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

Preprint ISSN 0946 – 8633

Passing to the limit in a Wasserstein gradient flow: From diffusion to reaction

Steffen Arnrich[1], Alexander Mielke[2,3], Mark A. Peletier[4], Giuseppe Savaré[5], Marco Veneroni[6]

[1] Department of Mathematics and Computer Sciences, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
E-Mail: s.arnrich@tue.nl

[2] Weierstraß-Institut, Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: alexander.mielke@wias-berlin.de

[3] Institut für Mathematik, Humboldt-Universität zu Berlin, Rudower Chaussee 25
12489 Berlin-Adlershof, Germany

[4] Institute for Complex Molecular Systems and Department of Mathematics and Computer Sciences, Technische Universiteit Eindhoven, P.O. Box 513
5600 MB Eindhoven, The Netherlands
E-Mail: m.a.peletier@tue.nl

[5] Università di Pavia, Dipartimento di Matematica “F. Casorati”, Via Ferrate 1
27100 Pavia, Italy
E-Mail: giuseppe.savare@unipv.it

[6] McGill University, Department of Mathematics and Statistics, Burnside Hall
805 Sherbrooke W. Montreal, QC, H3A, Canada
E-Mail: marco.veneroni@mail.mcgill.ca

submitted: February 18, 2011

No. 1593

Berlin 2011

2010 *Mathematics Subject Classification.* 35K67, 35B25, 35B27, 49S99, 35K10, 35K20, 35K57, 60F10, 70F40, 70G75, 37L05.

Key words and phrases. Fokker–Planck equation, transport equation, metric evolution, Gamma convergence.

The research of M. A. Peletier has received funding from the ITN “FIRST” of the Seventh Framework Programme of the European Community (grant agreement number 238702).

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 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

ABSTRACT. We study a singular-limit problem arising in the modelling of chemical reactions. At finite $\varepsilon > 0$, the system is described by a Fokker-Planck convection-diffusion equation with a double-well convection potential. This potential is scaled by $1/\varepsilon$, and in the limit $\varepsilon \to 0$, the solution concentrates onto the two wells, resulting into a limiting system that is a pair of ordinary differential equations for the density at the two wells.

This convergence has been proved in Peletier, Savaré, and Veneroni, *SIAM Journal on Mathematical Analysis*, 42(4):1805–1825, 2010, using the linear structure of the equation. In this paper we re-prove the result by using solely the Wasserstein gradient-flow structure of the system. In particular we make no use of the linearity, nor of the fact that it is a second-order system.

The first key step in this approach is a reformulation of the equation as the minimization of an action functional that captures the propety of being a *curve of maximal slope* in an integrated form. The second important step is a rescaling of space. Using only the Wasserstein gradient-flow structure, we prove that the sequence of rescaled solutions is pre-compact in an appropriate topology. We then prove a Gamma-convergence result for the functional in this topology, and we identify the limiting functional and the differential equation that it represents. A consequence of these results is that solutions of the ε-problem converge to a solution of the limiting problem.

CONTENTS

1. INTRODUCTION

In a seminal paper in 1940, Kramers introduced a model of chemical reactions in which the system is represented by a Brownian particle in a potential energy landscape [Kra40]. In this model the wells of the potential energy correspond to stable states of the system, and a reaction event is the passage of the particle from one well to another. By analyzing the probability of such a reaction event in terms of system parameters, Kramers was able to improve existing formulas for the macroscopically observed reaction rate.

Although Kramers does not state it in these terms, the central result in [Kra40] is a convergence result in the limit of *large activation energy*. In [PSV10] we provided a first rigorous proof of this result in the case of Brownian particles without inertia. The present paper can be considered a sequel to [PSV10], in which we address a question that was left unanswered in [PSV10].

The issue hinges on the fact that the system of [PSV10] is a gradient flow of a free-energy functional with respect to the *Wasserstein* metric. The proof of the main result made no use of this structure, however, and this led us to ask, *Can we prove the same result using the structure of the Wasserstein gradient flow?*

This question is interesting for a number of reasons. The first is that the Wasserstein gradient flow is a *natural* and *physically meaningful* structure for this problem—we explain in Section 7 what we mean by this. It can actually be argued that it is more natural than the linear structure that we used in the proof in [PSV10], and therefore it is also natural to ask whether this structure can be used.

The second reason is that the Wasserstein gradient-flow structure is known to arise in an impressively wide range of models and systems (e.g. [CG04, AGS05, Sav07, BCC08, MMS09, GST09, Gig10, CDF$^+$], just to name a few), and therefore any method that uses only the properties of this structure has the potential of application to a wide range of problems. Consequently, our approach here is to limit our use of information to those properties that follow directly from the gradient-flow structure.

