Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 0946 – 8633

On gradient structures for Markov chains and the passage to Wasserstein gradient flows

Karoline Disser, Matthias Liero

submitted: December 19, 2013

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: karoline.disser@wias-berlin.de
matthias.liero@wias-berlin.de

No. 1899
Berlin 2013

2010 *Mathematics Subject Classification.* 35K10, 35K20, 37L05, 49M25, 60F99, 65M08, 60J27, 70G75.

Key words and phrases. Wasserstein gradient flow, relative entropy, finite-volume scheme, entropy/entropy-dissipation formulation, gradient structures, Markov chain.

Acknowledgments. The work of K.D. was supported by ERC-2010-AdG no. 267802 “Analysis of Multiscale Systems Driven by Functionals”. The work of M.L. was supported by DFG Research Center “Mathematics for Key Technologies” MATHEON under project D22.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We study the approximation of Wasserstein gradient structures by their finite-dimensional analog. We show that simple finite-volume discretizations of the linear Fokker-Planck equation exhibit the recently established entropic gradient-flow structure for reversible Markov chains. Then we reprove the convergence of the discrete scheme in the limit of vanishing mesh size using only the involved gradient-flow structures. In particular, we make no use of the linearity of the equations nor of the fact that the Fokker-Planck equation is of second order.

1 Introduction

In this work we consider gradient structures for reversible continuous-time Markov chains on finite sets arising from finite-volume discretizations of drift-diffusion equations. We treat the most simple drift-diffusion problem, namely the linear Fokker-Planck equation with a drift coming from a given, sufficiently smooth potential Φ:

$$\dot{U} = \operatorname{div}(\nabla U + U\nabla\Phi) \qquad \text{in }]0,T[\times\Omega, \tag{1.1}$$

under suitable initial conditions and no-flux boundary conditions on a domain Ω. Since the seminal work of Otto [Ott98, JKO98, Ott01] it is known that the Fokker-Planck equation can be interpreted as a (metric) gradient flow in the space of probability measures $\mathcal{X}$ equipped with the Wasserstein distance $\mathrm{d_W}$ and with the relative entropy $\mathcal{E} : \mathcal{X} \to \mathbb{R}$ as driving functional. Recently, in [Mie11] it was shown that general reaction-diffusion systems with reactions of mass-action type, satisfying the detailed-balance condition, can also be written as gradient systems with respect to the relative entropy. The dissipation mechanism is given in terms of a so-called Onsager operator, i.e., the evolution of the system can be written in the form

$$\dot{u} = -\mathcal{K}(u)\mathrm{D}\mathcal{E}(u) =: -\nabla_{\mathcal{G}}\,\mathcal{E}(u) \quad \Leftrightarrow \quad \mathcal{G}(u)\dot{u} = -\mathrm{D}\mathcal{E}(u), \tag{1.2}$$

where the Onsager operator $\mathcal{K}(u)$ is a symmetric, positive semidefinite and, in general, state-dependent operator which maps thermodynamic forces to rates (see also [Ons31]). Moreover, $\mathcal{G}(u) = \mathcal{K}(u)^{-1}$ is the associated metric tensor and $\nabla_{\mathcal{G}}$ the metric gradient.

The reversible Markov chains discussed in this paper are special cases of reversible reactions, namely exchange reactions with a rate matrix $A \in \mathbb{R}^{n\times n}$ that lead to the linear ODE system $\dot{u} = Au$ in the state space

$$X_n = \{u \in \mathbb{R}^n |\, u_i > 0,\ \textstyle\sum_{i=1}^n u_i = 1\}.$$

Here, reversibility means that there exists a unique positive steady state $w \in X_n$ such that

$$A_{ij}w_j = A_{ji}w_i \quad \text{for all } i,\, j \in \{1,\dots,n\}.$$

(we include the irreducibility – the uniqueness of w – into the definition of reversibility). Thus, reversible Markov chains are a special case of more general reaction-diffusion systems with the gradient structure in the sense of (1.2) given in terms of the relative entropy E_n and the Onsager matrix $K_n(u)$:

$$E_n(u) = \sum_{i=1}^n u_i \log(u_i/w_i) \quad \text{and} \quad K_n(u) = \sum_{i<j} A_{ij}w_j \Lambda(\tfrac{u_i}{w_i}, \tfrac{u_j}{w_j})(e_i - e_j)\otimes(e_i - e_j),$$

where $\Lambda(a,b) = (a-b)/\log(a/b)$ denotes the logarithmic mean. By duality theory and chain rule, the finite-dimensional metric gradient flow in X_n can be equivalently formulated as entropy/entropy-dissipation balance

$$E_n(u(T)) + \int_0^T \big[R_n(u,\dot u) + R_n^*(u, -\mathrm{D}E_n(u))\big]\,\mathrm{d}t = E_n(u(0)), \tag{1.3}$$

where $R_n(u,\cdot)$ and $R_n^*(u,\cdot)$ are Legendre duals. We note that this entropic gradient structure for reversible Markov chains was found independently in [Mie13a, Maa11, CH*12]. We will review some of their results in Section 2.

In our case, (1.3) arises from a two-point flux finite-volume discretization scheme for the Fokker-Planck equation with the transmission rate coefficients containing the drift and the geometric information of the mesh. We (re)prove the convergence of the scheme by establishing a Γ-convergence-type result for the discrete entropy and dissipation functionals, hereby relying only on the structure in (1.3). In the limit of vanishing mesh size we obtain the integrated Wasserstein formulation of the Fokker-Planck equation, namely

$$\mathcal{E}(U(T)) + \frac{1}{2}\int_0^T \int_\Omega \Big[U|V|^2 + \frac{|U' + U\Phi'|^2}{U}\Big]\,\mathrm{d}x\,\mathrm{d}t \le \mathcal{E}(U(0)) \tag{1.4}$$

(we refer to Subsection 2.4 for a precise definition). Thus, we show that the Markov gradient structure gives a discrete counterpart to the continuous Wasserstein formulation. A crucial point here is that we do not exploit classical a priori estimates or compactness properties in Sobolev spaces.

We highlight that a related result was recently established in [GiM13]. It is based on the results in [Maa11] which characterize the Riemannian distances d_n on the manifolds X_n induced by the metric tensors $G_n(u) = K_n(u)^{-1}$. It is shown in [GiM13] that the metric spaces (X_n, d_n) converge in the sense of Gromov-Hausdorff to the (continuous) Wasserstein space $(\mathcal{X}, \mathrm{d}_\mathrm{W})$. Combined with the abstract convergence result in [Gig10] for metric gradient flows with geodesically convex and

Γ-converging driving functionals the limit passage for the discretized pure diffusion equation on the d-dimensional torus is shown.

In contrast, our result does not use the geodesic convexity of the relative entropies, a property that is hard to show and seems to hold in our setting only for (almost) equidistant discretizations (see [Mie13a, ErM12]). This makes our approach interesting for the extension to driving functionals which are even in the equidistant case not geodesically convex, see e.g. [ErM13] for an important example. Moreover, we allow for a drift given by the gradient of the potential Φ, opposed to the diffusion-only case in [GiM13].

Unfortunately, we are not able to cover the general higher-dimensional case. The problem lies in the construction of suitable interpolants for the discrete gradients and velocity fields, whose convergence is compatible with our variational approach. We will investigate this case in future work.

In [MaO13] another interesting numerical scheme for a nonlinear drift-diffusion equation on an interval is considered using the Lagrangian formulation of the problem. The discretization is based on the time-incremental formulation of the equation's gradient flow structure with respect to the (continuous) Wasserstein distance. In particular, no gradient structure of the discrete problems is exploited.

The limit passage in terms of gradient structures we present here is interesting for a number of reasons. The first is that the Wasserstein gradient structure provides a natural and physically meaningful formulation of the problem (see e.g. [AD*11] for the connection to large deviation principles for particle systems). It would be interesting to investigate if this meaning is reflected in the discretized structure.

The second reason is that Wasserstein gradient structures for diffusion problems and the related structures for reaction-diffusion systems introduced in [Mie11] can be found in a wide range of problems. Therefore, any method that uses only the structural properties of the systems has the potential for a wide application and helps to devise more efficient numerical schemes for reaction-diffusion problems.

Finally, it is in general of great interest to use variational tools such as Γ-convergence to pass to the limit in nonlinear time evolving systems that are driven by functionals, see e.g. [SaS04, MRS08, MiS11, LiS13, Lie13].

The outline of the paper is as follows. In Section 2 we briefly discuss the general setting of gradient systems upon which our result is based. We explain the well-known Wasserstein formulation of the Fokker-Planck equation in (1.4) and present the framework and the notation for gradient systems introduced in [Mie11] for reversible Markov chains using Onsager operators.

The Markov chains we consider here arise from finite-volume schemes of the Fokker-Planck equation. We discuss their derivation in Section 3. In particular, the Markov chains satisfy the reversibility condition by construction and we can rewrite the finite-volume scheme as a gradient-flow equation with respect to the relative

entropy. The discrete steady state is given by the discretization of its continuous counterpart. Moreover, we comment on the possibility of using different upwinding schemes such as the renowned Scharfetter-Gummel scheme. Finally, the main result is given in Theorem 3.1.

The limit passage is shown in Section 4. The crucial point is to establish lower liminf estimates for the discrete entropies and dissipation functionals in terms of the Wasserstein gradient structure. The first step consists of constructing suitable interpolants for the discrete quantities, where suitable means that the convergence of the interpolants is compatible with the variational convergence of the entropy and dissipation functionals. The lower estimate for the relative entropies follows easily from their lower semicontinuity properties with respect to weak* convergence in the space of probability measures. In order to prove the corresponding lower bound for the discrete dissipation functionals we exploit a useful lower semicontinuity result for the Wasserstein distance, which is an adaption of [AGS05, Theorem 5.4.4].

2 Abstract Gradient Flows

We first briefly discuss gradient flows on a smooth n-dimensional Riemannian manifold X in order to fix some notation and give heuristics for the well-known Wasserstein gradient structure of the Fokker-Planck equation and for the recent entropic gradient flow approach to its discretization. For a similar presentation we refer to [AM*12].

