

## ON GRADIENT STRUCTURES FOR MARKOV CHAINS AND THE PASSAGE TO WASSERSTEIN GRADIENT FLOWS

KAROLINE DISSER AND MATTHIAS LIERO

Weierstrass Institute  
Mohrenstraße 39  
10117 Berlin, Germany

(Communicated by Benedetto Piccoli)

**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 simplest drift-diffusion problem, namely the linear Fokker-Planck equation with a drift coming from a given, sufficiently smooth potential  $\Phi$ :

$$\dot{U} = (U' + U\Phi)' \quad \text{in } ]0, T[ \times \Omega, \quad (1)$$

under suitable initial conditions and no-flux boundary conditions on an interval  $\Omega \subset \mathbb{R}$ . Since the seminal work of Otto [32, 19, 33] 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  $d_W$  and with the relative entropy  $\mathcal{E} : \mathcal{X} \rightarrow \mathbb{R}$  as driving functional. Recently, in [26] 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

$$\dot{u} = -\mathcal{K}(u)D\mathcal{E}(u) =: -\nabla_{\mathcal{G}} \mathcal{E}(u) \quad \Leftrightarrow \quad \mathcal{G}(u)\dot{u} = -D\mathcal{E}(u), \quad (2)$$

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

---

2010 *Mathematics Subject Classification.* 35K10, 35K20, 37L05, 49M25, 60F99, 65M08, 60J27, 70G75, 82B35.

*Key words and phrases.* Wasserstein gradient flow, discrete gradient flow structures, entropy/entropy-dissipation formulation, evolutionary  $\Gamma$ -convergence, Markov chains.

The first author was supported by the European Research Council via “ERC-2010-AdG no. 267802 (*Analysis of Multiscale Systems Driven by Functionals*)”. The second author has received funding from the DFG Research Center MATHEON in project D22.

also [31]). Moreover,  $\mathcal{G}(u)$  denotes the metric tensor corresponding to  $\mathcal{K}(u)$  and  $\nabla_{\mathcal{G}}$  denotes 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

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

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 (2) given in terms of the relative entropy  $E_n$  and the Onsager matrix  $K_n(u)$ :

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

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

$$E_n(u(T)) + \int_0^T [R_n(u, \dot{u}) + R_n^*(u, -DE_n(u))] dt = E_n(u(0)), \quad (3)$$

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 [27, 23, 8]. We will review some of their results in Section 2.

In our case, (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  $\Gamma$ -convergence-type result for the discrete entropy and dissipation functionals, hereby relying only on the structure in (3) (see e.g. [11, Theorem 4.2] for a convergence result based on classical PDE methods). In the limit of vanishing mesh size we obtain the integrated Wasserstein formulation of the Fokker-Planck equation, namely

$$\mathcal{E}(U(T)) + \frac{1}{2} \int_0^T \int_{\Omega} \left[ U|V|^2 + \frac{|U' + U\Phi'|^2}{U} \right] dx dt \leq \mathcal{E}(U(0)) \quad (4)$$

(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 [17]. It is based on the results in [23] which characterize the Riemannian distances  $d_n$  on the manifolds  $X_n$  induced by the metric tensors  $G_n(u) = K_n(u)^{-1}$ , where by definition and the positivity of  $u \in X_n$ ,  $\mathcal{K}(u)$  is invertible with respect to the subspace of  $\mathbb{R}^n$  of rates preserving mass. It is shown in [17] that the metric spaces  $(X_n, d_n)$  converge in the sense of Gromov-Hausdorff to the (continuous) Wasserstein space  $(\mathcal{X}, d_W)$ . Combined with the abstract convergence result in [16] for metric gradient flows with

geodesically convex and  $\Gamma$ -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 [27, 9]). This makes our approach interesting for the extension to driving functionals which are even in the equidistant case not geodesically convex, see e.g. [10] for an important example. Moreover, we allow for a drift given by the gradient of the potential  $\Phi$ , opposed to the diffusion-only case in [17].

In [25] 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. [1] 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 [26] 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  $\Gamma$ -convergence to pass to the limit in nonlinear time evolving systems that are driven by functionals, see e.g. [34, 29, 30, 22, 20].

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 (4) and present the framework and the notation for gradient systems introduced in [26] 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 [2, 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 [3].

**2.1. Riemannian point of view.** On a smooth  $n$ -dimensional manifold  $X$  we consider a differentiable energy functional  $E : X \rightarrow \mathbb{R}$  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  $\mathbb{T}_u X \simeq \mathbb{R}^n$  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  $\mathbb{T}_u X \times \mathbb{T}_u^* X$ . The gradient flow of  $E$  in  $X$  is then given by the equation

$$\dot{u}(t) = -\nabla_G E(u(t)) \in \mathbb{T}_{u(t)} X. \quad (5)$$

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  $\mathbb{T}_u X$  such that

$$\langle DE(u), v \rangle = g_u(v, \nabla_G E(u)) \quad \text{for all } v \in \mathbb{T}_u X.$$

It is given by  $G(u)^{-1}DE(u)$  with  $DE(u) \in \mathbb{T}_u^* X$  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  $\frac{d}{dt}E(u) = \langle DE(u), \dot{u} \rangle$  the formulation

$$E(u(T)) + \frac{1}{2} \int_0^T [g_u(\dot{u}, \dot{u}) + g_u(\nabla_G E(u), \nabla_G E(u))] dt \leq E(u(0)), \quad (6)$$

is equivalent to (5) for smooth solutions  $u : [0, T] \rightarrow X$ .

**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

$$\dot{u} = -K(u)DE(u), \quad (7)$$

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 [31]. To this system we associate the dissipation potential

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

and the dual dissipation potential given by Legendre transform,

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

Then, as before, an equivalent formulation of the gradient flow equation (7) is

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

where the dissipation functional

$$J(u; 0, T) = \int_0^T [R(u, \dot{u}) + R^*(u, -DE(u))] dt \quad (8)$$

gives the total dissipation along a curve. If  $K(u)^{-1} =: G(u)$  defines a Riemannian metric on  $X$ , this is equivalent to (6).

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 [28]).