As a third reason, this work fits into a general endeavour to use gradient-flow structures to pass to the limit in nonlinear time-evolving systems (see e.g. [SS04, Ste08, MRS08, MS09, ASZ09, Ser09]). The inherent convexity and lower-semicontinuity properties of this type of formulation provide handles for such limit passages that are similar to the well-known results for elliptic systems—as we show below.

1.1. **Kramers' problem.**

The motion of a Brownian particle in a one-dimensional potential landscape is described by the initial boundary-value problem (often called a Fokker-Planck or Smoluchovski equation [Ris84, p. 8])

$$\partial_t \rho_\varepsilon = \tau_\varepsilon \partial_\xi \Big(\partial_\xi \rho_\varepsilon + \frac{1}{\varepsilon} \rho_\varepsilon \partial_\xi H \Big), \qquad t \geq 0,\ \xi \in \Xi := [-1,1], \tag{1.1a}$$

$$\partial_\xi \rho_\varepsilon + \frac{1}{\varepsilon} \rho_\varepsilon \partial_\xi H = 0, \qquad t \geq 0,\ \xi = \pm 1. \tag{1.1b}$$

The unknown function ρ_ε is a time-dependent measure in $\mathcal{M}(\Xi)$ (the space of finite, nonnegative, Borel measures on the closed interval $\Xi = [-1,1]$), and this equation is to be interpreted in an appropriate weak form.

In this paper we take the potential energy H to be a double-well potential, with wells in $\xi = \pm 1$, and we follow the choice of [PSV10] to truncate the domain at the wells, i.e. we take $\Xi = [-1,1]$ as the spatial domain (see Figure 1). For definiteness we assume that H is smooth, even, maximal at 0 with $H(0) = 1$, and minimal at ± 1 with $H(\pm 1) = 0$. Each of these assumptions can be relaxed, but that is not the purpose of this paper.

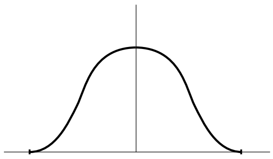

FIGURE 1. A typical function H

In (1.1) two important constants appear. The potential H is scaled by $1/\varepsilon$, which creates the situation of *large activation energy*: the energy barrier separating the two wells is large in the limit $\varepsilon \to 0$. As a consequence, the rate at which a particle passes from one well to the other is exponentially small as $\varepsilon \to 0$; with the coefficient τ_ε, ÃŁwhich is defined in (2.2) below and which tends to infinity as $\varepsilon \to 0$, we adapt the time scale to make the rate of transition asymptotically $O(1)$.

The asymptotically large 'hump' of the potential H/ε causes any solution of (1.1) to become singular in the limit $\varepsilon \to 0$. This is well illustrated by the unique stationary solution of unit mass,

$$\gamma_\varepsilon = Z_\varepsilon^{-1} \mathrm{e}^{-H/\varepsilon} \mathcal{L}^1|_{[-1,1]}, \tag{1.2}$$

where Z_ε is a normalization constant and $\mathcal{L}^1$ is the one-dimensional Lebesgue measure (see Figure 2). Since $H(\xi) > 0$ for all $\xi \neq \pm 1$, the measure γ_ε becomes strongly concentrated at the wells $\xi = \pm 1$ as

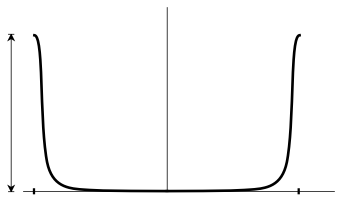

FIGURE 2. The measure γ_ε, illustrated by plotting its Lebesgue density

$\varepsilon \to 0$:

$$\gamma_\varepsilon \overset{*}{\rightharpoonup} \gamma_0 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1. \tag{1.3}$$

In [PSV10] we proved a number of results. The first is that the sequence ρ_ε converges[1], in the sense of measures, to a limit measure ρ_0, whose support is restricted to the two points $\xi = \pm 1$:

$$\rho_\varepsilon \overset{*}{\rightharpoonup} \rho_0 = \frac{1}{2}u_0^- \delta_{-1} + \frac{1}{2}u_0^+ \delta_1.$$

The densities $u_0^\pm : [0,T] \to \mathbb{R}$ of this limit measure ρ_0 satisfy the limit equation

$$\partial_t u_0^- = k(u_0^+ - u_0^-) \tag{1.4a}$$

$$\partial_t u_0^+ = k(u_0^- - u_0^+). \tag{1.4b}$$

where the rate constant k is given in terms of the potential function H by

$$k = \frac{1}{\pi}\sqrt{|H''(0)|H''(1)}. \tag{1.5}$$

[1]The result of [PSV10] uses a slightly different definition of τ_ε, which is asympotically equivalent to the one this paper, (2.2).