2.1 Riemannian point of view

On a smooth n-dimensional manifold X we consider a differentiable energy functional $E : X \to \mathbb{R}$ and a Riemannian metric g, i.e., a family of state-dependent inner products $g_u(v_1, v_2) = \langle G(u)v_1, v_2\rangle$ on the tangent spaces $\mathsf{T}_u X \simeq \mathbb{R}^n$ at u via a Riemannian tensor $G(u)$. Here, $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$ which gives the dual pairing on $\mathsf{T}_u X \times \mathsf{T}_u^* X$. The gradient flow of E in X is then given by the equation

$$\dot{u}(t) = -\nabla_G E(u(t)) \in \mathsf{T}_{u(t)} X. \tag{2.1}$$

Here and in the following we use overdots to denote time differentiation. The gradient $\nabla_G E(u)$ is characterized as the unique element of $\mathsf{T}_u X$ such that

$$\langle \mathrm{D}E(u), v\rangle = g_u(v, \nabla_G E(u)) \qquad \text{for all } v \in \mathsf{T}_u X.$$

It is given by $G(u)^{-1}\mathrm{D}E(u)$ with $\mathrm{D}E(u) \in \mathsf{T}_u^* X$ being the (Fréchet) differential of E. We can use the equivalence

$$\dot{u} = -\nabla_G E(u) \Leftrightarrow g_u(\dot{u}, -\nabla_G E(u)) \geq \frac{1}{2} g_u(\dot{u}, \dot{u}) + \frac{1}{2} g_u(\nabla_G E(u), \nabla_G E(u))$$

for the inner products to show that by the chain rule $\frac{\mathrm{d}}{\mathrm{d}t}E(u) = \langle \mathrm{D}E(u), \dot{u}\rangle$ the formulation

$$E(u(T)) + \frac{1}{2}\int_0^T \left[g_u(\dot{u},\dot{u}) + g_u(\nabla_G E(u), \nabla_G E(u))\right] \mathrm{d}t \le E(u(0)), \tag{2.2}$$

is equivalent to (2.1) for smooth solutions $u : [0,T] \to X$.

2.2 Onsager point of view

When considering Markov chains on finite domains, we will take on an "inverse" perspective. Given an equation in the form

$$\dot{u} = -K(u)\mathrm{D}E(u), \tag{2.3}$$

where $K(u)$ is a linear, symmetric, and positive definite operator for all $u \in X$, we call $K(u)$ an *Onsager operator* and (X, E, K) an *Onsager system* in relation to Onsager's famous principle [Ons31]. To this system we associate the dissipation potential

$$R(u; v) = \frac{1}{2}\langle K(u)^{-1}v, v\rangle \qquad \text{for } v \in \mathsf{T}_u X.$$

and the dual dissipation potential given by Legendre transform,

$$R^*(u;\xi) = \sup\left\{\langle \xi, v\rangle - R(u;v) \,|\, v \in \mathsf{T}_u X\right\} = \frac{1}{2}\langle \xi, K(u)\xi\rangle.$$

Then, as before, an equivalent formulation of the gradient flow equation (2.3) is

$$E(u(T)) + J(u; 0, T) \le E(u(0)),$$

where the dissipation functional

$$J(u; 0, T) = \int_0^T \left[R(u, \dot{u}) + R^*(u, -\mathrm{D}E(u))\right] \mathrm{d}t \tag{2.4}$$

gives the total dissipation along a curve. If $K(u)^{-1} =: G(u)$ defines a Riemannian metric on X, this is equivalent to (2.2).

The advantage of the Onsager formulation is that in the case of the Wasserstein and Markov chain gradient flows the operator K is explicitly given. Moreover, the Onsager structure allows in general for easy and thermodynamically consistent modeling e.g. of reaction-diffusion systems (see [Mie13b]).

2.3 Abstract metric setting

The above approaches in terms of the functional J can be generalized to an infinite-dimensional and non-smooth setting, given a topological space $\mathcal{X}$ and a suitable metric $\mathrm{d} : \mathcal{X} \times \mathcal{X} \to [0, \infty[$. The square norm of the gradient in (2.2) can be generalized by the square of the so-called *metric slope* $|\partial\mathcal{E}|_{\mathrm{d}}$ of $\mathcal{E} : \mathcal{X} \to]-\infty, \infty]$, given by

$$|\partial\mathcal{E}|_{\mathrm{d}}(u) := \limsup_{\widetilde{u}\to u} \frac{\big(\mathcal{E}(u) - \mathcal{E}(\widetilde{u})\big)_+}{\mathrm{d}(u, \widetilde{u})}. \tag{2.5}$$

Moreover, the square norm of the velocity vector of a curve u can be generalized to the square of the metric velocity $|\dot{u}|_{\mathrm{d}}$ of u via

$$|\dot{u}|_{\mathrm{d}}(t) := \lim_{h\to 0} \frac{\mathrm{d}(u(t), u(t+h))}{|h|}, \tag{2.6}$$

which exists for a.e. $t \in]0, T[$ if $u \in \mathrm{AC}(0, T; (\mathcal{X}, \mathrm{d}))$ is an absolutely continuous curve in $\mathcal{X}$ with respect to the distance d, see [AGS05, Theorem 1.1.2].

The dissipation functional J given in (2.4) is generalized to

$$\mathcal{J}(u; 0, T) := \frac{1}{2}\int_0^T \big[|\dot{u}|_{\mathrm{d}}^2 + |\partial\mathcal{E}|_{\mathrm{d}}(u)^2\big]\,\mathrm{d}t. \tag{2.7}$$

We use the following definition of a metric gradient flow and refer to [AGS05] for an extensive survey on this topic.

Definition 2.1 *Let $(\mathcal{X}, d)$ be a metric space, $\mathcal{E} : \mathcal{X} \to]-\infty, \infty]$ and $\mathcal{J}$ as in (2.7). Then a curve $u \in \mathrm{AC}(0, T; (\mathcal{X}, \mathrm{d}))$ is called a solution of the gradient system $(\mathcal{X}, \mathcal{E}, \mathrm{d})$ if $\mathcal{E}(u(0)) < \infty$ and*

$$\mathcal{E}(u(T)) + \mathcal{J}(u; 0, T) \le \mathcal{E}(u(0)). \tag{2.8}$$

In particular, the gradient flow satisfies the identity, if the additional property

$$\mathcal{E}(\widetilde{u}(s_1)) + \mathcal{J}(\widetilde{u}; s_0, s_1) \ge \mathcal{E}(\widetilde{u}(s_0)) \tag{2.9}$$

holds for all $\widetilde{u} \in \mathrm{C}([s_0, s_1]; \mathcal{X})$. This is guaranteed if the metric slope $|\partial\mathcal{E}|_{\mathrm{d}}$ is a *strong upper gradient* for $\mathcal{E}$ in $(\mathcal{X}, \mathrm{d})$, cf. [AGS05, Def. 1.2.1].

In the subsequent section, we briefly recall the Wasserstein formulation of the Fokker-Planck equation in the space of probability measures with respect to the relative entropy functional. Then we consider suitable finite-volume discretizations of this equation which carry an analog gradient structure in the space of probability measures on the finite domain.

2.4 Wasserstein formulation of the Fokker-Planck equation

In a bounded domain $\Omega \subset \mathbb{R}^d$ we consider the Fokker-Planck equation

$$\dot{U} = \operatorname{div}(\nabla U + U\nabla\Phi) \qquad \text{in } \Omega, \tag{2.10}$$

subject to the non-flux boundary condition $(\nabla U + U\nabla\Phi)\cdot\nu = 0$. Here, for simplicity, the drift potential Φ is assumed to satisfy

$$\Phi \in \mathrm{C}^1(\Omega;\mathbb{R}) \cap \mathrm{C_b}(\Omega), \tag{2.11}$$

such that the system has the steady state $W(x) = c_\Phi \mathrm{e}^{-\Phi(x)}$, bounded from above and strictly positive. Here, $c_\Phi := 1/\int_\Omega \mathrm{e}^{-\Phi(x)}\,\mathrm{d}x$ is the normalization constant. In order to introduce the gradient structure of [JKO98, Ott01], we define the space of probability measures on $\overline{\Omega}$

$$\mathcal{X} := \{\mu \in \mathcal{M}_{\geq 0}(\overline{\Omega}) \,|\, \mu(\overline{\Omega}) = 1\},$$

where $\mathcal{M}_{\geq 0}(\overline{\Omega})$ denotes the set of nonnegative Borel measures on $\overline{\Omega}$. The relative entropy functional on $\mathcal{X}$ with respect to the equilibrium density W is defined as

$$\mathcal{E}(\mu) := \begin{cases} \int_\Omega U \log(U/W)\,\mathrm{d}x & \text{if } \mu = U\,\mathrm{d}x, \\ +\infty & \text{otherwise.} \end{cases}$$

In the following, we will simply write $\mathcal{E}(\mu) = \mathcal{E}(U)$ if μ is absolutely continuous with density U and do not distinguish between measures and densities in this case. The space $\mathcal{X}$ is endowed with the weak* topology of measures and is (pseudo-)metrized by the 2-Wasserstein distance $\mathrm{d_W}$. It is well-known that $\mathrm{d_W}$ admits two interesting characterizations. The first is based on the theory of optimal transport, while the second is given in the form of a dynamical characterization and is well adapted to the gradient-flow setting. We briefly recall this characterization, which gives rise to a Riemannian structure in $\mathcal{X}$, see [BeB00, JKO98].

Definition 2.2 *Let $\mu :]s_0, s_1[\to \mathcal{X}$ be a family of measures and $V :]s_0, s_1[\times \Omega \to \mathbb{R}^d$ a measurable velocity field such that*

$$\int_{s_0}^{s_1} \int_\Omega |V(s,x)| \mu(s,\mathrm{d}x)\,\mathrm{d}s < +\infty.$$

We say that $(\mu, V) \in \mathrm{CE}$ (CE for Continuity Equation), if μ and V satisfy the continuity equation in the sense of distributions, i.e.

$$\int_{s_0}^{s_1} \int_{\mathbb{R}^d} (\dot{\Psi} + V \cdot \nabla\Psi)\mu(t,\mathrm{d}x)\,\mathrm{d}t = 0,$$

for all $\Psi \in \mathrm{C}_c^\infty(]s_0, s_1[\times\mathbb{R}^d)$, where μ and V are trivially extended by 0 outside of Ω.