**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  $d : \mathcal{X} \times \mathcal{X} \rightarrow [0, \infty[$ . The square norm of the gradient in (6) can be generalized by the square of the so-called *metric slope*  $|\partial\mathcal{E}|_d$  of  $\mathcal{E} : \mathcal{X} \rightarrow ]-\infty, \infty]$ , given by

$$|\partial\mathcal{E}|_d(u) := \limsup_{\tilde{u} \rightarrow u} \frac{(\mathcal{E}(u) - \mathcal{E}(\tilde{u}))_+}{d(u, \tilde{u})}. \tag{9}$$

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

$$|\dot{u}|_d(t) := \lim_{h \rightarrow 0} \frac{d(u(t), u(t+h))}{|h|}, \tag{10}$$

which exists for a.e.  $t \in ]0, T[$  if  $u \in \text{AC}(0, T; (\mathcal{X}, d))$  is an absolutely continuous curve in  $\mathcal{X}$  with respect to the distance  $d$ , see [2, Theorem 1.1.2].

The dissipation functional  $J$  given in (8) is generalized to

$$\mathcal{J}(u; 0, T) := \frac{1}{2} \int_0^T [|\dot{u}|_d^2 + |\partial\mathcal{E}|_d(u)^2] dt. \tag{11}$$

We use the following definition of a metric gradient flow and refer to [2] for an extensive survey on this topic.

**Definition 2.1.** Let  $(\mathcal{X}, d)$  be a metric space,  $\mathcal{E} : \mathcal{X} \rightarrow ]-\infty, \infty]$  and  $\mathcal{J}$  as in (11). Then a curve  $u \in \text{AC}(0, T; (\mathcal{X}, d))$  is called a solution of the gradient system  $(\mathcal{X}, \mathcal{E}, d)$  if  $\mathcal{E}(u(0)) < \infty$  and

$$\mathcal{E}(u(T)) + \mathcal{J}(u; 0, T) \leq \mathcal{E}(u(0)). \tag{12}$$

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

$$\mathcal{E}(\tilde{u}(s_1)) + \mathcal{J}(\tilde{u}; s_0, s_1) \geq \mathcal{E}(\tilde{u}(s_0)) \tag{13}$$

holds for all  $\tilde{u} \in C([s_0, s_1]; \mathcal{X})$ . This is guaranteed if the metric slope  $|\partial\mathcal{E}|_d$  is a *strong upper gradient* for  $\mathcal{E}$  in  $(\mathcal{X}, d)$ , cf. [2, Definition 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

$$\dot{U} = \text{div}(\nabla U + U \nabla \Phi) \quad \text{in } \Omega, \tag{14}$$

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

$$\Phi \in C^1(\Omega; \mathbb{R}) \cap C_b(\Omega), \tag{15}$$

such that the system has the steady state  $W(x) = c_\Phi e^{-\Phi(x)}$ , bounded from above and strictly positive. Here,  $c_\Phi := 1/\int_\Omega e^{-\Phi(x)} dx$  is the normalization constant. In order to introduce the gradient structure of [19, 33], we define the space of probability measures on  $\bar{\Omega}$

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

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

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

In the following, we will simply write  $\mathcal{E}(\mu) = \mathcal{E}(U)$  if  $\mu$  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  $d_W$ . It is well-known that  $d_W$  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 [4, 19].

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

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

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

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

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

For two measures  $\mu_0$  and  $\mu_1$  in  $\mathcal{X}$  the distance  $d_W$  can be defined in terms of couples  $(\mu, V)$  by the famous Benamou-Brenier characterization [4]

$$d_W(\mu_0, \mu_1)^2 = \min \left\{ \int_0^1 \int_\Omega |V(t, x)|^2 \mu(t, dx) dt \mid (\mu, V) \in \text{CE}, \mu(0) = \mu_0, \mu(1) = \mu_1 \right\}.$$

In particular, this identifies  $V$  as (Wasserstein) velocity field tangent to the curve  $\mu$ . Moreover, this interpretation is reflected in the characterization of the metric time derivative in (10). We quote the following result from [2, Theorem 8.3.1].

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

Following Otto's formalism we can associate an Onsager operator  $\mathcal{K}(\mu)$  with the Wasserstein distance. In particular, the velocity fields  $V$  are chosen to be the gradient of a function  $\Xi$  such that  $\dot{\mu} = -\text{div}(\mu \nabla \Xi) =: \mathcal{K}(\mu)\xi$ . 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:

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

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

$$\begin{aligned} \mathcal{E}(U(T)) + \mathcal{J}(U; 0, T) &\leq \mathcal{E}(U(0)), \quad \text{where} \\ \mathcal{J}(U; 0, T) &= \frac{1}{2} \int_0^T \int_{\Omega} \left[ U|V|^2 + \frac{1}{U} |W \nabla(U/W)|^2 \right] dx dt, \quad (16) \\ \text{with } (U, V) &\in \text{CE}. \end{aligned}$$

Here,  $\mathcal{F}(U) = \frac{1}{2} \int_{\Omega} |W \nabla(U/W)|^2 / U dx$  is the *Fisher information*, which gives the square of the metric slope of  $\mathcal{E}$  (see (9)). Note that  $\mathcal{F}(U) = 2 \int_{\Omega} W |\nabla \sqrt{U/W}|^2 dx$ , so that  $\mathcal{F}(U)$  is finite if and only if  $U \in W^{1,1}(\Omega)$ . We refer to [2] for a comprehensive survey on the theory of Wasserstein gradient flows.