This limit system corresponds to the natural modelling of the monomolecular reaction $A \leftrightharpoons B$ at the continuum level.

A second result states a stronger form of convergence, and also highlights the role of the *density* u_ε of the measure ρ_ε with respect to γ_ε, i.e.

$$u_\varepsilon = \frac{\mathrm{d}\rho_\varepsilon}{\mathrm{d}\gamma_\varepsilon}, \tag{1.6}$$

which satisfies the dual equation

$$\partial_t u_\varepsilon = \tau_\varepsilon\Big(\partial_{\xi\xi} u_\varepsilon - \frac{1}{\varepsilon}\partial_\xi u_\varepsilon \partial_\xi H\Big). \tag{1.7}$$

Figure 3 illustrates the relationship between ρ_ε and u_ε. As it turns out, u_ε is much better behaved than ρ_ε

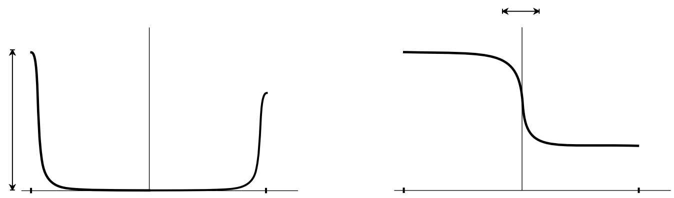

FIGURE 3. A comparison of ρ_ε and the density $u_\varepsilon = \mathrm{d}\rho_\varepsilon/\mathrm{d}\gamma_\varepsilon$.

in the limit $\varepsilon \to 0$: if the initial datum for u_ε is bounded above and below, then the same holds for u_ε by the comparison principle, since constants are solutions of (1.7). In addition, u_ε becomes locally constant away from $\xi = 0$ (see part 3 of Theorems 3.1 and 3.2 below). This is reflected in a stronger form of convergence for u_ε, proved in [PSV10], which implies in particular that nonlinear functions of u_ε also converge.

The aim of this paper is to derive similar convergence statements by different methods, specifically, by using only the structure of the *Wasserstein gradient flow*. Before describing this structure for the specific case of (1.1), we first recall the general structure of a gradient flow in a smooth and finite-dimensional setting.

1.2. **Gradient flows in a smooth Riemannian setting.** Let us consider a smooth d-dimensional Riemannian manifold $\mathcal{Z}$, a C^1 energy functional $\mathcal{E} : \mathcal{Z} \to \mathbb{R}$, and a quadratic dissipation potential ψ induced by the Riemannian metric on $\mathcal{Z}$. In local coordinates, we can identify $\mathcal{Z}$ (and the tangent space $\mathrm{T}_z(\mathcal{Z})$ at each point $z \in \mathcal{Z}$) with $\mathbb{R}^d$ endowed with a smooth Riemannian tensor $G(z) : \mathbb{R}^d(= \mathrm{T}_z(\mathcal{Z})) \to \mathbb{R}^d(= \mathrm{T}^*_z(\mathcal{Z}))$ in the form $\psi(\dot z; z) = \frac{1}{2}\langle G(z)\dot z, \dot z\rangle$.

The gradient flow of $\mathcal{E}$ in $\mathcal{Z}$ is then given in the form

$$\dot z(t) = v(t) \in \mathrm{T}_{z(t)}(\mathcal{Z}), \quad \text{where} \quad v(t) = -\nabla_G \mathcal{E}(z(t)) \quad \text{or} \quad G(z(t))v(t) = -\mathrm{D}\mathcal{E}(z(t)). \tag{1.8}$$

Here and elsewhere in this paper we use overdots for time differentiation and D for the Fréchèt derivative of a function (an element of $\mathrm{T}^*_z(\mathcal{Z})$ in the Riemannian setting). The gradient $\nabla_G\mathcal{E}$ is defined as usual via the metric as $z \mapsto G(z)^{-1}\mathrm{D}\mathcal{E}(z)$. It will sometimes be easier to use the dual dissipation potential ψ^* given via the Legrendre transform with respect to $\dot z$, namely $\psi^*(\eta; z) = \frac{1}{2}\langle \eta, G(z)^{-1}\eta\rangle$. Then the gradient flow (1.8) takes the form

$$\dot z = \mathrm{D}\psi^*(-\mathrm{D}\mathcal{E}(z); z). \tag{1.9}$$

Here and below the derivatives $\mathrm{D}\psi$ and $\mathrm{D}\psi^*$ are only taken with respect to the first variable.