For two measures μ_0 and μ_1 in $\mathcal{X}$ the distance $\mathrm{d_W}$ can be defined in terms of couples (μ, V) by the famous Benamou-Brenier characterization [BeB00]

$$\mathrm{d_W}(\mu_0,\mu_1)^2 = \min\Big\{ \int_0^1 \int_\Omega |V(t,x)|^2 \mu(t,\mathrm{d}x)\,\mathrm{d}t \;\Big|\; (\mu,V)\in \mathrm{CE},\ \mu(0){=}\mu_0,\ \mu(1){=}\mu_1\Big\}.$$

In particular, this identifies V as (Wasserstein) velocity field tangent to the curve μ. Moreover, this interpretation is reflected in the characterization of the metric time derivative in (2.6). We quote the following result from [AGS05, Theorem 8.3.1].

Proposition 2.3 *If $\mu \in \mathrm{C}(s_0,s_1;\mathcal{X})$ and V are such that $(\mu,V)\in\mathrm{CE}$, then $\mu \in \mathrm{AC}(s_0,s_1;X)$ and $|\dot\mu|_{\mathrm{d_W}}(t) \le \int_\Omega |V(t,x)|^2\mu(t,\mathrm{d}x)$ for almost all $t\in\,]s_0,s_1[$.*

Following Otto's formalism we can associate an Onsager operator with the Wasserstein distance. In particular, the velocity fields V are chosen to be the gradient of a function Ξ such that $\dot\mu = -\mathrm{div}(\mu\nabla\Xi) =: \mathcal{K}(\mu)\xi$. Then, the metric tensor $\mathcal{G}(\mu) = \mathcal{K}(\mu)^{-1}$ is induced by the identification $\dot\mu \mapsto \nabla\Xi_{\dot\mu}$ as follows:

$$\langle \mathcal{G}(\mu)\nu_1,\nu_2\rangle = \int_\Omega \nabla\Xi_{\nu_1}\cdot\nabla\Xi_{\nu_2}\mu(\mathrm{d}x).$$

In the sense of Definition 2.1, we recover the gradient system $(\mathcal{X},\mathcal{E},\mathcal{G})$ for the Fokker-Planck equation,

$$\begin{aligned}\mathcal{E}(U(T)) + \mathcal{J}(U;0,T) \le \mathcal{E}(U(0)),\quad &\text{where}\\ \mathcal{J}(U;0,T) = \frac12\int_0^T\int_\Omega \Big[U|V|^2 + \frac1U |W\nabla(U/W)\big|^2\Big]\,\mathrm{d}x\,\mathrm{d}t,\qquad &(2.12)\\ \text{with}\qquad (U,V)\in\mathrm{CE}.\end{aligned}$$

Here, $\mathcal{F}(U) = \frac12\int_\Omega |W\nabla(U/W)|^2/U\,\mathrm{d}x$ is the *Fisher information*, which gives the square of the metric slope of $\mathcal{E}$ (see (2.5)). Note that $\mathcal{F}(U) = 2\int_\Omega W|\nabla\sqrt{U/W}|^2\,\mathrm{d}x$, so that $\mathcal{F}(U)$ is finite if and only if $U\in \mathrm{W}^{1,1}(\Omega)$. We refer to [AGS05] for a comprehensive survey on the theory of Wasserstein gradient flows.

We call (2.12) the *entropy/entropy-dissipation formulation* of the Fokker-Planck equation in (2.10). Note that dissipation has the physical dimension of energy over time. Since the entropy has the dimension energy over temperature, we introduce the term "entropy-dissipation" to reflect this conceptual difference to energy dissipation.

2.5 Entropic gradient structure for reversible Markov chains

It was shown in [Maa11, Mie13a, CH*12] that the structure of the entropic Wasserstein gradient flows can be carried over from the continuum equation to Markov chains on finite domains. We will first briefly discuss a general class of reversible

Markov chains in this context and then later on consider processes which appear as finite-volume discretizations of equation (2.12) only. On $\{1,\dots,n\}$, $n \in \mathbb{N}$, we introduce the space

$$X_n := \{\, u \in \mathbb{R}^n \mid u_i > 0,\ \textstyle\sum_{i=1}^n u_i = 1 \,\}.$$

The evolution of a time-continuous Markov chain is given by the ODE system with transmission matrix $A \in \mathbb{R}^{n\times n}$, where A_{ij} is the rate for a particle moving from position j to i, viz.

$$\dot{u} = Au, \qquad \text{where } A_{ij} \geq 0 \text{ for } i \neq j \text{ and } A_{ii} = -\sum_{j:j\neq i} A_{ji}. \tag{2.13}$$

Many different gradient structures for the Markov chain (2.13) can be written down (see Remark 2.4). However, we are interested in a discrete Wasserstein-type gradient structure with respect to the discrete relative entropy functional. This entropic gradient structure was discovered in [Mie11, Sect. 3.1] and independently in [Maa11, CH*12]. The geodesic convexity of the relative entropy with respect to the Markov gradient structures was studied in [Mie13a] (see also [LiM12]). In order to state them we make two basic assumptions: (i) The Markov process is *irreducible*, i.e., for all i and j there is a path connecting both states.. (ii) The matrix A satisfies the *reversibility condition* also called *detailed-balance condition*, i.e.,

$$\pi_{ij} := A_{ij}w_j = A_{ji}w_i \qquad \text{for all } i,j \in \{1,\dots,n\}. \tag{2.14}$$

Clearly, conditions (i) and (ii) imply the existence of a unique strictly positive steady state $w \in X_n$ such that $Aw = 0$. With this we define the discrete relative entropy functional via

$$E_n(u) = \sum_{i=1}^n u_i \log(u_i/w_i).$$

Using reversibility and the calculation rules for the logarithm it is straightforward to check that (2.13) has the Onsager structure

$$\dot{u} = -K_n(u)\mathrm{D}E_n(u), \tag{2.15}$$

where $\mathrm{D}E_n(u) = (\log \rho_i)_i$ is the differential of E_n on X_n with $\rho_i := u_i/w_i$ denoting the relative density. The Onsager operator $K_n(u)$ is given via

$$K_n(u) = \sum_{i<j} \pi_{ij}\Lambda(\rho_i,\rho_j)(e_i - e_j)\otimes(e_i - e_j) \in \mathbb{R}^{n\times n}_{\mathrm{sym},\geq 0} \tag{2.16}$$

with e_i being the standard unit vectors in $\mathbb{R}^n$. Moreover, the function $\Lambda :]0,\infty[^2 \to]0,\infty[$ is the logarithmic mean of a and b and is given by

$$\Lambda(a,b) = \frac{a-b}{\log a - \log b} \qquad \text{for } a \neq b \qquad \text{and} \qquad \Lambda(a,a) = a. \tag{2.17}$$

Clearly, for each $u \in X_n$, $K_n(u)$ is a symmetric and positive semidefinite matrix with $\ker K_n(u) = \operatorname{span}\{(1,\dots,1)^{\mathsf{T}}\}$.

As in Subsection 2.2, we define the dual dissipation potential associated with the Onsager operator $K_n(u)$ via

$$R_n^*(u;\xi) = \frac{1}{2}\langle \xi, K_n(u)\xi\rangle = \frac{1}{2}\sum_{i<j} \pi_{ij}\Lambda(\rho_i,\rho_j)(\xi_i-\xi_j)^2.$$

The dissipation potential R_n is given by Legendre transform

$$R_n(u;v) = R_n^*(u;\xi_v) \qquad \text{with } v = K_n(u)\xi_v.$$

As before, it follows that an equivalent formulation of (2.15) is given by

$$E_n(u(T)) + \int_0^T \big[R_n(u;\dot u) + R_n^*(u;-\mathrm{D}E_n(u))\big]\,\mathrm{d}t \le E_n(u(0)). \tag{2.18}$$

As in the Wasserstein case, we call (2.18) the (discrete) entropy/entropy-dissipation formulation of the Markov chain equation in (2.13). Note that this structure on the open simplex X_n is also Riemannian in the sense of Subsection 2.1 with tensor $G_n(u) = K_n(u)^{-1}$ and inner products $g_u^n(v_1,v_2) = \langle v_1, G_n(u)v_2\rangle$ on the tangent space

$$\mathbb{R}^{n-1} \simeq \big\{v\in\mathbb{R}^n \mid \textstyle\sum_{i=1}^n v_i = 0\big\} \simeq \mathbb{R}^n \setminus \ker K_n(u).$$

As in Definition 2.2 for the Wasserstein case we introduce the discrete continuity equation and write

$$(u,\xi)\in \mathrm{CE}_n \quad \text{if } u\in \mathrm{C}^1_{\mathrm{p}}(0,1;X_n), \quad \dot u = K_n(u)\xi. \tag{2.19}$$

Here, $\mathrm{C}^1_{\mathrm{p}}(0,1;X_n)$ denotes the piecewise C^1 curves in X_n with respect to the euclidian metric. In particular, the matrix $K_n(u)$ induces a distance d_n, which is given by a discrete version of the Benamou-Brenier formula

$$\mathrm{d}_n(u_0,u_1)^2 = \min\big\{\textstyle\int_0^1 \big\langle \xi, K_n(u)\xi\big\rangle\,\mathrm{d}s \,\big|\, \dot u = K_n(u)\xi\big\},$$

where the minimization is over all curves $u\in \mathrm{C}^1_{\mathrm{p}}(0,1;X_n)$ connecting u_0 and u_1. We refer to [Maa11] for an extensive and rigorous study of X_n equipped with this structure.

Remark 2.4 *The ODE system in* (2.13) *is induced by many different gradient systems if the reversibility condition* (2.14) *holds. Indeed, for $\phi:\mathbb{R}_+\to\mathbb{R}$ strictly convex and twice differentiable we consider the driving functional given by $E_n^\phi(u) = \sum_{i=1}^n \phi(u_i/w_i)w_i$. Moreover, we define the Onsager matrix via*

$$K_n^\phi(u) = \sum_{i<j} \pi_{ij}\Theta(\tfrac{u_i}{w_i},\tfrac{u_j}{w_j})(e_i-e_j)\otimes(e_i-e_j),$$

where $\Theta(a,b) = (a-b)/(\phi'(a)-\phi'(b))$ and $\Theta(a,a) = 1/\phi''(a)$. Then, it is easy to check that the system $(X_n, E_n^\phi, K_n^\phi)$ also induces (2.13).

3 Discretization scheme in the one-dimensional case

In this section we discuss the finite-volume discretization of the one-dimensional Fokker-Planck equation in (2.10) using a simple two-point flux scheme. In particular, we highlight that the ODE system arising from this discretization scheme exhibits the Markov chain gradient structure detailed in Section 2.5.