We call (16) the *entropy/entropy-dissipation formulation* of the Fokker-Planck equation in (14). 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 [23, 27, 8] 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 (16) only. On  $\{1, \dots, n\}$ ,  $n \in \mathbb{N}$ , we introduce the space

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

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.

$$\dot{u} = Au, \quad \text{where } A_{ij} \geq 0 \text{ for } i \neq j \text{ and } A_{ii} = - \sum_{j:j \neq i} A_{ji}. \quad (17)$$

Many different gradient structures for the Markov chain (17) 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 [26, Section 3.1] and independently in [23, 8]. The geodesic convexity of the relative entropy with respect to the Markov gradient structures was studied in [27] (see also [21]). 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.,

$$\pi_{ij} := A_{ij} w_j = A_{ji} w_i \quad \text{for all } i, j \in \{1, \dots, n\}. \quad (18)$$

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 (17) has the Onsager structure

$$\dot{u} = -K_n(u)DE_n(u), \quad (19)$$

where  $DE_n(u) = (\log \rho_i)_i$  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

$$K_n(u) = \sum_{i < j} \pi_{ij} \Lambda(\rho_i, \rho_j) (e_i - e_j) \otimes (e_i - e_j) \in \mathbb{R}_{\text{sym}, \geq 0}^{n \times n} \quad (20)$$

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

$$\Lambda(a, b) = \frac{a - b}{\log a - \log b} \quad \text{for } a \neq b \quad \text{and} \quad \Lambda(a, a) = a. \quad (21)$$

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

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

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

As before, it follows that an equivalent formulation of (19) is given by

$$E_n(u(T)) + \int_0^T [R_n(u; \dot{u}) + R_n^*(u; -DE_n(u))] dt \leq E_n(u(0)). \quad (22)$$

As in the Wasserstein case, we call (22) the (discrete) entropy/entropy-dissipation formulation of the Markov chain equation in (17). 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

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

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

$$(u, \xi) \in \text{CE}_n \quad \text{if } u \in C_p^1(0, 1; X_n), \quad \dot{u} = K_n(u) \xi. \quad (23)$$

Here,  $C_p^1(0, 1; X_n)$  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

$$d_n(u_0, u_1)^2 = \min \left\{ \int_0^1 \langle \xi, K_n(u) \xi \rangle ds \mid \dot{u} = K_n(u) \xi \right\},$$

where the minimization is over all curves  $u \in C_p^1(0, 1; X_n)$  connecting  $u_0$  and  $u_1$ . We refer to [23] for an extensive and rigorous study of  $X_n$  equipped with this structure.

**Remark 2.4.** The ODE system in (17) is induced by many different gradient systems if the reversibility condition (18) holds. Indeed, for  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  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

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

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 (17).

**3. Discretization scheme.** In this section we discuss the finite-volume discretization of the Fokker-Planck equation in (14) 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. [5, 6, 18, 7, 15, 13, 12]). 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 [11] for a survey on finite-volume methods and to [14] 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  $0 = x_1^n < x_2^n < \dots < x_n^n = 1$ . Given such a partition, we introduce the  $n+1$  midpoints between the vertices  $x_i^n$  via

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

In view of finite-volume schemes in higher dimensions we shall call  $\sigma_i^n$  (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  $\omega_i^n$  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  $h^n \rightarrow 0$  as  $n \rightarrow \infty$ .

As in Subsection 2.4 we consider a potential  $\Phi \in C^1(\Omega) \cap C_b(\Omega)$  with which we associate the equilibrium density  $W(x) = c_\Phi e^{-\Phi(x)}$ . We rewrite the Fokker-Planck equation in (14) using  $W$  and find

$$\dot{U} = (W(U/W)')' \text{ in } \Omega \quad \text{and} \quad (U/W)'(t, x) = 0 \text{ for } x \in \{0, 1\}. \quad (24)$$

Integrating the equation in (24) over the control volume  $\omega_i^n$  gives

$$\frac{d}{dt} \int_{\omega_i^n} U(t, x) dx = f_i - f_{i-1},$$

where  $f_i = (W(U/W)')(\sigma_i^n)$  denotes the flux across the edge  $\sigma_i^n$ . We expect  $u_i(t)$  to approximate  $\int_{\omega_i^n} U(t, x) dx$  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  $n \rightarrow \infty$  by definition. Moreover,  $\kappa_i$  are given transmission coefficients satisfying the consistency condition  $\kappa_i (x_{i+1}^n - x_i^n) / W(\sigma_i) \rightarrow 1$ . Many different choices for  $\kappa_i$  are

possible (see Subsection 3.2 for a discussion). Following [27, Section 5] we use for simplicity the geometric mean of  $W(x_{i+1}^n)$  and  $W(x_i^n)$ , i.e.,

$$\kappa_i = \frac{\sqrt{W(x_{i+1}^n)W(x_i^n)}}{x_{i+1}^n - x_i^n}, \quad \text{for } i = 1, \dots, n-1. \tag{25}$$

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 (24) as

$$\dot{u}_i = \alpha_{i-1}u_{i-1} - (\alpha_i + \beta_{i-1})u_i + \beta_i u_{i+1} \quad \text{for } i = 2, \dots, n-1, \tag{26}$$

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

$$(u_0^n)_i := \int_{\omega_i^n} U_0(x) \, dx, \tag{27}$$

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  $i$ th 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  $u \in C_p^1(0, T; X_n)$ , cf. (23), which is given by

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

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

$$R_n(u, \dot{u}) = \frac{1}{2} \sum_{i=1}^{n-1} \frac{q_i^2}{\kappa_i \Lambda(\rho_{i+1}, \rho_i)} \quad \text{with } q_i = \sum_{k=1}^i \dot{u}_k \tag{28}$$

and the latter as

$$F_n(u) = \frac{1}{2} \sum_{i=1}^{n-1} \frac{\kappa_i (\rho_{i+1} - \rho_i)^2}{\Lambda(\rho_{i+1}, \rho_i)}. \tag{29}$$

Following Definition 2.1, we say that  $u^n \in C([0, T]; X_n)$  is a solution for the gradient flow problem associated with (26) if

$$E_n(u^n(T)) + J_n(u^n; 0, T) \leq E_n(u_0^n). \tag{30}$$

By classical Markov theory, for every  $n \in \mathbb{N}$ , given a partition  $\Pi_n$  and an initial value  $u_0^n \in X_n$ , a solution  $u^n$  of (30) exists and even equality is satisfied in (30). We prove the convergence of the piecewise constant interpolant associated with  $u^n$  to a solution  $U$  of (16) using only the gradient structure of (30). In particular, the main result reads as follows.