Solutions of (1.9) in a time interval (a,b) can be characterized as minimizers of the action functional

$$\mathcal{A}(z;a,b) := \int_a^b \psi(\dot z + \nabla_G \mathcal{E}(z); z)\,\mathrm{d}t = \frac{1}{2}\int_a^b \langle G(z)(\dot z + \nabla_G\mathcal{E}(z)), \dot z + \nabla_G \mathcal{E}(z)\rangle\,\mathrm{d}t, \tag{1.10}$$

defined on C^1 curves with values in $\mathcal{Z}$. Expanding the integrand and observing that

$$\langle G(z)\dot z, \nabla_G\mathcal{E}(z)\rangle = \langle \dot z, \mathrm{D}\mathcal{E}(z)\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(z),$$

we see that $\mathcal{A}$ has the structure

$$\begin{aligned}\mathcal{A}(z;a,b) &= \mathcal{E}(z(b)) - \mathcal{E}(z(a)) + \mathcal{J}(z;a,b),\\ \mathcal{J}(z;a,b) &:= \int_a^b \Big[\psi\big(\dot z(t); z(t)\big) + \psi^*\big(-\mathrm{D}\mathcal{E}(z(t)); z(t)\big)\Big]\,\mathrm{d}t.\end{aligned} \tag{1.11}$$

Note that for every curve z we have $\mathcal{A}(z;a,b)\geq 0$, while $\mathcal{A}(z;a,b)=0$ if and only if z satisfies (1.8).

1.3. **Gradient flows in a metric setting.** The functionals $\mathcal{J}$ and $\mathcal{A}$ can be generalized to infinite-dimensional and non-smooth settings given by a space $\mathcal{Z}$ with a lower semicontinuous (pseudo-, i.e. possibly taking the value $+\infty$) distance $\mathsf{d} : \mathcal{Z}\times\mathcal{Z}\to[0,+\infty]$. In such a space both tangent spaces and derivatives might not exist. Instead one can turn to two metric concepts, the metric slope $|\partial\mathcal{E}|$ of the functional $\mathcal{E}$ and the metric velocity $|\dot z|$ of a curve. The metric slope generalizes $(2\psi^*(-\mathrm{D}\mathcal{E}(z);z))^{-1/2}$ and is defined by

$$|\partial\mathcal{E}|(z) := \limsup_{w\to z}\frac{(\mathcal{E}(z)-\mathcal{E}(w))_+}{\mathsf{d}(z,w)}. \tag{1.12}$$

Instead of defining a dissipation potential ψ on the tangent space of an arbitrary point of $\mathcal{Z}$, one considers the class $AC(a,b;(\mathcal{Z},\mathsf{d}))$ of absolutely continuous curves (with respect to the distance d) and their metric velocity

$$|\dot z|(t) := \lim_{h\to 0}\frac{\mathsf{d}(z(t),z(t+h))}{|h|} \quad \text{if } z\in AC(a,b;(\mathcal{Z},\mathsf{d})), \tag{1.13}$$

which exists for a.e. $t\in(a,b)$ [AGS05, Th. 2.1.2].

Using these concepts, the natural generalization of $\mathcal{J}$ in (1.11) is

$$\mathcal{J}(z;a,b) := \int_0^T \Big[\frac{1}{2}|\dot z|^2(t) + \frac{1}{2}|\partial\mathcal{E}|^2(z(t))\Big]\,\mathrm{d}t \quad \text{if } z\in AC(a,b;(\mathcal{Z},\mathsf{d})), \tag{1.14}$$

trivially extended by $+\infty$ if z is not absolutely continuous. Assuming that the slope is a *strong upper gradient* for $\mathcal{E}$ [AGS05, Ch. 2], it is not difficult to prove that

$$\mathcal{J}(z;a,b) \geq \mathcal{E}(z(a)) - \mathcal{E}(z(b)) \quad \text{for every curve } z\in C([a,b];\mathcal{Z}) \text{ with } \mathcal{E}(z(a))<+\infty. \tag{1.15}$$

Comparing with the classical case outlined in Section 1.2 we deduce the following common structure:

Definition 1.1. *Let $\mathcal{Z}$ be a topological space, $\mathcal{E} : \mathcal{Z}\to(-\infty,+\infty]$ be a functional, and let $\mathcal{J}(\,\cdot\,;a,b)$ be a nonnegative (extended) real functional defined on $C([0,T];\mathcal{Z})$ for all $0\leq a<b\leq T$, and satisfying*

$$\mathcal{E}(z(b)) + \mathcal{J}(z;a,b) \geq \mathcal{E}(z(a)) \quad \text{for every } z\in C([a,b];\mathcal{Z}). \tag{1.16}$$

Writing

$$\mathcal{A}(z) := \mathcal{E}(z(T)) - \mathcal{E}(z(0)) + \mathcal{J}(z;0,T),$$

we define a curve $z\in C([0,T];\mathcal{Z})$ to be a solution of the gradient flow system $(\mathcal{Z},\mathcal{E},\mathcal{J})$ if $\mathcal{E}(z(0))<\infty$ and $\mathcal{A}(z)=0$.

This formulation of a gradient flow, in terms of the functional $\mathcal{A}$, will be the basis for the rest of this paper. It clearly contains the classical case of a gradient system $(\mathcal{Z}, \mathcal{E}, \mathrm{d})$, for which $\mathcal{J}$ can be defined via (1.12)–(1.14), and the metric-space outlined above, and it is sufficiently general to contain also the structure of the limiting problem (see Section 1.7).

1.4. **A first gradient-flow structure for** (1.1)**: the Hilbertian approach of** [PSV10]**.** We now turn to the specific case of this paper, equation (1.1). It is well known [Bre73] that equation (1.1) in the density formulation (1.7) is the gradient flow of the Dirichlet form

$$\mathcal{E}_\varepsilon^{\mathsf{lin}}(u) := \frac{\tau_\varepsilon}{2} \int_{-1}^{1} |\partial_\xi u|^2 \, \mathrm{d}\gamma_\varepsilon, \tag{1.17}$$

in the weighted Hilbert space $\mathcal{Z}_\varepsilon^{\mathsf{lin}} = L^2(\Xi; \gamma_\varepsilon)$. In this approach the quadratic dissipation potential (which is also the squared metric velocity) of a curve u is

$$\psi_\varepsilon^{\mathsf{lin}}(\dot{u}; u) = \frac{1}{2} \int_{-1}^{1} \dot{u}^2 \, \mathrm{d}\gamma_\varepsilon = \frac{1}{2} \|\dot{u}\|^2_{L^2(\Xi;\gamma_\varepsilon)}, \tag{1.18}$$

and does not depend on u, so that the resulting space has a flat Hilbertian geometry.

The limit ODE (1.4) has a similar linear structure, given by

$$\begin{aligned} \mathcal{Z}_0^{\mathsf{lin}} \;&=\; L^2(\{-1,1\}; \gamma_0) \;=\; \big\{ \, (u_0^+, u_0^-) \in \mathbb{R}^2 : u_0^\pm = u_0(\pm 1) \, \big\}, \\ \mathcal{E}_0^{\mathsf{lin}}(u_0) \;&=\; \frac{k}{2}(u_0^+ - u_0^-)^2, \quad \psi_0^{\mathsf{lin}}(\dot{u}_0; u_0) \;=\; \frac{1}{4}|\dot{u}_0^+|^2 + \frac{1}{4}|\dot{u}_0^-|^2 \;=\; \frac{1}{2}\int_{\{-1,1\}} |\dot{u}_0|^2 \, \mathrm{d}\gamma_0. \end{aligned} \tag{1.19}$$

The rigorous transition from (1.18) to (1.19) is established in [PSV10] in a more general setting where diffusion in physical space is allowed as well. The analysis in [PSV10] depends in a crucial way on the linearity of the problem.

1.5. **An alternative gradient-flow structure for** (1.1)**: the Wasserstein approach of** [JKO97]**.** As was discovered in the seminal work by Otto and co-workers [JKO97, Ott01], equation (1.1) has another relevant gradient structure. It relies on the interpretation of ρ as a mass distribution which is transported such as to reduce the free energy.

In order to describe this point of view, we introduce

$$\mathcal{Z}^{\mathsf{meas}} := \mathcal{M}(\Xi), \qquad \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho) := \int_\Xi u \log u \, \mathrm{d}\gamma_\varepsilon - \rho(\Xi) \log \rho(\Xi), \quad \text{where } u := \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_\varepsilon}, \tag{1.20}$$

with the convention that $\mathcal{E}^{\mathsf{free}}(\rho) = +\infty$ if ρ is not absolutely continuous with respect to γ_ε. The space $\mathcal{Z}^{\mathsf{meas}}$ is endowed with the usual weak-$*$ convergence of measures (i.e. convergence in duality with continuous functions) and can be metrized by the L^2-*Wasserstein distance* d_W.