Finite-volume methods are well adapted to drift-diffusion problems as they automatically conserve the local numerical fluxes between cells and hence the total mass. Moreover, they can be built to also conserve the positivity of solutions. These features make finite-volume methods quite attractive when modeling problems for which the flux is of importance, such as in fluid mechanics, semiconductor device simulation, heat and mass transfer, etc. (see e.g. [Bes12, BrF13, HDW12, CHGJ11, Gär11, FF*11, EyH08]). The good properties of the finite-volume method are due to its balance approach: a local balance is stated on each control volume. By the divergence formula, an integral formulation of the fluxes over the boundary of the control volume is then obtained. We refer to [EGH00] for a survey on finite-volume methods and to [FLL11] for a discussion of the properties of some finite-volume schemes.

3.1 Finite-volume discretization

In the open interval $\Omega =]0,1[$ we consider for each $n \in \mathbb{N}$ a partition $\Pi_n = \{x_i^n\}_{i=1}^n$ such that $0 = x_1^n < x_2^n < \ldots < x_n^n = 1$. Given such a partition, we introduce the $n+1$ midpoints between the vertices x_i^n via

$$\sigma_0^n = 0, \quad \sigma_n^n = 1, \qquad \sigma_i^n := \tfrac{1}{2}(x_{i+1}^n + x_i^n) \qquad \text{for } i = 1, \ldots, n-1.$$

In view of finite-volume schemes in higher dimensions we shall call σ_i^n (Voronoi) edge between x_i^n and x_{i+1}^n. In particular, the open interval $\omega_i^n =]\sigma_{i-1}^n, \sigma_i^n[$ denotes the Voronoi control volume with respect to the vertex x_i^n. We denote the length of the control volume ω_i^n by $h_i^n = \sigma_i^n - \sigma_{i-1}^n$ and set $h^n = \max_i h_i^n$, the fineness of the partition. In particular, we assume that the partitions satisfy $h^n \to 0$ as $n \to \infty$.

As in Subsection 2.4 we consider a potential $\Phi \in \mathrm{C}^1(\Omega) \cap \mathrm{C}_\mathrm{b}(\Omega)$ with which we associate the equilibrium density $W(x) = c_\Phi \mathrm{e}^{-\Phi(x)}$. We rewrite the Fokker-Planck equation in (2.10) using W and find

$$\dot{U} = \big(W(U/W)'\big)' \text{ in } \Omega \quad \text{and} \quad (U/W)'(t,x) = 0 \text{ for } x \in \{0,1\}. \tag{3.1}$$

Integrating the equation in (3.1) over the control volume ω_i^n gives

$$\tfrac{\mathrm{d}}{\mathrm{d}t} \textstyle\int_{\omega_i^n} U(t,x)\,\mathrm{d}x = f_i - f_{i-1},$$

where $f_i = (W(U/W)')(\sigma_i^n)$ denotes the flux across the edge σ_i^n. We expect $u_i(t)$ to approximate $\int_{\omega_i^n} U(t,x)\,\mathrm{d}x$ and that the fluxes f_i can be approximated by $\kappa_i(\frac{u_{i+1}}{w_{i+1}} - \frac{u_i}{w_i})$. Here, w denotes the discrete counterpart of the steady state W which we set to $w_i = c_n W(x_i^n)h_i^n$ with c_n such that $w \in X_n$. In particular, c_n tends to 1 as $n \to \infty$ by definition. Moreover, κ_i are given transmission coefficients satisfying the consistency condition $\kappa_i(x_{i+1}^n - x_i^n)/W(\sigma_i) \to 1$. Many different choices for κ_i are possible (see Subsection 3.2 for a discussion). Following [Mie13a, Sect. 5] we use for simplicity the geometric mean of $W(x_{i+1}^n)$ and $W(x_i^n)$, i.e.,

$$\kappa_i = \frac{\sqrt{W(x_{i+1}^n)W(x_i^n)}}{x_{i+1}^n - x_i^n}, \qquad \text{for } i = 1,\dots,n-1. \tag{3.2}$$

which obviously gives a consistent discretization scheme in the above sense.

Introducing the rate coefficients $\alpha_i = \kappa_i/w_i$ and $\beta_i = \kappa_i/w_{i+1}$ for $i = 1,\dots,n-1$ we can write the discretization of the Fokker-Planck equation in (3.1) as

$$\dot{u}_i = \alpha_{i-1}u_{i-1} - (\alpha_i+\beta_{i-1})u_i + \beta_i u_{i+1} \qquad \text{for } i = 2,\dots,n-1, \tag{3.3}$$

while at the boundary we have $\dot{u}_1 = \beta_1 u_2 - \alpha_1 u_1$ and $\dot{u}_n = \alpha_{n-1}u_{n-1} - \beta_{n-1}u_n$. For the initial value, given $U_0 \in \mathcal{X}$ such that $\mathcal{E}(U_0) < \infty$, we define

$$(u_0^n)_i := \int_{\omega_i^n} U_0(x)\,\mathrm{d}x, \tag{3.4}$$

which yields $u_0^n \in X_n$.

From this discretization scheme we obtain a Markov chain with tridiagonal transmission matrix $Q \in \mathbb{R}^{n\times n}$, viz.

$$Q = \begin{pmatrix} -\alpha_1 & \alpha_1 & 0 & \dots & 0 \\ \beta_1 & -(\alpha_2+\beta_1) & \alpha_2 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \beta_{n-2} & -(\alpha_{n-1}+\beta_{n-2}) & \alpha_{n-1} \\ 0 & \dots & 0 & \beta_{n-1} & -\beta_{n-1} \end{pmatrix}.$$

Obviously, the detailed balance condition $\alpha_i w_i = \beta_i w_{i+1}$ is automatically satisfied. Hence, we are in the situation of Subsection 2.5 and can provide an entropic gradient structure for the finite-volume discretization. In particular, the Onsager matrix takes the form

$$K_n(u) := \sum_{i=1}^{n-1} \kappa_i \Lambda(\rho_{i+1},\rho_i)(e_{i+1}-e_i)\otimes(e_{i+1}-e_i),$$

where as before $e_i \in \mathbb{R}^n$ is the ith unit vector, $\rho_i = u_i/w_i$ are the relative densities, and $\Lambda(a,b)$ denotes the logarithmic mean of a and b. With the Onsager matrix

$K_n(u)$ we can associate the total dissipation functional for a curve $u \in \mathrm{C}^1_{\mathrm{p}}(0,T;X_n)$, cf. (2.19), which is given by

$$J_n(u;0,T) := \int_0^T \big[R_n(u,\dot u) + F_n(u)\big]\,\mathrm{d}t,$$

where $R_n(u,\dot u) = \frac{1}{2}\langle K_n(u)^{-1}\dot u, \dot u\rangle$ and $F_n(u) = \frac{1}{2}\langle \mathrm{D}E_n(u), K_n(u)\mathrm{D}E_n(u)\rangle$ are the discrete dissipation potential and the discrete Fisher information, respectively. The former can be written as

$$R_n(u,\dot u) = \frac{1}{2}\sum_{i=1}^{n-1} \frac{q_i^2}{\kappa_i \Lambda(\rho_{i+1},\rho_i)} \qquad \text{with} \quad q_i = \sum_{k=1}^{i} \dot u_k \tag{3.5}$$

and the latter as

$$F_n(u) = \frac{1}{2}\sum_{i=1}^{n-1} \frac{\kappa_i(\rho_{i+1}-\rho_i)^2}{\Lambda(\rho_{i+1},\rho_i)}. \tag{3.6}$$

Following Definition 2.1, we say that $u^n \in \mathrm{C}([0,T];X_n)$ is a solution for the gradient flow problem associated with (3.3) if

$$E_n\big(u^n(T)\big) + J_n(u^n;0,T) \le E_n\big(u_0^n\big). \tag{3.7}$$

By classical Markov theory, for every $n \in \mathbb{N}$, given a partition Π_n and an initial value $u_0^n \in X_n$, a solution u^n of (3.7) exists and even equality is satisfied in (3.7). We prove the convergence of the piecewise constant interpolant associated with u^n to a solution U of (2.12) using only the gradient structure of (3.7). In particular, the main results reads as follows.

Theorem 3.1 *Let the potential Φ be given as in (2.11), let the initial value $U_0 \in \mathcal{X}$ be such that $\mathcal{E}(U_0) < \infty$, and consider a sequence of partitions $(\Pi_n)_{n\in\mathbb{N}}$ such that $h^n \to 0$ as $n \to \infty$. Moreover, let $(u^n)_{n\in\mathbb{N}}$ be a sequence of solutions of the entropy/entropy-dissipation formulation (3.7) with initial values $(u_0^n)_{n\in\mathbb{N}}$ as in (3.4). Then, denoting by $U^n(t,x) = u_i^n(t)/h_i^n$ for $x \in \omega_i^n$ the piecewise constant interpolant in $\mathcal{X}$ one has up to subsequences $U^n \overset{*}{\rightharpoonup} U$ in $\mathcal{M}_{\ge 0}([0,T]\times\overline{\Omega})$ with $U \in \mathrm{AC}(0,T;\mathcal{X})$ solving*

$$\mathcal{E}\big(U(T)\big) + \frac{1}{2}\int_0^T \int_\Omega \Big[U|V|^2 + \frac{|U' + U\Phi'|^2}{U}\Big]\,\mathrm{d}x\,\mathrm{d}t \le \mathcal{E}(U_0),$$

where $(U,V) \in \mathrm{CE}$, i.e., V gives the metric time derivative of U.

3.2 On the choice of transmission coefficients

In the last section we chose the transmission coefficients κ_i to be given by the geometric mean in the form

$$\kappa_i = \frac{\theta_{\mathrm{geo}}\big(W(x_{i+1}^n), W(x_i^n)\big)}{x_{i+1}^n - x_i^n} \quad \text{with} \quad \theta_{\mathrm{geo}}(a,b) = \sqrt{ab}.$$

However, other choices of mean functions θ are possible. Indeed, the specific form of θ is crucial for the properties of the finite volume discretization (see e.g. [FLL11]). Numerical stability for the drift-diffusion equations is increased substantially by the use of upwind schemes in general. Here, "upwind" refers to the property that the direction of the drift is respected by the discretization.