**Theorem 3.1.** *Let the potential  $\Phi$  be given as in (15), 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  $h^n \rightarrow 0$  as  $n \rightarrow \infty$ . Moreover, let  $(u^n)_{n \in \mathbb{N}}$  be a sequence of solutions of the entropy/entropy-dissipation formulation (30) with initial values  $(u_0^n)_{n \in \mathbb{N}}$  as in (27). 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  $U^n \xrightarrow{*} U$  in  $\mathcal{M}_{\geq 0}([0, T] \times \bar{\Omega})$  with  $U \in \text{AC}(0, T; \mathcal{X})$  solving*

$$\mathcal{E}(U(T)) + \frac{1}{2} \int_0^T \int_{\Omega} \left[ U|V|^2 + \frac{|U' + U\Phi'|^2}{U} \right] dx dt \leq \mathcal{E}(U_0),$$

where  $(U, V) \in \text{CE}$ , 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  $\kappa_i$  to be given by the geometric mean in the form

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

However, other choices of mean functions  $\theta$  are possible. Indeed, the specific form of  $\theta$  is crucial for the properties of the finite volume discretization (see e.g. [14]). 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  $\theta$  to have the natural property

$$\min\{a, b\} \leq \theta(a, b) \leq \max\{a, b\}. \tag{31}$$

An important example with excellent stability properties is given by the Scharfetter-Gummel scheme (see e.g. [24, 5]), which is widely used in semiconductor device simulation. In our setting the Scharfetter-Gummel scheme reads

$$\theta_{\text{SG}}(a, b) = \Lambda\left(\frac{1}{a}, \frac{1}{b}\right)^{-1},$$

where  $\Lambda$  denotes as before the logarithmic mean. We easily check that this particular choice is also admissible in the sense of (31).

**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 (30). This question has two parts: (i) Do (interpolants of) solutions of (30) with uniformly bounded initial entropies have

beneficial compactness properties that allow us to extract subsequences converging in a suitable topology  $\tau$ ? (ii) Using this topology can we show the lower estimate

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

where  $J_n$  is the dissipation functional in the discrete case and  $\mathcal{J}$  its Wasserstein counterpart in (16)? 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  $\Omega$ .** 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  $\Omega$ . 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

$$U^n(x) = U_i := u_i/h_i^n \quad \text{for } x \in \omega_i^n. \quad (32)$$

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  $W^n \rightarrow W$  in  $L^\infty(\Omega)$ , as  $W$  is uniformly continuous on  $\Omega$ . In particular, this definition allows us to rewrite the discrete relative entropy as

$$E_n(u) = \mathcal{E}_n(U^n) := \int_{\Omega} U^n \log(U^n/W^n) dx. \quad (33)$$

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  $\Lambda(a, b) \leq (a+b)/2$  it is natural to define interpolants by taking the maximum value with respect to adjacent control volumes, i.e.

$$\tilde{U}^n(x) := \max\{U_{i-1}, U_i, U_{i+1}\} \quad \text{for } x \in \omega_i^n, \quad i = 2, \dots, n-1, \quad (34a)$$

while at the boundary cells we set

$$\tilde{U}^n(x) := \begin{cases} \max\{U_1, U_2\} & \text{for } x \in \omega_1^n, \\ \max\{U_{n-1}, U_n\} & \text{for } x \in \omega_n^n. \end{cases} \quad (34b)$$

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

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

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 (28) as the piecewise constant interpolants  $G^n, Q^n$  via

$$G^n(x) = \frac{\rho_{i+1} - \rho_i}{x_{i+1}^n - x_i^n} \quad \text{and} \quad Q^n(x) = q_i \quad \text{for } x \in ]x_i^n, x_{i+1}^n[, \quad i = 1, \dots, n-1. \quad (35)$$

We are now in a position to state the following result which gives a lower estimate for the discrete dissipation functional  $R_n$  in (28) and the discrete Fisher information  $F_n$  in (29) for fixed  $n$ .

**Lemma 4.1.** *For a curve  $u \in C_p^1(0, T; X_n)$  let  $\tilde{U}^n$ ,  $G^n$  and  $Q^n$  be the interpolants defined in (34) and (35), then for all  $t \in [0, T]$ ,*

$$R_n(u(t), \dot{u}(t)) \geq \frac{\gamma_n}{2} \int_{\Omega} \frac{|Q^n(t)|^2}{\tilde{U}^n(t)} dx \quad \text{and} \quad F_n(u(t)) \geq \frac{1}{2} \int_{\Omega} \frac{(\tilde{W}^n)^2}{\tilde{U}^n(t)} |G^n(t)|^2 dx$$

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

*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  $\kappa_i$  in (25) we obtain

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

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

$$\begin{aligned} R_n(u, \dot{u}) &= \frac{1}{4} \frac{q_1^2}{\lambda_1} + \frac{1}{4} \sum_{i=2}^{n-1} \left( \frac{q_{i-1}^2}{\lambda_{i-1}} + \frac{q_i^2}{\lambda_i} \right) + \frac{1}{4} \frac{q_{n-1}^2}{\lambda_{n-1}} \\ &\geq \frac{\gamma_n}{2} \sum_{i=1}^n \int_{\sigma_{i-1}^n}^{\sigma_i^n} \frac{|Q^n|^2}{\tilde{U}^n} dx = \frac{\gamma_n}{2} \int_{\Omega} \frac{|Q^n|^2}{\tilde{U}^n} dx, \end{aligned}$$

where we have used that  $x_i^n - \sigma_{i-1}^n = (x_i^n - x_{i-1}^n) / 2$  for  $i = 2, \dots, n$  and  $\sigma_i^n - x_i^n = (x_{i+1}^n - x_i^n) / 2$  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

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

is satisfied. □

Combining (33) and Lemma 4.1 we arrive at the following proposition which is the starting point for the limit passage  $n \rightarrow \infty$  in the subsequent section.