This distance d_W admits two nice characterizations: the first one involves optimal transport (see e.g. [Vil03, AGS05]), while the second one is related to the dynamical interpretation discovered by Benamou and Brenier [BB00] and is well adapted to the gradient-flow setting.

In the latter point of view, we introduce the class $\mathsf{CE}(a, b; \Xi)$ (**C**ontinuity **E**quation) given by couples $\rho \in C([a,b]; \mathcal{Z}^{\mathsf{meas}})$ and $\nu \in \mathcal{M}((a,b) \times \Xi; \mathbb{R})$ such that

$$\partial_t \rho + \partial_\xi \nu = 0 \;\text{ in the sense of distributions in } \mathcal{D}'((a,b) \times \mathbb{R}).$$

Here we trivially extend ρ by zero outside of Ξ. Often $\nu = \rho v$ for some Borel velocity field $v : (a,b)\times\Xi \to \mathbb{R}$, in which case the conditions above reduce to

$$\begin{aligned}&\int_a^b\int_\Xi |v(t,\xi)|\,\rho(t,\mathrm{d}\xi)\,\mathrm{d}t < +\infty \quad \text{and}\\ &\partial_t\rho + \partial_\xi(\rho v) = 0 \ \text{ in the sense of distributions in } \mathcal{D}'((a,b)\times\mathbb{R}).\end{aligned} \tag{1.21}$$

For those couples $(\rho,\nu) \in \mathsf{CE}(a,b;\Xi)$ such that there exists such a velocity field v with $\nu = \rho v$, the distance d_W can be defined in terms of v, by

$$\mathsf{d}_W^2(\rho_0,\rho_1) = \min\Big\{\int_0^1\int_{\bar D} |\boldsymbol{v}(t,\boldsymbol{x})|^2\,\rho(t,\mathrm{d}\boldsymbol{x})\,\mathrm{d}t : (\rho,\rho\boldsymbol{v}) \in \mathsf{CE}(0,1;\bar D), \rho(0)=\rho_0, \rho(1)=\rho_1\Big\}, \tag{1.22}$$

which illustrates how we can interpret v as the 'Wasserstein velocity' of the curve ρ. Note how finiteness of d_W requires that $\nu \ll \rho$ and $\mathrm{d}\nu/\mathrm{d}\rho \in L^2(\rho)$, implying that $\mathsf{CE}(a,b;\Xi)$ is a larger space than $AC([a,b];\mathsf{d}_W)$; indeed, our choice to work with $\mathsf{CE}(a,b;\Xi)$ stems from the fact that in the limit $\varepsilon = 0$ the objects will still be elements of $\mathsf{CE}(a,b;\Xi)$, but no longer of $AC([a,b];\mathsf{d}_W)$.

Recalling (1.13), it is natural to introduce the dissipation potential

$$\psi_\varepsilon^{\mathsf{Wass}}(v;\rho) := \frac{1}{2\tau_\varepsilon}\int_\Xi v(\xi)^2\,\rho(\mathrm{d}\xi), \quad \text{for } \rho\in\mathcal{M}(\Xi),\ v\in L^2(\Xi;\rho). \tag{1.23}$$

This expression suggests to interpret $L^2(\Xi;\rho(t,\cdot))$ as the 'Wasserstein tangent space' at the measure $\rho(t,\cdot)$, and in [AGS05, Ch. 8] this suggestion is made rigorous.

The corresponding (squared) slope of $\mathcal{E}_\varepsilon^{\mathsf{free}}$ [AGS05, §10.4.4] defined by (1.12) is the *Fisher information*

$$|\partial\mathcal{E}^{\mathsf{free}}(\rho)|^2 := \int_\Xi \Big|\frac{\partial_\xi u}{u}\Big|^2\,\mathrm{d}\rho = 4\int_\Xi \big|\partial_\xi\sqrt{u}\big|^2\,\mathrm{d}\gamma_\varepsilon, \quad \text{if } u = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_\varepsilon} \text{ with } \sqrt{u}\in W^{1,2}(-1,1).$$

