1}^d (\partial_{x_k} W(x_i^h) + \mathcal{O}(h)) \frac{f(x_i^h + h\mathbf{e}_k) + f(x_i^h - h\mathbf{e}_k)}{2} \\ &= \frac{1}{2} \Delta W(x_i^h) f(x_i^h) + \mathcal{O}(h). \end{aligned}$$

Thanks to the invariance of (83) with respect to $h \mapsto -h$, all odd powers of h in the expansion need to vanish. So the error $\mathcal{O}(h)$ is actually even $\mathcal{O}(h^2)$.

Further, since \sinh is an odd function, the expansion of \mathfrak{W} directly implies that

$$\sinh(\mathfrak{W}(x_i^h \pm h\mathbf{e}_k) - \mathfrak{W}(x_i^h)) = \pm \frac{h}{2} \partial_{x_k} W(x_i^h) + \frac{h^2}{4} \partial_{x_k}^2 W(x_i^h) + \mathcal{O}(h^3).$$

This suffices to obtain

$$\begin{aligned} (84) &= \sum_{k=1}^d \partial_{x_k} W(x_i^h) \frac{f(x_i^h + h\mathbf{e}_k) - f(x_i^h - h\mathbf{e}_k)}{2h} \\ &\quad + \sum_{k=1}^d \partial_{x_k}^2 W(x_i^h) \frac{f(x_i^h + h\mathbf{e}_k) + f(x_i^h - h\mathbf{e}_k)}{4} + \mathcal{O}(h) \\ &= \nabla W(x_i^h) \cdot \nabla f(x_i^h) + \frac{1}{2} \Delta W(x_i^h) f(x_i^h) + \mathcal{O}(h). \end{aligned}$$

Again, the error is actually $\mathcal{O}(h^2)$ because of the symmetry of (84). This together yields (55).

REFERENCES

- [1] 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.
- [2] 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.
- [3] P. M. Bleher, J. L. Lebowitz, and E. R. Speer. Existence and positivity of solutions of a fourth-order nonlinear PDE describing interface fluctuations. *Comm. Pure Appl. Math.*, 47(7):923–942, 1994.
- [4] M. Bukal, E. Emmrich, and A. Jüngel. Entropy-stable and entropy-dissipative approximations of a fourth-order quantum diffusion equation. *Numer. Math.*, 127(2):365–396, 2014.
- [5] P. Caputo, P. Dai Pra, and G. Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009.
- [6] J. A. Carrillo, A. Jüngel, and S. Tang. Positive entropic schemes for a nonlinear fourth-order parabolic equation. *Discrete Contin. Dyn. Syst. Ser. B*, 3(1):1–20, 2003.
- [7] J. A. Carrillo and G. Toscani. Long-time asymptotics for strong solutions of the thin film equation. *Comm. Math. Phys.*, 225(3):551–571, 2002.
- [8] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.*, 203(3):969–1008, 2012.
- [9] S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.
- [10] P. Degond, F. Méhats, and C. Ringhofer. Quantum energy-transport and drift-diffusion models. *J. Stat. Phys.*, 118(3-4):625–667, 2005.