It will become clear in the proof of the convergence result in the subsequent section that we only need the mean function θ to have the natural property

$$\min\{a,b\} \le \theta(a,b) \le \max\{a,b\}. \tag{3.8}$$

An important example with excellent stability properties is given by the Scharfetter-Gummel scheme (see e.g. [Mar86, Bes12]), which is widely used in semiconductor device simulation. In our setting the Scharfetter-Gummel scheme reads

$$\theta_{\mathrm{SG}}(a,b) = \Lambda\big(\tfrac{1}{a}, \tfrac{1}{b}\big)^{-1},$$

where Λ denotes as before the logarithmic mean. We easily check that this particular choice is also admissible in the sense of (3.8).

4 Limit passage

In this section we connect the gradient structures of the Fokker-Planck equation in Subsection 2.4 and of Markov chains in Subsection 2.5 by proving the limit passage stated in Theorem 3.1. As described above, the motivating question is whether we may pass to the limit in the entropy/entropy-dissipation formulation for the Markov chain in (3.7). This question has two parts: (i) Do (interpolants of) solutions of (3.7) with uniformly bounded initial entropies have beneficial compactness properties that allow us to extract subsequences converging in a suitable topology τ? (ii) Using this topology can we show the lower estimate

$$u^n \xrightarrow{\tau} U: \quad \liminf_{n\to\infty} J_n(u^n;0,T) \ge \mathcal{J}(U;0,T),$$

where J_n is the dissipation functional in the discrete case and $\mathcal{J}$ its Wasserstein counterpart in (2.12)? Our answers to these questions are affirmative. In particular, we use only information associated with the gradient structures for the proofs.

4.1 Interpolation of discrete quantities on Ω

The first step in the convergence proof is the embedding of the discrete problems introduced in the last section into the continuum setting. The crucial point is the construction of interpolants defined on the whole of Ω. We show that the interpolants converge in a suitable sense to limits which permit lower liminf estimates for the discrete dissipation functionals and relative entropies.

With a given vector $u = (u_1, \dots, u_n) \in X_n$ we associate the piecewise constant interpolant U^n via

$$U^n(x) = U_i := u_i/h_i^n \qquad \text{for } x \in \omega_i^n. \tag{4.1}$$

Analogously, we define the piecewise constant interpolant $W^n(x) = W_i := w_i/h_i^n$ for $x \in \omega_i^n$, such that $W^n \to W$ in $\mathrm{L}^\infty(\Omega)$, as W is uniformly continuous on Ω. In particular, this definition allows us to rewrite the discrete relative entropy as

$$E_n(u) = \mathcal{E}_n(U^n) := \int_\Omega U^n \log(U^n/W^n)\,\mathrm{d}x. \tag{4.2}$$

In order to prove lower limits for the dissipation potentials we have to find a suitable estimate for the logarithmic mean $\Lambda(\rho_{i+1}, \rho_i)$. Here, in view of the elementary estimate $\Lambda(a,b) \le (a+b)/2$ it is natural to define interpolants by taking the maximum value with respect to adjacent control volumes, i.e.

$$\widetilde{U}^n(x) := \max\{U_{i-1}, U_i, U_{i+1}\} \quad \text{for } x \in \omega_i^n, \quad i = 2, \dots, n-1, \tag{4.3a}$$

while at the boundary cells we set

$$\widetilde{U}^n(x) := \begin{cases} \max\{U_1, U_2\} & \text{for } x \in \omega_1^n, \\ \max\{U_{n-1}, U_n\} & \text{for } x \in \omega_n^n. \end{cases} \tag{4.3b}$$

Replacing max with min and U_i with W_i in the definition above we define the interpolant $x \mapsto \widetilde{W}^n(x)$ analogously. It is easy to check that $\widetilde{W}^n \to W$ in $\mathrm{L}^\infty(\Omega)$, too. Using the definitions of $\widetilde{U}^n$ and $\widetilde{W}^n$ we arrive at the estimate

$$\widetilde{W}^n(x)/\widetilde{U}^n(x) \le 1/\Lambda(\rho_{i\pm1}, \rho_i) \qquad \text{for } x \in \omega_i^n, \quad i = 2, \dots, n-1.$$

Finally, we introduce the discrete gradient of the relative density $\rho_i = u_i/w_i$ and the interpolant Q^n for the discrete fluxes q_i in (3.5) as the piecewise constant interpolants G^n, Q^n via

$$G^n(x) = \frac{\rho_{i+1} - \rho_i}{x_{i+1}^n - x_i^n} \quad \text{and} \quad Q^n(x) = q_i \quad \text{for } x \in \left]x_i^n, x_{i+1}^n\right[, \quad i = 1, \dots, n-1. \tag{4.4}$$

We are now in a position to state the following result which gives a lower estimate for the discrete dissipation functional R_n in (3.5) and the discrete Fisher information F_n in (3.6) for fixed n.

Lemma 4.1 *For a curve $u \in \mathrm{C}^1_{\mathrm{p}}(0,T;X_n)$ let $\widetilde{U}^n$, G^n and Q^n be the interpolants defined in* (4.3) *and* (4.4)*, then for all $t \in [0,T]$,*

$$R_n(u(t),\dot{u}(t)) \geq \frac{\gamma_n}{2}\int_\Omega \frac{|Q^n(t)|^2}{\widetilde{U}^n(t)}\,\mathrm{d}x \quad \textit{and} \quad F_n(u(t)) \geq \frac{1}{2}\int_\Omega \frac{(\widetilde{W}^n)^2}{\widetilde{U}^n(t)}|G^n(t)|^2\,\mathrm{d}x$$

with $\gamma_n \to 1$ for $n \to \infty$.

Proof: For notational simplicity we drop the dependence on t in the following. We use the following estimate for the logarithmic mean

$$\Lambda(\rho_{i+1},\rho_i) \leq \frac{1}{2}\left(\frac{U_{i+1}}{W_{i+1}}+\frac{U_i}{W_i}\right) \leq \frac{\max\{U_{i+1},U_i\}}{2}\left(\frac{1}{W_{i+1}}+\frac{1}{W_i}\right).$$

In particular, using the definition of the transmission coefficients κ_i in (3.2) we obtain

$$\lambda_i := \kappa_i\,\Lambda(\rho_{i+1},\rho_i) \leq \frac{1}{2}\left(\sqrt{\frac{W_{i+1}}{W_i}}+\sqrt{\frac{W_i}{W_{i+1}}}\right)\frac{\max\{U_{i+1},U_i\}}{x^n_{i+1}-x^n_i}.$$

Due to the assumptions on the potential Φ the term in the parentheses tends to 2 uniformly in i. Hence, we can assume that $\gamma_n\lambda_i \leq \max\{U_{i+1},U_i\}/(x^n_{i+1}-x^n_i)$ with constants γ_n satisfying $\gamma_n \to 1$ as $n \to \infty$. Summing over cells instead of edges (hence counting each edge twice) we estimate

$$\begin{aligned} R_n(u,\dot{u}) &= \frac{1}{4}\frac{q_1^2}{\lambda_1}+\frac{1}{4}\sum_{i=2}^{n-1}\left(\frac{q_{i-1}^2}{\lambda_{i-1}}+\frac{q_i^2}{\lambda_i}\right)+\frac{1}{4}\frac{q_{n-1}^2}{\lambda_{n-1}} \\ &\geq \frac{\gamma_n}{2}\sum_{i=1}^{n}\int_{\sigma^n_{i-1}}^{\sigma^n_i}\frac{|Q^n|^2}{\widetilde{U}^n}\,\mathrm{d}x = \frac{\gamma_n}{2}\int_\Omega \frac{|Q^n|^2}{\widetilde{U}^n}\,\mathrm{d}x, \end{aligned}$$

where we have used that $x^n_i - \sigma^n_{i-1} = (x^n_i - x^n_{i-1})/2$ for $i = 2,\dots,n$ and $\sigma^n_i - x^n_i = (x^n_{i+1}-x^n_i)/2$ for $i = 1,\dots,n-1$.

The proof of the estimate for the discrete Fisher information follows along the same lines noting that the estimate

$$\frac{1}{\kappa_i}\Lambda(\rho_{i+1},\rho_i) \leq \frac{1}{x^n_{i+1}-x^n_i}\frac{\max\{U_{i+1},U_i\}}{\min\{W_{i+1},W_i\}^2} \qquad \text{for } i = 1,\dots,n-1$$

is satisfied.

∎

Combining (4.2) and Lemma 4.1 we arrive at the following proposition which is the starting point for the limit passage $n \to \infty$ in the subsequent section.

Proposition 4.2 *Let $u \in \mathrm{C}^1_{\mathrm{p}}(0,T,X_n)$ denote a solution of the discrete entropy/entropy-dissipation formulation in* (3.7)*. Then, the interpolants U^n, $\widetilde{U}^n$, G^n and Q^n in* (4.1)*,* (4.3) *and* (4.4) *satisfy the entropy/entropy-dissipation formulation*

$$\mathcal{E}_n\big(U^n(T)\big) + \frac{1}{2}\int_0^T \int_\Omega \frac{1}{\widetilde{U}^n}\big[|Q^n|^2 + (\widetilde{W}^n)^2|G^n|^2\big]\,\mathrm{d}x\,\mathrm{d}t \le \mathcal{E}_n(U_0^n) \tag{4.5}$$

with Q^n associated with $\dot{U}^n$ in virtue of

$$\dot{U}^n(t,x_i^n) = \frac{Q^n(t,\sigma_i) - Q^n(t,\sigma_{i-1})}{\sigma_i - \sigma_{i-1}} \quad \text{for } i = 1,\dots,n \tag{4.6}$$

with boundary conditions $Q^n(t,\sigma_0) = Q^n(t,\sigma_n) \equiv 0$.

4.2 Proof of the main result

In this section we provide the actual proof of the limit passage stated in Theorem 3.1. The main step relies in establishing a lower liminf estimate for the entropies and dissipation functionals J_n. The notion of convergence that we use here is that of weak* convergence in the space of probability measures $\mathcal{X}$. More precisely, we write $\mu^n \overset{*}{\rightharpoonup} \mu$ in $\mathcal{X}$ if

$$\forall\, \Psi \in \mathrm{C_b}(\Omega): \quad \int_\Omega \Psi(x)\,\mu^n(\mathrm{d}x) \longrightarrow \int_\Omega \Psi(x)\,\mu(\mathrm{d}x) \quad \text{as } n \to \infty.$$

Note that since the domain Ω is bounded, $\mathcal{X}$ is compact within this topology by Prokhorov's theorem.