This corresponds to the choice of the dual dissipation potential

$$(\psi_\varepsilon^{\mathsf{Wass}})^*(\eta;\rho) := \frac{\tau_\varepsilon}{2}\int_\Xi |\eta(\xi)|^2\,\rho(\mathrm{d}\xi)$$

and of the 'Wasserstein gradient' $\nabla_W\mathcal{E}_\varepsilon^{\mathsf{free}}$ of the entropy given (at least formally) by

$$\nabla_W\mathcal{E}_\varepsilon^{\mathsf{free}}(\rho) := \frac{\partial_\xi u}{u} = \partial_\xi\log u, \quad u = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_\varepsilon}.$$

This construction is equivalent to (1.1): in fact, at least for smooth densities,

$$\partial_\xi\rho + \frac{1}{\varepsilon}\rho\partial_\xi H = \rho\,\partial_\xi\log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma_\varepsilon}\Big)$$

so that (1.1) has the gradient flow structure (1.8) in the Wasserstein sense:

$$\partial_t\rho + \partial_\xi(\rho\,v) = 0, \qquad v = -\tau_\varepsilon\nabla_W\mathcal{E}_\varepsilon^{\mathsf{free}}(\rho),$$

Motivated by these remarks, we introduce the functional $\mathcal{J}_\varepsilon^{\mathsf{Wass}}$,

$$\mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho;a,b) := \int_a^b\Big(\int_\Xi\frac{1}{2\tau_\varepsilon}v^2\,\rho(t,\mathrm{d}\xi) + 2\tau_\varepsilon\int_\Xi\big|\partial_\xi\sqrt{u}\big|^2\,\mathrm{d}\gamma_\varepsilon\Big)\,\mathrm{d}t,$$
$$\text{if } (\rho,\rho v)\in\mathsf{CE}(a,b;\Xi) \text{ and } \rho = u\gamma_\varepsilon, \tag{1.24}$$

and the corresponding Wasserstein action functional

$$\mathcal{A}_\varepsilon^{\mathsf{Wass}}(\rho;a,b) := \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho(a)) - \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho(b)) + \mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho;a,b), \tag{1.25}$$

which satisfies the admissibility condition (1.16). In analogy to Definition 1.1, a Wasserstein solution ρ of (1.1) in the time interval $[0,T]$ is a curve in $\mathcal{M}(\Xi)$ with $\mathcal{E}_\varepsilon^{\mathsf{free}}(\rho(0)) < +\infty$ and $\mathcal{A}_\varepsilon(\rho;0,T) = 0$.

1.6. **Our main results.** In this work we prove various results on the connection between the Wasserstein gradient structure $(\mathcal{Z}^{\mathsf{meas}}, \mathcal{E}_\varepsilon^{\mathsf{free}}, \mathcal{J}_\varepsilon^{\mathsf{Wass}})$ and a gradient structure $(\mathcal{Z}_0^{\mathsf{meas}}, \mathcal{E}_0^{\mathsf{free}}, \mathcal{J}_0)$ for the limit system (1.4). As described above, the motivating question is whether we may pass to the limit in the gradient-flow equation $\mathcal{A}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon) = 0$. This question falls apart into two sub-questions:

1 *Compactness:* Do solutions of $\mathcal{A}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon;0,T) = 0$ with uniformly bounded initial entropy $\mathcal{E}_\varepsilon^{\mathsf{free}}(\rho_\varepsilon(0))$ have beneficial compactness properties, allowing us to extract a subsequence that converges in a suitable topology, say σ?

2 *Liminf inequality:* Is there a limit functional $\mathcal{J}_0$ such that

$$\rho_\varepsilon \xrightarrow{\sigma} \rho_0 \quad \Rightarrow \quad \liminf_{\varepsilon\to 0} \mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon;a,b) \geq \mathcal{J}_0(\rho_0;a,b)?$$

And if so, does it satisfy the admissibility condition (1.16), i.e.

$$\mathcal{E}_0^{\mathsf{free}}(\rho(a)) - \mathcal{E}_0^{\mathsf{free}}(\rho(b)) + \mathcal{J}_0(\rho;a,b) \geq 0$$

for every $0 \leq a < b \leq T$ and $\rho \in C([a,b];\mathcal{Z}_0^{\mathsf{meas}})$ with $\mathcal{E}_0^{\mathsf{free}}(\rho(a)) < +\infty$?

Our answer to these questions is indeed affirmative. Question 1 is answered by Theorem 3.2, which establishes that any sequence ρ_ε such that $\mathcal{E}_\varepsilon^{\mathsf{free}}(\rho_\varepsilon(0))$ and $\mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon;0,T)$ are bounded, is compact in several topologies. These boundedness assumptions are natural, and only use information associated with the gradient-flow structure.