- [11] J. Denzler and R. J. McCann. Nonlinear diffusion from a delocalized source: affine self-similarity, time reversal, & nonradial focusing geometries. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 25(5):865–888, 2008.
- [12] B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn. Dynamics of an anchored Toom interface. *J. Phys. A*, 24(20):4805–4834, 1991.
- [13] B. Derrida, J. L. Lebowitz, E. R. Speer, and H. Spohn. Fluctuations of a stationary nonequilibrium interface. *Phys. Rev. Lett.*, 67(2):165–168, 1991.
- [14] K. Disser and M. Liero. On gradient structures for Markov chains and the passage to Wasserstein gradient flows. *Netw. Heterog. Media*, 10(2):233–253, 2015.
- [15] B. Düring, D. Matthes, and J. P. Milišić. A gradient flow scheme for nonlinear fourth order equations. *Discrete Contin. Dyn. Syst. Ser. B*, 14(3):935–959, 2010.
- [16] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.
- [17] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.
- [18] M. Erbar, J. Maas, and P. Tetali. Discrete Ricci curvature bounds for Bernoulli–Laplace and random transposition models. *Ann. Fac. Sci. Toulouse Math.*, 24(4):781–799, 2015.
- [19] M. Fathi and J. Maas. Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.*, to appear, *arXiv:1501.00562*, 2015.
- [20] A. Fiebach and A. Glitzky. Uniform estimate of the relative free energy by the dissipation rate for finite volume discretized reaction-diffusion systems. In *Finite volumes for complex applications. VII. Methods and theoretical aspects*, volume 77 of *Springer Proc. Math. Stat.*, pages 275–283. Springer, Cham, 2014.
- [21] U. Gianazza, G. Savaré, and G. Toscani. The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.*, 194(1):133–220, 2009.
- [22] N. Gigli and J. Maas. Gromov-Hausdorff convergence of discrete transportation metrics. *SIAM J. Math. Anal.*, 45(2):879–899, 2013.
- [23] A. Glitzky. Uniform exponential decay of the free energy for Voronoi finite volume discretized reaction-diffusion systems. *Math. Nachr.*, 284(17-18):2159–2174, 2011.
- [24] G. Grün. On the convergence of entropy consistent schemes for lubrication type equations in multiple space dimensions. *Math. Comp.*, 72(243):1251–1279 (electronic), 2003.
- [25] G. Grün and M. Rumpf. Nonnegativity preserving convergent schemes for the thin film equation. *Numer. Math.*, 87(1):113–152, 2000.
- [26] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [27] A. Jüngel. *Transport equations for semiconductors*, volume 773 of *Lecture Notes in Physics*. Springer-Verlag, Berlin, 2009.
- [28] A. Jüngel and D. Matthes. The Derrida-Lebowitz-Speer-Spohn equation: existence, nonuniqueness, and decay rates of the solutions. *SIAM J. Math. Anal.*, 39(6):1996–2015, 2008.
- [29] A. Jüngel and R. Pinnau. A positivity-preserving numerical scheme for a nonlinear fourth order parabolic system. *SIAM J. Numer. Anal.*, 39(2):385–406 (electronic), 2001.
- [30] A. Jüngel and G. Toscani. Exponential time decay of solutions to a nonlinear fourth-order parabolic equation. *Z. Angew. Math. Phys.*, 54(3):377–386, 2003.
- [31] A. Jüngel and I. Violet. First-order entropies for the Derrida-Lebowitz-Speer-Spohn equation. *Discrete Contin. Dyn. Syst. Ser. B*, 8(4):861–877, 2007.
- [32] M. Liero and A. Mielke. Gradient structures and geodesic convexity for reaction-diffusion systems. *Phil. Trans. Royal Soc. A.*, 2012.
- [33] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
- [34] D. Matthes, A. Jüngel, and G. Toscani. Convex Sobolev inequalities derived from entropy dissipation. *Arch. Ration. Mech. Anal.*, 199(2):563–596, 2011.
- [35] D. Matthes, J. McCann, and G. Savaré. A family of nonlinear fourth order equations of gradient flow type. *Comm. Partial Differential Equations*, 34(10-12):1352–1397, 2009.
- [36] D. Matthes and H. Osberger. A convergent Lagrangian discretization for a nonlinear fourth order equation. *Found. Comput. Math. (online first)*, pages 1–54, 2015.
- [37] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
- [38] R. J. McCann and C. Seis. The spectrum of a family of fourth-order nonlinear diffusions near the global attractor. *Communications in Partial Differential Equations*, 40(2):191–218, 2015.

- [39] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.
- [40] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48(1-2):1–31, 2013.
- [41] H. Osberger. Long-time behaviour of a fully discrete Lagrangian scheme for a family of fourth order. *ArXiv e-prints*, January 2015.
- [42] F. Otto and M. Westdickenberg. Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Anal.*, 37(4):1227–1255 (electronic), 2005.
- [43] A. Unterreiter, A. Arnold, P. Markowich, and G. Toscani. On generalized Csiszár-Kullback inequalities. *Monatsh. Math.*, 131(3):235–253, 2000.
- [44] L. Zhornitskaya and A. L. Bertozzi. Positivity-preserving numerical schemes for lubrication-type equations. *SIAM J. Numer. Anal.*, 37(2):523–555, 2000.

JAN MAAS, INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA (IST AUSTRIA),
AM CAMPUS 1, 3400 KLOSTERNEUBURG, AUSTRIA
E-mail address: `jan.maas@ist.ac.at`

DANIEL MATTHES, TECHNISCHE UNIVERSITÄT MÜNCHEN, ZENTRUM MATHEMATIK/M8, BOLTZMANNSTR. 3, D-85747 GARCHING, GERMANY
E-mail address: `matthes@ma.tum.de`