After having embedded the discrete solutions of the Markov chain the next step consists of describing the compactness properties of the interpolants U^n. More precisely, we show that we can extract (not relabeled) subsequences which converge pointwise for each $t \in [0,T]$ in $\mathcal{X}$. In particular, we show that the limit is continuous with respect to the 1-Wasserstein distance. Here, we follow the proof of [AM*12, Theorem 3.1] which is based on the dual formulation of the 1-Wasserstein distance:

$$\mathrm{d_{W_1}}(\mu_1,\mu_2) = \sup\Big\{ \int_\Omega \Psi(x)\mu_1(\mathrm{d}x) - \int_\Omega \Psi(x)\mu_2(\mathrm{d}x) \;\Big|\; \Psi \in \mathrm{C}^{0,1}(\Omega),\ \|\Psi'\|_\infty \le 1\Big\}.$$

We show that the family of discrete solutions is equicontinuous with respect to $\mathrm{d_{W_1}}$ and apply a metric Arzelà-Ascoli theorem. To shorten notation we introduce the open set $\Omega_T =]0,T[\times \Omega$.

Proposition 4.3 *With the same assumptions as in Theorem 3.1 let u^n be the solution of the discrete entropy/entropy-dissipation formulation in* (3.7) *and U^n the associated piecewise constant interpolant, then, up to subsequences, $U^n(t) \overset{*}{\rightharpoonup} \mu(t)$ in $\mathcal{X}$ for all $t \in [0,T]$ and the limit measure satisfies $\mu \in \mathrm{C}(0,T;\mathcal{X})$.*

Proof: We consider $\Psi \in \mathrm{C}^{0,1}(\Omega)$ satisfying $|\Psi'(x)| \le 1$ for every $x \in \Omega$. As before we set $\Psi_i = \Psi(x_i^n)$ and define the piecewise constant interplant $\Psi^n(x) = \Psi_i$ for $x \in \omega_i^n$. The dual formulation of the 1-Wasserstein distance gives for $0 \le t_0 < t_1 \le T$

$$\begin{aligned}\mathrm{d}_{\mathrm{W}_1}\big(U^n(t_0), U^n(t_1)\big) &\le \int_\Omega \Psi^n(x) U^n(t_0,x)\,\mathrm{d}x - \int_\Omega \Psi^n(x) U^n(t_1,x)\,\mathrm{d}x + \varepsilon_n \\ &= \int_{t_0}^{t_1}\int_\Omega \Psi^n(x)\dot U^n(t,x)\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_n,\end{aligned}$$

where $\varepsilon_n = 2\|\Psi - \Psi^n\|_\infty$. Using summation by parts, the discrete continuity equation in (4.6) then yields

$$\begin{aligned}&\int_{t_0}^{t_1}\int_\Omega \Psi^n(x)\dot U^n(t,x)\,\mathrm{d}x\,\mathrm{d}t = -\int_{t_0}^{t_1}\int_\Omega Q^n(t,x)\Psi'(x)\,\mathrm{d}x\,\mathrm{d}t \\ &\le \int_{t_0}^{t_1}\sum_{i=1}^{n-1}(x_{i+1}^n - x_i^n)|q_i(t)|\,\mathrm{d}t \le C\sqrt{t_1-t_0}\left(\int_0^T \sum_{i=1}^{n-1}\frac{|q_i(t)|^2}{\kappa_i \Lambda(\rho_{i+1}(t),\rho_i(t))}\,\mathrm{d}t\right)^{1/2}.\end{aligned}$$

Here, we used Hölder's inequality and the conservation of total mass $\sum_{i=1}^n u_i = 1$. Indeed, exploiting $\Lambda(a,b) \le (a+b)/2$ we obtain the estimate

$$\sum_{i=1}^{n-1}(x_{i+1}^n - x_i^n)\sqrt{W_{i+1}W_i}\,\Lambda(\rho_{i+1},\rho_i) \le \sum_{i=1}^{n-1}\frac{x_{i+1}^n - x_i^n}{2}\left(\sqrt{\frac{W_{i+1}}{W_i}}\frac{u_i}{h_i^n} + \sqrt{\frac{W_i}{W_{i+1}}}\frac{u_{i+1}}{h_{i+1}^n}\right).$$

Noting that $h_{i+1}^n \ge (x_{i+1}^n - x_i^n)/2$ and $h_i^n \ge (x_{i+1}^n - x_i^n)/2$ we see that the sum is uniformly bounded. Finally, since $R_n(u,\dot u)$ is uniformly bounded in $\mathrm{L}^1(0,T)$ we have shown the equicontinuity of $t \mapsto U^n(t) \in \mathcal{X}$. Applying the metric Arzelà-Ascoli theorem [AGS05, Proposition 3.3.1] we obtain a (not relabeled) subsequence with $U^n(t) \overset{*}{\rightharpoonup} \mu(t)$ in $\mathcal{X}$ for all $t \in [0,T]$ and $\mu \in \mathrm{C}(0,T;\mathcal{X})$. ∎

Using the lower semicontinuity properties of the relative entropies (see e.g. [AGS05, Lemma 9.4.3]) and the convergence $W_n \to W$ in $\mathrm{L}^\infty(\Omega)$ we obtain the following corollary.

Corollary 4.4 *The limit μ in Proposition 4.3 satisfies*

$$\text{for all } t \in [0,T]: \quad \liminf_{n\to\infty} E_n\big(u^n(t)\big) \ge \mathcal{E}\big(\mu(t)\big).$$

It is easy to see that the limiting measures $\mu(t)$ are absolutely continuous with respect to the Lebesgue measure on Ω and we denote $\mu(t,\mathrm{d}x) = U(t,x)\,\mathrm{d}x$. Indeed, let us consider the time-reversed curves $\widehat u^n(t) = u^n(T-t)$, which satisfy $J_n(\widehat u^n;0,T) = J_n(u^n;0,T)$. Using the property (2.9) and the uniform boundedness

of the initial entropies and the dissipation functional we get $\sup_n E_n(u^n(t)) < \infty$ for every $t \in [0,T]$.

Moreover, we can canonically identify the curves U^n with elements in $\mathcal{M}_{\geq 0}(\overline{\Omega_T})$. In particular, we also have the weak* convergence of U^n in this space, i.e.,

$$\forall \Psi \in \mathrm{C}(\overline{\Omega_T}): \qquad \lim_{n\to\infty} \int_0^T \int_\Omega \Psi U^n \,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_\Omega \Psi U \,\mathrm{d}x\,\mathrm{d}t. \tag{4.7}$$

Also note that the construction of the initial values u_0^n in (3.4) yields $U_0^n \overset{*}{\rightharpoonup} U_0$ in $\mathcal{X}$. Moreover, U_0^n is a recovery sequence for $\mathcal{E}_n \overset{\Gamma}{\to} \mathcal{E}$, namely, using Jensen's inequality we obtain

$$\limsup_{n\to\infty} E_n(u_0^n) = \limsup_{n\to\infty} \mathcal{E}_n(U_0^n) \leq \mathcal{E}(U_0) < \infty. \tag{4.8}$$

In the following lemma we show that the interpolant $\widetilde{U}^n$ defined in (4.3) also converges weakly* to U.

Lemma 4.5 *With the same assumptions as in Theorem 3.1 let U^n and $\widetilde{U}^n$ be the interpolants, given via* (4.1) *and* (4.3)*, associated with the solution u^n of the discrete entropy/entropy-dissipation formulation in* (3.7)*, then $U^n - \widetilde{U}^n \to 0$ in $\mathrm{L}^1(\Omega_T)$ for $n \to \infty$.*

Proof: We again start out from a pointwise estimate in time (omitting again the dependence on t if possible). Denoting $\widetilde{U}_i = \widetilde{U}^n(x)$ for $x \in \omega_i^n$ we compute

$$\big\|U^n - \widetilde{U}^n\big\|_{\mathrm{L}^1(\Omega)} = \sum_{i=1}^n h_i^n \big|U_i - \widetilde{U}_i\big| \leq \sum_{i\in I_+} h_i^n \big|U_i - U_{i+1}\big| + \sum_{i\in I_-} h_i^n \big|U_i - U_{i-1}\big|,$$

where the index sets $I_\pm$ correspond to the cases in which the maximum in (4.3) is attained at the left and right of i, respectively. The relative densities ρ_i satisfy $U_i = W_i\rho_i$. Hence, the discrete analog of the product rule $f_{i+1}g_{i+1} - f_ig_i = f_{i+1}(g_{i+1}-g_i) + g_i(f_{i+1}-f_i)$ leads to the estimate

$$\begin{aligned}\sum_{i\in I_+} h_i^n \big|U_i - U_{i+1}\big| &\leq \sum_{i\in I_+} h_i^n \Big(\rho_i \big|W_{i+1} - W_i\big| + W_{i+1}\big|\rho_{i+1} - \rho_i\big|\Big) \\ &\leq \sum_{i\in I_+} \Big(u_i \big|\tfrac{W_i}{W_{i+1}} - 1\big| + h^n W_{i+1} |\rho_i - \rho_{i+1}|\Big),\end{aligned}$$

where we used that $u_i = h_i^n U_i$ and $h^n = \max_i h_n^i$. The first term vanishes as $n \to \infty$ due to the continuity of $x \mapsto W(x)$ and $\sum_{i=1}^n u_i^n = 1$, which holds at every time. The second term can be estimated from above using the inequality,

$$h^n \sum_{i\in I_+} W_{i+1} |\rho_i - \rho_{i+1}| \leq h^n \left(\sum_{i\in I_+} \frac{\kappa_i(\rho_{i+1}-\rho_i)^2}{\Lambda(\rho_{i+1},\rho_i)}\right)^{1/2} \left(\sum_{i\in I_+} \frac{W_{i+1}^2}{\kappa_i} \Lambda(\rho_{i+1},\rho_i)\right)^{1/2} \tag{4.9}$$

The first term is the square root of the discrete Fisher information $F_n(u)$. It is uniformly bounded in $\mathrm{L}^2(0,T)$ by assumption. Using the definition of κ_i in (3.2) and that $\sum_{i=1}^n u_i = 1$, we find that the second term is even uniformly bounded in time. More precisely, we compute

$$\begin{aligned}\frac{W_{i+1}^2}{\kappa_i}\Lambda(\rho_{i+1},\rho_i) &\le \frac{(x_{i+1}^n - x_i^n)W_{i+1}^2}{2\sqrt{W_{i+1}W_i}}\Big(\frac{U_{i+1}}{W_{i+1}} + \frac{U_i}{W_i}\Big)\\ &\le \frac{x_{i+1}^n - x_i^n}{2h_{i+1}^n}\Big(\frac{W_{i+1}}{W_i}\Big)^{\frac12}u_{i+1} + \frac{x_{i+1}^n - x_i^n}{2h_i^n}\Big(\frac{W_{i+1}}{W_i}\Big)^{\frac32}u_i,\end{aligned}$$

where we used the elementary estimate $\Lambda(a,b) \le (a+b)/2$ again. Hence, we have shown the strong convergence $U^n - \widetilde{U}^n \to 0$ in $\mathrm{L}^1(\Omega_T)$. ∎

Corollary 4.6 *The interpolant $\widetilde{U}^n$ in* (4.3) *associated with the solution u of the discrete entropy/entropy-dissipation formulation in* (3.7) *converges to U as in* (4.7).