**Proposition 4.2.** *Let  $u \in C_p^1(0, T, X_n)$  denote a solution of the discrete entropy/entropy-dissipation formulation in (30). Then, the interpolants  $U^n$ ,  $\tilde{U}^n$ ,  $G^n$  and  $Q^n$  in (32), (34) and (35) satisfy the entropy/entropy-dissipation formulation*

$$\mathcal{E}_n(U^n(T)) + \frac{1}{2} \int_0^T \int_{\Omega} \frac{1}{\tilde{U}^n} [ |Q^n|^2 + (\tilde{W}^n)^2 |G^n|^2 ] dx dt \leq \mathcal{E}_n(U_0^n) \quad (36)$$

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

$$\dot{U}^n(t, x_i^n) = \frac{Q^n(t, \sigma_i) - Q^n(t, \sigma_{i-1})}{\sigma_i - \sigma_{i-1}} \quad \text{for } i = 1, \dots, n \quad (37)$$

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  $\mu^n \xrightarrow{*} \mu$  in  $\mathcal{X}$  if

$$\forall \Psi \in C_b(\Omega) : \int_{\Omega} \Psi(x) \mu^n(dx) \longrightarrow \int_{\Omega} \Psi(x) \mu(dx) \quad \text{as } n \rightarrow \infty.$$

Note that since the domain  $\Omega$  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 [3, Theorem 3.1] which is based on the dual formulation of the 1-Wasserstein distance:

$$d_{W_1}(\mu_1, \mu_2) = \sup \left\{ \int_{\Omega} \Psi(x) \mu_1(dx) - \int_{\Omega} \Psi(x) \mu_2(dx) \mid \Psi \in C^{0,1}(\Omega), \|\Psi'\|_{\infty} \leq 1 \right\}.$$

We show that the family of discrete solutions is equicontinuous with respect to  $d_{W_1}$  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 (30) and  $U^n$  the associated piecewise constant interpolant, then, up to subsequences,  $U^n(t) \xrightarrow{*} \mu(t)$  in  $\mathcal{X}$  for all  $t \in [0, T]$  and the limit measure satisfies  $\mu \in C(0, T; \mathcal{X})$ .*

*Proof.* We consider  $\Psi \in C^{0,1}(\Omega)$  satisfying  $|\Psi'(x)| \leq 1$  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  $0 \leq t_0 < t_1 \leq T$

$$\begin{aligned} d_{W_1}(U^n(t_0), U^n(t_1)) &\leq \int_{\Omega} \Psi^n(x) U^n(t_0, x) dx - \int_{\Omega} \Psi^n(x) U^n(t_1, x) dx + \varepsilon_n \\ &= \int_{t_0}^{t_1} \int_{\Omega} \Psi^n(x) \dot{U}^n(t, x) dx dt + \varepsilon_n, \end{aligned}$$

where  $\varepsilon_n = 2\|\Psi - \Psi^n\|_{\infty}$ . Using summation by parts, the discrete continuity equation in (37) then yields

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

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

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

Noting that  $h_{i+1}^n \geq (x_{i+1}^n - x_i^n)/2$  and  $h_i^n \geq (x_{i+1}^n - x_i^n)/2$  we see that the sum is uniformly bounded. Finally, since  $R_n(u, \dot{u})$  is uniformly bounded in  $L^1(0, T)$  we

have shown the equicontinuity of  $t \mapsto U^n(t) \in \mathcal{X}$ . Applying the metric Arzelà-Ascoli theorem [2, Proposition 3.3.1] we obtain a (not relabeled) subsequence with  $U^n(t) \xrightarrow{*} \mu(t)$  in  $\mathcal{X}$  for all  $t \in [0, T]$  and  $\mu \in C(0, T; \mathcal{X})$ .  $\square$

Using the lower semicontinuity properties of the relative entropies (see e.g. [2, Lemma 9.4.3]) and the convergence  $W_n \rightarrow W$  in  $L^\infty(\Omega)$  we obtain the following corollary.

**Corollary 4.4.** *The limit  $\mu$  in Proposition 4.3 satisfies*

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

It is easy to see that the limiting measures  $\mu(t)$  are absolutely continuous with respect to the Lebesgue measure on  $\Omega$  and we denote  $\mu(t, dx) = U(t, x) dx$ . 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 (13) 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.,

$$\forall \Psi \in C(\overline{\Omega_T}) : \quad \lim_{n \rightarrow \infty} \int_0^T \int_\Omega \Psi U^n \, dx \, dt = \int_0^T \int_\Omega \Psi U \, dx \, dt. \quad (38)$$

Also note that the construction of the initial values  $u_0^n$  in (27) yields  $U_0^n \xrightarrow{*} U_0$  in  $\mathcal{X}$ . Moreover,  $U_0^n$  is a recovery sequence for  $\mathcal{E}_n \xrightarrow{\Gamma} \mathcal{E}$ , namely, using Jensen's inequality we obtain

$$\limsup_{n \rightarrow \infty} E_n(u_0^n) = \limsup_{n \rightarrow \infty} \mathcal{E}_n(U_0^n) \leq \mathcal{E}(U_0) < \infty. \quad (39)$$

In the following lemma we show that the interpolant  $\widetilde{U}^n$  defined in (34) 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 (32) and (34), associated with the solution  $u^n$  of the discrete entropy/entropy-dissipation formulation in (30), then  $U^n - \widetilde{U}^n \rightarrow 0$  in  $L^1(\Omega_T)$  for  $n \rightarrow \infty$ .*

*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

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

where the index sets  $I_\pm$  correspond to the cases in which the maximum in (34) is attained at the left and right of  $i$ , respectively. The relative densities  $\rho_i$  satisfy  $U_i = W_i \rho_i$ . Hence, the discrete analog of the product rule  $f_{i+1} g_{i+1} - f_i g_i = f_{i+1}(g_{i+1} - g_i) + g_i(f_{i+1} - f_i)$  leads to the estimate

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

where we used that  $u_i = h_i^n U_i$  and  $h^n = \max_i h_n^i$ . The first term vanishes as  $n \rightarrow \infty$  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,

$$h^n \sum_{i \in I_+} W_{i+1} |\rho_i - \rho_{i+1}| \leq h^n \left( \sum_{i \in I_+} \frac{\kappa_i (\rho_{i+1} - \rho_i)^2}{\Lambda(\rho_{i+1}, \rho_i)} \right)^{1/2} \left( \sum_{i \in I_+} \frac{W_{i+1}^2}{\kappa_i} \Lambda(\rho_{i+1}, \rho_i) \right)^{1/2} \tag{40}$$

The first term is the square root of the discrete Fisher information  $F_n(u)$ . It is uniformly bounded in  $L^2(0, T)$  by assumption. Using the definition of  $\kappa_i$  in (25) 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

$$\begin{aligned} \frac{W_{i+1}^2}{\kappa_i} \Lambda(\rho_{i+1}, \rho_i) &\leq \frac{(x_{i+1}^n - x_i^n) W_{i+1}^2}{2\sqrt{W_{i+1} W_i}} \left( \frac{U_{i+1}}{W_{i+1}} + \frac{U_i}{W_i} \right) \\ &\leq \frac{x_{i+1}^n - x_i^n}{2h_{i+1}^n} \left( \frac{W_{i+1}}{W_i} \right)^{\frac{1}{2}} u_{i+1} + \frac{x_{i+1}^n - x_i^n}{2h_i^n} \left( \frac{W_{i+1}}{W_i} \right)^{\frac{3}{2}} u_i, \end{aligned}$$

where we used the elementary estimate  $\Lambda(a, b) \leq (a+b)/2$  again. Hence, we have shown the strong convergence  $U^n - \tilde{U}^n \rightarrow 0$  in  $L^1(\Omega_T)$ .  $\square$

**Corollary 4.6.** *The interpolant  $\tilde{U}^n$  in (34) associated with the solution  $u$  of the discrete entropy/entropy-dissipation formulation in (30) converges to  $U$  as in (38).*

We are now in position to prove the lower liminf estimate for the dissipation functional. The proof of the following proposition is based on [2, 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 (30) and  $U$  the limit in (38), then*

$$\liminf_{n \rightarrow \infty} \int_0^T [R_n(u^n, \dot{u}^n) + F_n(u^n)] dt \geq \frac{1}{2} \int_0^T \int_{\Omega} \left[ U|V|^2 + \frac{|U' + U\Phi'|^2}{U} \right] dx dt, \tag{41}$$

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

*Proof.* We define the velocity fields  $V^n = Q^n / \tilde{U}^n$  and  $\hat{V}^n = \tilde{W}^n G^n / \tilde{U}^n$ . In particular, due to (36) and the boundedness of the initial entropies we have

$$\sup_n \int_0^T \int_{\Omega} [ |V^n|^2 + |\hat{V}^n|^2 ] \tilde{U}^n dx dt < \infty. \tag{42}$$

We proceed in four steps:

1. *Extraction of converging subsequences.* Here, we only argue for  $V^n$ , the case of  $\hat{V}^n$  being analog. For brevity we denote  $y = (t, x) \in \Omega_T$ . Moreover, in the space  $\mathcal{M}_{\geq 0}(\overline{\Omega_T} \times \mathbb{R})$  we introduce the family of measures given by the push-forward  $\mu^n = (\text{id} \times V^n)_{\#} \tilde{U}^n$ . In particular, for all suitably integrable functions  $g : \overline{\Omega_T} \times \mathbb{R} \rightarrow \mathbb{R}$  we have

$$\int_{\mathbb{R}} \int_{\Omega_T} g(y, v) \mu^n(dy, dv) = \int_{\Omega_T} g(y, V^n(y)) \tilde{U}^n(y) dy.$$

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



$\pi_{\#}^1 \mu^n = \tilde{U}^n$  converges weakly in  $\mathcal{M}_{\geq 0}(\overline{\Omega_T})$ . Moreover, the second marginal satisfies due to (42)

$$\sup_n \int_{\mathbb{R}} |v| (\pi_{\#}^2 \mu^n)(dv) = \sup_n \int_{\Omega_T} |V^n(y)| \tilde{U}^n(y) dy < +\infty.$$

Hence, by [2, Lemma 5.2.2] the sequence  $\mu^n$  is relatively compact in  $\mathcal{M}(\overline{\Omega_T} \times \mathbb{R})$  and we can find a (not relabeled) subsequence and a limit  $\mu \in \mathcal{M}_{\geq 0}(\overline{\Omega_T} \times \mathbb{R})$  such that  $\mu^n \xrightarrow{*} \mu$  in  $\mathcal{M}_{\geq 0}(\overline{\Omega_T} \times \mathbb{R})$ . In particular, since by (42)  $v \mapsto |v|$  is uniformly integrable with respect to  $\mu^n$  and by [2, Proposition 5.1.10], this yields the convergence

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \int_{\Omega_T} \Psi(y)v \mu^n(dy, dv) = \int_{\mathbb{R}} \int_{\Omega_T} \Psi(y)v \mu(dy, dv), \tag{43}$$

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

Let us denote by  $\mu_y \in \mathcal{M}_{\geq 0}(\mathbb{R})$  the disintegration of  $\mu$  with respect to the limit measure  $\mu(dy) = U(y)dy$ , which is  $\mu$ -a.e. uniquely determined, see [2, Section 5.3]. In particular, for every bounded or nonnegative measurable function  $f : \Omega_T \times \mathbb{R} \rightarrow \mathbb{R}$ , the disintegration satisfies

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

Choosing  $f(y, v) = v$  gives the barycentric projection  $V(y) = \int_{\mathbb{R}} v \mu_y(dv)$  of the measure  $\mu$ . Now, using the definition of  $\mu^n$  and the convergence in (43) with  $\Psi \in C(\overline{\Omega_T})$  arbitrary we arrive at

$$\lim_{n \rightarrow \infty} \int_{\Omega_T} \Psi(y)V^n(y)\tilde{U}^n(y) dy = \int_{\Omega_T} \Psi(y)V(y)U(y) dy.$$

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

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{\Omega_T} g(V^n(y)) \tilde{U}^n(y) dy &= \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} \int_{\Omega_T} g(v) \mu^n(dy, dv) \\ &\geq \int_{\mathbb{R}} \int_{\Omega_T} g(v) \mu(dy, dv) \geq \int_{\Omega_T} g\left(\int_{\mathbb{R}} v \mu_y(dv)\right) U(y) dy \\ &= \int_{\Omega_T} g(V(y)) U(y) dy. \end{aligned}$$

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

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

3. *Identification of the limit V.* In this step we verify that the limits  $U$  and  $V$  satisfy the Wasserstein continuity equation, i.e.,  $(U, V) \in \text{CE}$ . To this end, let us consider a test function  $\Psi \in C_c^\infty(]0, T[ \times \mathbb{R})$  and define  $\Psi_i(t) = \Psi(t, x_i^n)$  as well as the piecewise constant interpolant  $\Psi^n$ , i.e., we have  $\Psi^n(t, x) = \Psi_i(t)$  for  $x \in \omega_i^n$ . Using integration by parts and (37) we obtain