We are now in position to proof the lower liminf estimate for the dissipation functional. The proof of the following proposition is based on [AGS05, Theorem 5.4.4.].

Proposition 4.7 *With the same assumptions as in Theorem 3.1 let u^n be the solution of the discrete entropy/entropy-dissipation formulation in* (3.7) *and U the limit in* (4.7)*, then*

$$\liminf_{n\to\infty}\int_0^T \big[R_n(u^n,\dot u^n) + F_n(u^n)\big]\,\mathrm{d}t \ge \frac12\int_0^T\int_\Omega \Big[U|V|^2 + \frac{|U' + U\Phi'|^2}{U}\Big]\,\mathrm{d}x\,\mathrm{d}t, \quad (4.10)$$

where $(U,V) \in \mathrm{CE}$, see Definition 2.2.

Proof: We define the velocity fields $V^n = Q^n/\widetilde{U}^n$ and $\widehat{V}^n = \widetilde{W}^n G^n/\widetilde{U}^n$. In particular, due to (4.5) and the boundedness of the initial entropies we have

$$\sup_n \int_0^T\int_\Omega \big[|V^n|^2 + |\widehat{V}^n|^2\big]\widetilde{U}^n\,\mathrm{d}x\,\mathrm{d}t < \infty. \quad (4.11)$$

We proceed in four steps:

1. Extraction of converging subsequences. Here, we only argue for V^n, the case of $\widehat{V}^n$ being analog. For brevity we denote $y = (t,x) \in \Omega_T$. Moreover, in the space $\mathcal{M}_{\ge0}(\overline{\Omega_T}\times\mathbb{R})$ we introduce the family of measures given by the push-forward $\boldsymbol{\mu}^n = (\mathrm{id}\times V^n)_\#\widetilde{U}^n$. In particular, for all suitably integrable functions $g : \overline{\Omega_T}\times\mathbb{R} \to \mathbb{R}$ we have

$$\int_{\mathbb{R}}\int_{\Omega_T} g(y,v)\,\boldsymbol{\mu}^n(\mathrm{d}y,\mathrm{d}v) = \int_{\Omega_T} g\big(y, V^n(y)\big)\,\widetilde{U}^n(y)\,\mathrm{d}y.$$

We aim to show that the family of measures $\boldsymbol{\mu}^n$ converges weakly* (up to subsequences) to a limit $\boldsymbol{\mu} \in \mathcal{M}_{\geq 0}(\overline{\Omega_T}\times\mathbb{R})$ and that a limit velocity field V can be recovered from $\boldsymbol{\mu}$. For this we denote by $\pi^1(y,v) = y$ and $\pi^2(y,v) = v$ the canonical projections onto Ω_T and $\mathbb{R}$, respectively. We note that the first marginal of $\pi^1_\#\boldsymbol{\mu}^n = \widetilde{U}^n$ converges weakly in $\mathcal{M}_{\geq 0}(\overline{\Omega_T})$. Moreover, the second marginal satisfies due to (4.11)

$$\sup_n \int_{\mathbb{R}} |v| \left(\pi^2_\#\boldsymbol{\mu}^n\right)(\mathrm{d}v) = \sup_n \int_{\Omega_T} \left|V^n(y)\right| \widetilde{U}^n(y)\,\mathrm{d}y < +\infty.$$

Hence, by [AGS05, Lemma 5.2.2] the sequence $\boldsymbol{\mu}^n$ is relatively compact in $\mathcal{M}(\overline{\Omega_T}\times\mathbb{R})$ and we can find a (not relabeled) subsequence and a limit $\boldsymbol{\mu} \in \mathcal{M}_{\geq 0}(\overline{\Omega_T}\times\mathbb{R})$ such that $\boldsymbol{\mu}^n \overset{*}{\rightharpoonup} \boldsymbol{\mu}$ in $\mathcal{M}_{\geq 0}(\overline{\Omega_T}\times\mathbb{R})$. In particular, since by (4.11) $v \mapsto |v|$ is uniformly integrable with respect to $\boldsymbol{\mu}^n$ and by [AGS05, Proposition 5.1.10], this yields the convergence

$$\lim_{n\to\infty} \int_{\mathbb{R}}\int_{\Omega_T} \Psi(y) v\, \boldsymbol{\mu}^n(\mathrm{d}y,\mathrm{d}v) = \int_{\mathbb{R}}\int_{\Omega_T} \Psi(y) v\, \boldsymbol{\mu}(\mathrm{d}y,\mathrm{d}v), \tag{4.12}$$

where $\Psi \in \mathrm{C}(\overline{\Omega_T})$ is an arbitrary test function.

Let us denote by $\mu_y \in \mathcal{M}_{\geq 0}(\mathbb{R})$ the disintegration of $\boldsymbol{\mu}$ with respect to the limit measure $\mu(\mathrm{d}y)=U(y)\mathrm{d}y$, which is μ-a.e. uniquely determined, see [AGS05, Sect. 5.3]. In particular, for every bounded or nonnegative measurable function $f : \Omega_T\times\mathbb{R} \to \mathbb{R}$, the disintegration satisfies

$$\int_{\mathbb{R}}\int_{\Omega_T} f(y,v)\, \boldsymbol{\mu}(\mathrm{d}y,\mathrm{d}v) = \int_{\Omega_T} \left[\int_{\mathbb{R}} f(y,v)\, \mu_x(\mathrm{d}v)\right] U(y)\,\mathrm{d}y.$$

Choosing $f(y,v) = v$ gives the barycentric projection $V(y) = \int_{\mathbb{R}} v\, \mu_y(\mathrm{d}v)$ of the measure $\boldsymbol{\mu}$. Now, using the definition of $\boldsymbol{\mu}^n$ and the convergence in (4.12) with $\Psi \in \mathrm{C}(\overline{\Omega_T})$ arbitrary we arrive at

$$\lim_{n\to\infty} \int_{\Omega_T} \Psi(y) V^n(y) \widetilde{U}^n(y)\,\mathrm{d}y = \int_{\Omega_T} \Psi(y) V(y)\, U(y)\,\mathrm{d}y.$$

2. Lower liminf estimate. Let $g : \mathbb{R} \to [0,\infty[$ be convex. Part (d) of Lemma 5.1.12 in [AGS05] and Jensen's inequality yield

$$\begin{aligned}
\liminf_{n\to\infty} \int_{\Omega_T} g\big(V^n(y)\big)\, \widetilde{U}^n(y)\,\mathrm{d}y &= \liminf_{n\to\infty} \int_{\mathbb{R}}\int_{\Omega_T} g(v)\, \boldsymbol{\mu}^n(\mathrm{d}y,\mathrm{d}v) \\
&\geq \int_{\mathbb{R}}\int_{\Omega_T} g(v)\, \boldsymbol{\mu}(\mathrm{d}y,\mathrm{d}v) \geq \int_{\Omega_T} g\left(\int_{\mathbb{R}} v\, \mu_y(\mathrm{d}v)\right) U(y)\,\mathrm{d}y \\
&= \int_{\Omega_T} g(V(y))\, U(y)\,\mathrm{d}y.
\end{aligned}$$

In particular, for $g(v) = |v|^2$ we obtain

$$\liminf_{n\to\infty} \int_0^T \big[R_n(u,\dot u) + F_n(u)\big]\,\mathrm{d}t \geq \frac{1}{2}\int_0^T \int_\Omega U(t,x)\Big[|V(t,x)|^2 + |\widehat V(t,x)|^2\Big]\,\mathrm{d}x\,\mathrm{d}t.$$

3. Identification of the limit V. In this step we verify that the limits U and V satisfy the Wasserstein continuity equation, i.e., $(U,V) \in \mathrm{CE}$. To this end, let us consider a test function $\Psi \in \mathrm{C}_c^\infty(]0,T[\times\mathbb{R})$ and define $\Psi_i(t) = \Psi(t,x_i^n)$ as well as the piecewise constant interpolant Ψ^n, i.e., we have $\Psi^n(t,x) = \Psi_i(t)$ for $x \in \omega_i^n$. Using integration by parts and (4.6) we obtain

$$-\int_0^T \int_\Omega U^n(t,x)\dot\Psi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \sum_{i=1}^n \Psi_i(t)\big\{Q^n(\sigma_i,t) - Q^n(\sigma_{i-1},t)\big\}\,\mathrm{d}t - \varepsilon_n,$$

where $\varepsilon_n = \|\dot\Psi - \dot\Psi^n\|_\infty \to 0$ as $n\to\infty$. Hence, summation by parts leads to the identity

$$\begin{aligned} -\int_0^T \int_\Omega U^n(t,x)\dot\Psi(t,x)\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_n &= \int_0^T \sum_{i=1}^{n-1} Q^n(t,\sigma_i)\big\{\Psi_i(t) - \Psi_{i+1}(t)\big\}\,\mathrm{d}t \\ &= -\int_0^T \int_\Omega V^n(t,x)\Psi'(t,x)\widetilde U^n(t,x)\,\mathrm{d}t, \end{aligned} \tag{4.13}$$

where we have used the boundary conditions $Q^n(t,\sigma_0) = Q^n(t,\sigma_n) = 0$. Now, using the convergence of U^n, $\widetilde U^n$ and V^n as in (4.12), we can pass to the limit in (4.13) to find

$$\int_0^T \int_\Omega U(t,x)\dot\Psi(t,x)\,\mathrm{d}x\,\mathrm{d}t = \int_0^T \int_\Omega V(t,x)\Psi'(t,x)U(t,x)\,\mathrm{d}t,$$

which is the verification of the continuity equation. Notice that although Ω is bounded, the continuity equation is posed on the whole real line $\mathbb{R}$ and therefore provides a weak formulation of the Neumann boundary conditions.