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

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

$$\begin{aligned} - \int_0^T \int_\Omega U^n(t, x) \dot{\Psi}(t, x) \, dx \, dt + \varepsilon_n &= \int_0^T \sum_{i=1}^{n-1} Q^n(t, \sigma_i) \{ \Psi_i(t) - \Psi_{i+1}(t) \} \, dt \\ &= - \int_0^T \int_\Omega V^n(t, x) \Psi'(t, x) \tilde{U}^n(t, x) \, dt, \end{aligned} \tag{44}$$

where we have used the boundary conditions  $Q^n(t, \sigma_0) = Q^n(t, \sigma_n) = 0$ . Now, using the convergence of  $U^n$ ,  $\tilde{U}^n$  and  $V^n$  as in (43), we can pass to the limit in (44) to find

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

which is the verification of the continuity equation. Notice that although  $\Omega$  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  $\Psi \in C_c^\infty(]0, T[ \times \overline{\Omega})$  and using (35) we compute

$$\begin{aligned} \int_0^T \int_\Omega \widehat{V}^n \Psi \tilde{U}^n \, dx \, dt &= \int_0^T \sum_{i=1}^{n-1} \frac{\rho_{i+1} - \rho_i}{x_{i+1}^n - x_i^n} \int_{x_i^n}^{x_{i+1}^n} \widetilde{W}^n \Psi \, dx \, dt \\ &= \int_0^T \sum_{i=1}^{n-1} (\rho_{i+1}(t) - \rho_i(t)) \left\{ W(\sigma_i) \Psi(t, \sigma_i) \right. \\ &\quad \left. + \frac{1}{x_{i+1}^n - x_i^n} \int_{x_i^n}^{x_{i+1}^n} [\widetilde{W}^n(x) \Psi(t, x) - W(\sigma_i) \Psi(t, \sigma_i)] \, dx \right\} \, dt. \end{aligned}$$

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

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

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

$$\int_0^T \int_\Omega \widehat{V} \Psi U \, dx \, dt = - \int_0^T \int_\Omega U \frac{(W\Psi)'}{W} \, dx \, dt = - \int_0^T \int_\Omega U(\Psi' - \Phi' \Psi) \, dx \, dt.$$

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

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

initial entropies in (39) we arrive at

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

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

**4.3. Recovery sequences for the dissipation functional.** For the limit passage in the proof of Theorem 3.1 we only need to prove the lower estimate for the dissipation functional. However, to complete the picture we provide the construction of recovery sequences to show that the converse estimate is also satisfied.

**Proposition 4.8** (Recovery sequence). *For all  $\mu \in \text{AC}^2(0, T; \mathcal{X})$  with  $\mu(dx) = U dx$  there exists a sequence  $\widehat{u}^n \in C^1([0, T]; X_n)$  such that  $\widehat{U}^n(t) \rightarrow \mu(t)$  in  $\mathcal{X}$  and*

$$\limsup_{n \rightarrow \infty} J_n(\widehat{u}^n; 0, T) \leq \mathcal{J}(U; 0, T). \tag{45}$$

*Proof.* Let  $U \in \text{AC}^2(0, T; \mathcal{X})$  be such that  $\mathcal{J}(U; 0, T)$  is finite, which yields  $U \in L^1(0, T; W^{1,1}(\Omega))$ , cf. (16). By adapting the mollification argument of Lemma 8.1.9 in [2] we find smooth approximations  $U_\delta$  and  $V_\delta$  satisfying  $U_\delta \in C^1(0, T; C^\infty(\Omega))$  and  $(U_\delta, V_\delta) \in \text{CE}$ . Moreover, we have  $U_\delta > 0$ ,  $U_\delta(t) \rightarrow U(t)$  in  $\mathcal{X}$ , for  $\delta \rightarrow 0$ , as well as the estimates

$$\int_0^T \int_\Omega V_\delta(t, x)^2 U_\delta(t, x) \, dx \, dt \leq \int_0^T \int_\Omega V(t, x)^2 U(t, x) \, dx \, dt$$

and

$$\int_0^T \int_\Omega \frac{W(x)^2}{U_\delta(t, x)} ((U_\delta(t, x)/W(x))')^2 \, dx \, dt \leq \int_0^T \int_\Omega \frac{W(x)^2}{U(t, x)} ((U(t, x)/W(x))')^2 \, dx \, dt$$

(see also [2, Lemma 8.1.10]). Hence, by a simple diagonal argument it suffices to construct a recovery sequence for every smooth  $U \in C^1(0, T; C^\infty(\Omega))$ .

In this case we define  $\widehat{u}_i^n(t) = \widehat{c}_n h_i^n U(t, x_i^n)$ , where  $\widehat{c}_n$  is chosen such that  $\widehat{u}^n \in X_n$ , i.e.  $\sum_{i=1}^n \widehat{u}_i^n = 1$ . Using  $\Lambda(a, b) \geq \sqrt{ab}$  and denoting  $\widehat{\rho}_i^n = \widehat{u}_i^n/w_i^n$  we proceed similar to the proof of Proposition 4.7 to verify that

$$\lim_{n \rightarrow \infty} \int_0^T \sum_{i=1}^{n-1} \frac{\kappa_i^n (\widehat{\rho}_{i+1}^n(t) - \widehat{\rho}_i^n(t))^2}{\Lambda(\widehat{\rho}_{i+1}^n(t), \widehat{\rho}_i^n(t))} \, dt = \int_0^T \int_\Omega \frac{W(x)^2}{U(t, x)} ((U(t, x)/W(x))')^2 \, dx \, dt. \tag{46}$$

Moreover, with  $\widehat{q}_i(t) := c_n \sum_{k=1}^i \widehat{u}_k^n(t)$  for  $i = 1, \dots, n-1$  we have that the associated piecewise constant interpolant defined by  $\widehat{Q}(t, x) = \widehat{q}_i(t)$  for  $x \in ]x_i^n, x_{i+1}^n]$  converges uniformly to  $Q$  with  $Q(t, x) = \int_0^x \dot{U}(t, y) \, dy$  for all  $t \in [0, T]$ . From this, we infer that

$$\lim_{n \rightarrow \infty} \int_0^T \sum_{i=1}^{n-1} \frac{\widehat{q}_i(t)^2}{\kappa_i^n \Lambda(\widehat{\rho}_{i+1}^n(t), \widehat{\rho}_i^n(t))} \, dt = \int_0^T \int_\Omega \frac{Q(t, x)^2}{U(t, x)} \, dx \, dt, \quad \text{with } \dot{U} = Q'.$$