4. Identification of the limit $\widehat V$. We show that $\widehat V = U'/U + \Psi'$, which identifies the continuum Fisher information. We fix again a smooth test vector field $\Psi \in \mathrm{C}_\mathrm{c}^\infty(]0,T[\times\overline\Omega)$ and using (4.4) we compute

$$\begin{aligned} \int_0^T \int_\Omega \widehat V^n \Psi \widetilde U^n\,\mathrm{d}x\,\mathrm{d}t &= \int_0^T \sum_{i=1}^{n-1} \frac{\rho_{i+1}-\rho_i}{x_{i+1}^n - x_i^n} \int_{x_i^n}^{x_{i+1}^n} \widetilde W^n \Psi\,\mathrm{d}x\,\mathrm{d}t \\ &= \int_0^T \sum_{i=1}^{n-1} \big(\rho_{i+1}(t) - \rho_i(t)\big)\Big\{W(\sigma_i)\Psi(t,\sigma_i) \\ &\qquad + \frac{1}{x_{i+1}^n - x_i^n}\int_{x_i^n}^{x_{i+1}^n} \big[\widetilde W^n(x)\Psi(t,x) - W(\sigma_i)\Psi(t,\sigma_i)\big]\,\mathrm{d}x\Big\}\,\mathrm{d}t. \end{aligned}$$

Due to the assumptions on the potential Φ and the smoothness of the test field Ψ the last term in the braces vanishes as $n \to \infty$. Hence, with summation by parts we obtain

$$\begin{aligned}\int_0^T \int_\Omega \widehat{V}^n \Psi \widetilde{U}^n \,\mathrm{d}x\,\mathrm{d}t &= -\int_0^T \sum_{i=1}^n \rho_i(t)\big\{W(\sigma_i)\Psi(t,\sigma_i) - W(\sigma_{i-1})\Psi(t,\sigma_{i-1})\big\}\,\mathrm{d}t + \varepsilon_n \\ &= \int_0^T \int_\Omega U^n \frac{(W\Psi)'}{W^n}\,\mathrm{d}x\,\mathrm{d}t + \varepsilon_n.\end{aligned}$$

Finally, passing to the limit $n \to \infty$ and exploiting that $1/W \in \mathrm{L}^\infty(\Omega)$ yields the identity

$$\int_0^T \int_\Omega \widehat{V}\Psi U \,\mathrm{d}x\,\mathrm{d}t = -\int_0^T \int_\Omega U \frac{(W\Psi)'}{W}\,\mathrm{d}x\,\mathrm{d}t = -\int_0^T \int_\Omega U(\Psi' - \Phi'\Psi)\,\mathrm{d}x\,\mathrm{d}t.$$

Hence, we obtain $\widehat{V} = U'/U + U\Phi'$. ∎

Remark 4.8 *For the limit passage we only exploit the lower liminf estimate for the dissipation functional J_n. However, to show Γ-convergence of J_n we additionally need to construct recovery sequences $\widehat{u}^n \in \mathrm{C}^1_{\mathrm{p}}(0,T;X_n)$ for every $\widehat{U} \in \mathrm{AC}(0,T;\mathcal{X})$ such that the associated interpolants converge in $\mathcal{M}_{\geq 0}(\overline{\Omega_T})$ and*

$$\limsup_{n\to\infty} J_n(\widehat{u}^n;0,T) \leq \mathcal{J}(\widehat{U};0,T).$$

Indeed, the existence of such a sequence is easy to show for sufficiently smooth functions $\widehat{U}$, e.g. by setting $\widehat{u}^n_i(t) = c_n \widehat{U}(t,x^n_i)h^n_i$. For the general case we then argue as usual by density.

Proof of Theorem 3.1: Since the limit satisfies $(U,V) \in \mathrm{CE}$, i.e., the Wasserstein continuity equation is satisfied, we have $U \in \mathrm{AC}(0,T;\mathcal{X})$ by Proposition 2.3. Moreover, combining Corollary 4.4, Proposition 4.7, and the limsup estimate for the initial entropies in (4.8) we arrive at

$$\begin{aligned}\mathcal{E}(U(T)) + \mathcal{J}(U;0,T) &\leq \liminf_{n\to\infty} \big[E_n\big(u^n(T)\big) + J_n(u^n;0,T)\big] \\ &\leq \limsup_{n\to\infty} E_n(u^n(0)) \leq \mathcal{E}(U_0).\end{aligned}$$

Hence, with Definition 2.1 we have shown that U is a solution for the Wasserstein formulation of the Fokker-Planck equation. ∎

Acknowledgment

The authors are grateful to Alexander Mielke for proposing this problem to them. Moreover, they thank him and Alexander Linke for stimulating discussions and many helpful comments.

References

[AD*11] S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations principle to the Wasserstein gradient flow: A new micro-macro passage. *Communications in Mathematical Physics*, 307(3), 791–815, 2011.

[AGS05] 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, 2005.

[AM*12] S. Arnrich, A. Mielke, M. A. Peletier, G. Savaré, and M. Veneroni. Passing to the limit in a Wasserstein gradient flow: from diffusion to reaction. *Calc. Var. Part. Diff. Eqns.*, 44, 419–454, 2012.

[BeB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3), 375–393, 2000.

[Bes12] M. Bessemoulin-Chatard. A finite volume scheme for convection-diffusion equations with nonlinear diffusion derived from the Scharfetter-Gummel scheme. *Numer. Math.*, 121(4), 637–670, 2012.

[BrF13] A. Bradji and J. Fuhrmann. Some abstract error estimates of a finite volume scheme for a nonstationary heat equation on general nonconforming multidimensional spatial meshes. *Appl. Math.*, 58(1), 1–38, 2013.

[CH*12] 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. Rational Mech. Anal.*, 203(3), 969–1008, 2012.

[CHGJ11] C. Chainais-Hillairet, M. Gisclon, and A. Jüngel. A finite-volume scheme for the multidimensional quantum drift-diffusion model for semiconductors. *Numer. Methods Partial Differential Equations*, 27(6), 1483–1510, 2011.

[EGH00] R. Eymard, T. Gallouët, and R. Herbin. The finite volume method. In *Handbook of Numerical Analysis*, volume VII, pages 715–1022. Amsterdam: North Holland, 2000.

[ErM12] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Rational Mech. Anal.*, 206(3), 997–1038, 2012.

[ErM13] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *arXiv:1212.1129*, 2013.

[EyH08] R. Eymard and J.-M. Hérard, editors. *Finite volumes for complex applications V*. ISTE, London, 2008.

[FF*11] J. Fořt, J. Fürst, Jiříand Halama, R. Herbin, and F. Hubert, editors. *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, volume 4 of *Springer Proceedings in Mathematics*, Heidelberg, 2011. Springer.

[FLL11] J. Fuhrmann, A. Linke, and H. Langmach. A numerical method for mass conservative coupling between fluid flow and solute transport. *Appl. Numer. Math.*, 61(4), 530–553, 2011.

[Gär11] K. Gärtner. Charge transport in semiconductors and a finite volume scheme. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, volume 4 of *Springer Proc. Math.*, pages 513–522. Springer, Heidelberg, 2011.

[Gig10] N. Gigli. On the heat flow on metric measure spaces: existence, uniqueness and stability. *Calc. Var. Partial Differential Equations*, 39(1-2), 101–120, 2010.

[GiM13] N. Gigli and J. Maas. Gromov-Hausdorff convergence of discrete transportation metrics. *SIAM J. Math. Anal.*, 45(2), 879–899, 2013.

[HDW12] D. Hilhorst, H. C. V. Do, and Y. Wang. A finite volume method for density driven flows in porous media. In *CEMRACS'11: Multiscale coupling of complex models in scientific computing*, volume 38 of *ESAIM Proc.*, pages 376–386. EDP Sci., Les Ulis, 2012.

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Analysis*, 29(1), 1–17, 1998.

[Lie13] M. Liero. Passing from bulk to bulk-surface evolution in the Allen-Cahn equation. *NoDEA Nonlinear Differential Equations Appl.*, 20(3), 919–942, 2013.

[LiM12] M. Liero and A. Mielke. Gradient structures and geodesic convexity for reaction-diffusion systems. *Phil. Trans. Royal Soc. A*, 2012. To appear. WIAS preprint 1701 (April 2012).

[LiS13] M. Liero and U. Stefanelli. Weighted inertia-dissipation-energy functionals for semilinear equations. *Boll. Unione Mat. Ital. (9)*, 6(1), 1–27, 2013.

[Maa11] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261, 2250–2292, 2011.

[MaO13] D. Matthes and H. Osberger. Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation. *ESAIM: Mathematical Modelling and Numerical Analysis*, eFirst, 12 2013.

[Mar86] P. A. Markowich. *The stationary semiconductor device equations*. Computational Microelectronics. Springer-Verlag, Vienna, 1986.

[Mie11] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24, 1329–1346, 2011.

[Mie13a] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Part. Diff. Eqns.*, 48(1), 1–31, 2013. Online doi:10.1007/s00526-012-0538-8.

[Mie13b] A. Mielke. Thermomechanical modeling of energy-reaction-diffusion systems, including bulk-interface interactions. *Discr. Cont. Dynam. Systems Ser. S*, 6(2), 479–499, 2013.

[MiS11] A. Mielke and U. Stefanelli. Weighted energy-dissipation functionals for gradient flows. *ESAIM Control Optim. Calc. Var.*, 17, 52–85, 2011.

[MRS08] A. Mielke, T. Roubíček, and U. Stefanelli. Γ-limits and relaxations for rate-independent evolutionary problems. *Calc. Var. Part. Diff. Eqns.*, 31, 387–416, 2008.

[Ons31] L. Onsager. Reciprocal relations in irreversible processes, I+II. *Physical Review*, 37, 405–426, 1931. (part II, 38:2265-227).

[Ott98] F. Otto. Dynamics of labyrinthine pattern formation in magnetic fluids: a mean-field theory. *Arch. Rational Mech. Anal.*, 141(1), 63–103, 1998.

[Ott01] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26, 101–174, 2001.

[SaS04] E. Sandier and S. Serfaty. Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Comm. Pure Appl. Math.*, LVII, 1627–1672, 2004.