Hence, combined with (46) this gives the upper estimate (45).  $\square$

**Acknowledgments.** 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

- [1] 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** (2011), 791–815.
- [2] 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.
- [3] 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** (2012), 419–454.
- [4] J.-D. Benamou and Y. Brenier, [A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem](#), *Numer. Math.*, **84** (2000), 375–393.
- [5] M. Bessemoulin-Chatard, [A finite volume scheme for convection-diffusion equations with non-linear diffusion derived from the Scharfetter-Gummel scheme](#), *Numer. Math.*, **121** (2012), 637–670.
- [6] 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** (2013), 1–38.
- [7] 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** (2011), 1483–1510.
- [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. Rational Mech. Anal.*, **203** (2012), 969–1008.
- [9] M. Erbar and J. Maas, [Ricci curvature of finite Markov chains via convexity of the entropy](#), *Arch. Rational Mech. Anal.*, **206** (2012), 997–1038.
- [10] M. Erbar and J. Maas, [Gradient flow structures for discrete porous medium equations](#), *Discrete Contin. Dyn. Syst.*, **34** (2014), 1355–1374.
- [11] R. Eymard, T. Gallouët and R. Herbin, The finite volume method, in *Handbook of Numerical Analysis. Vol. VII*, North Holland, Amsterdam, 2000, 713–1022.
- [12] R. Eymard and J.-M. Hérard, eds., *Finite Volumes for Complex Applications V*, ISTE, London, 2008.
- [13] J. Fořt, J. Fürst Jiří, H. R. Herbin and F. Hubert, eds., *Finite Volumes for Complex Applications. VI. Problems & perspectives. Volume 1, 2*, Springer Proceedings in Mathematics, 4, Springer, Heidelberg, 2011.
- [14] J. Fuhrmann, A. Linke and H. Langmach, [A numerical method for mass conservative coupling between fluid flow and solute transport](#), *Appl. Numer. Math.*, **61** (2011), 530–553.
- [15] K. Gärtner, [Charge transport in semiconductors and a finite volume scheme](#), in *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, Springer Proc. Math., 4, Springer, Heidelberg, 2011, 513–522.
- [16] N. Gigli, [On the heat flow on metric measure spaces: Existence, uniqueness and stability](#), *Calc. Var. Partial Differential Equations*, **39** (2010), 101–120.
- [17] N. Gigli and J. Maas, [Gromov-Hausdorff convergence of discrete transportation metrics](#), *SIAM J. Math. Anal.*, **45** (2013), 879–899.
- [18] 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*, ESAIM Proc., 38, EDP Sci., Les Ulis, 2012, 376–386.
- [19] R. Jordan, D. Kinderlehrer and F. Otto, [The variational formulation of the Fokker-Planck equation](#), *SIAM J. Math. Analysis*, **29** (1998), 1–17.
- [20] M. Liero, [Passing from bulk to bulk-surface evolution in the Allen-Cahn equation](#), *NoDEA Nonlinear Differential Equations Appl.*, **20** (2013), 919–942.
- [21] M. Liero and A. Mielke, [Gradient structures and geodesic convexity for reaction-diffusion systems](#), *Phil. Trans. Royal Soc. A*, **371** (2013), 20120346, 28pp.
- [22] M. Liero and U. Stefanelli, [Weighted inertia-dissipation-energy functionals for semilinear equations](#), *Boll. Unione Mat. Ital. (9)*, **6** (2013), 1–27.
- [23] J. Maas, [Gradient flows of the entropy for finite Markov chains](#), *J. Funct. Anal.*, **261** (2011), 2250–2292.
- [24] P. A. Markowich, *The Stationary Semiconductor Device Equations*, Computational Microelectronics, Springer-Verlag, Vienna, 1986.

- [25] D. Matthes and H. Osberger, [Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation](#), *ESAIM Math. Model. Numer. Anal.*, **48** (2014), 697–726.
- [26] A. Mielke, [A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems](#), *Nonlinearity*, **24** (2011), 1329–1346.
- [27] A. Mielke, [Geodesic convexity of the relative entropy in reversible Markov chains](#), *Calc. Var. Part. Diff. Eqns.*, **48** (2013), 1–31.
- [28] A. Mielke, [Thermomechanical modeling of energy-reaction-diffusion systems, including bulk-interface interactions](#), *Discr. Cont. Dynam. Systems Ser. S*, **6** (2013), 479–499.
- [29] A. Mielke, T. Roubíček and U. Stefanelli,  [\$\Gamma\$ -limits and relaxations for rate-independent evolutionary problems](#), *Calc. Var. Part. Diff. Eqns.*, **31** (2008), 387–416.
- [30] A. Mielke and U. Stefanelli, [Weighted energy-dissipation functionals for gradient flows](#), *ESAIM Control Optim. Calc. Var.*, **17** (2011), 52–85.
- [31] L. Onsager, Reciprocal relations in irreversible processes, I+II, *Physical Review*, **37** (1931), 405–426; Part II, **38**, 2265–2267.
- [32] F. Otto, [Dynamics of labyrinthine pattern formation in magnetic fluids: A mean-field theory](#), *Arch. Rational Mech. Anal.*, **141** (1998), 63–103.
- [33] F. Otto, [The geometry of dissipative evolution equations: The porous medium equation](#), *Comm. Partial Differential Equations*, **26** (2001), 101–174.
- [34] E. Sandier and S. Serfaty, [Gamma-convergence of gradient flows with applications to Ginzburg-Landau](#), *Comm. Pure Appl. Math.*, **57** (2004), 1627–1672.

Received March 2014; revised October 2014.

*E-mail address:* [karoline.disser@wias-berlin.de](mailto:karoline.disser@wias-berlin.de)

*E-mail address:* [matthias.liero@wias-berlin.de](mailto:matthias.liero@wias-berlin.de)