Question 2 is addressed by Theorems 4.1 and 6.1, which characterize the limit of the functionals $\mathcal{J}_\varepsilon^{\mathsf{Wass}}(\cdot;a,b)$ in terms of Gamma-convergence. If we denote by σ the topology mentioned above, then this convergence is characterized by the existence of functionals $\mathcal{J}_0$ and $\mathcal{E}_0^{\mathsf{free}}$ satisfying the two properties

1 Lower bound: for each family of curves $\rho_\varepsilon \xrightarrow{\sigma} \rho_0$ with $\sup_\varepsilon \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho_\varepsilon(a)) < +\infty$, we have

$$\mathcal{J}_0(\rho_0;a,b) \leq \liminf_{\varepsilon\to 0} \mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon;a,b) \quad \text{and} \quad \mathcal{E}_0^{\mathsf{free}}(\rho_0(b)) \leq \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho_\varepsilon(b)).$$

2 Recovery sequence: for each $\rho_0 \in C([a,b];\mathcal{Z}_0^{\mathsf{meas}})$ with $\mathcal{E}_0^{\mathsf{free}}(\rho_0(a)) < +\infty,\ \mathcal{J}_0(\rho_0;a,b) < +\infty$ there exists a sequence $\rho_\varepsilon \in C([a,b];\mathcal{Z}^{\mathsf{meas}})$ such that $\rho_\varepsilon \xrightarrow{\sigma} \rho_0$ and

$$\mathcal{J}_0(\rho_0;a,b) = \lim_{\varepsilon\to 0} \mathcal{J}_\varepsilon^{\mathsf{Wass}}(\rho_\varepsilon;a,b), \quad \mathcal{E}_0^{\mathsf{free}}(\rho_0(b)) = \lim_{\varepsilon\to 0} \mathcal{E}_\varepsilon^{\mathsf{free}}(\rho_\varepsilon(b)).$$

The limit structure $(\mathcal{Z}_0^{\mathsf{meas}}, \mathcal{E}_0^{\mathsf{free}}, \mathcal{J}_0)$ consists of measures ρ that are absolutely continuous with respect to γ_0 and thus supported in $\{-1,1\}$:

$$\rho = \frac{1}{2}u^-\delta_{-1} + \frac{1}{2}u^+\delta_1 \qquad \text{for some } u^\pm \geq 0.$$

The space $\mathcal{Z}_0^{\mathsf{meas}}$ and the energy $\mathcal{E}_0^{\mathsf{free}}$ are natural limits of the corresponding objects as $\varepsilon \to 0$:

$$
\begin{aligned}
&\mathcal{Z}_0^{\mathsf{meas}} = \Big\{\rho \in \mathcal{M}(\Xi) : \operatorname{supp}(\rho) \subset \{-1,1\}\Big\} \subset \mathcal{Z}^{\mathsf{meas}},\\
&\mathcal{E}_0^{\mathsf{free}}(\rho) = \int_{\{-1,1\}} \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_0} \log\Big(\frac{\mathrm{d}\rho}{\mathrm{d}\gamma_0}\Big)\,\mathrm{d}\gamma_0 - m\log m = \frac{1}{2}\Big(u^+\log u^+ + u^-\log u^-\Big) - m\log m\\
&\quad\text{where}\quad u^\pm = \frac{\mathrm{d}\rho}{\mathrm{d}\gamma_0}(\pm 1), \quad m = \rho(\Xi) = \frac{1}{2}(u^+ + u^-).
\end{aligned}
\tag{1.26}
$$

This limit energy $\mathcal{E}_0^{\mathsf{free}}$ is the Gamma-limit of $\mathcal{E}_\varepsilon^{\mathsf{free}}$ [ASZ09]. However, the limit functional $\mathcal{J}_0(\cdot; a, b)$, does not have the same duality structure as (1.11), and we discuss this next.

1.7. **The structure of $\mathcal{J}_0$.** In fact, since the limit problem is characterized by measures $\rho(t)$ concentrated at $\xi = \pm 1$, no effective mass transport is possible between $\xi = -1$ and $\xi = 1$. Assume for instance the case when ρ is sufficiently smooth, i.e. $\rho(t) = \frac{1}{2}\sum u^\pm(t)\delta_{\pm 1} = u\gamma_0$ for a couple $u^\pm \in C^1((a,b);\mathbb{R})$. Then the distributional time derivative of ρ is $\partial_t\rho = \frac{1}{2}\sum \dot u^\pm\delta_{\pm 1}$ and any signed measure ν supported in $\Xi \times [a,b]$ and solving the continuity equation

$$
\partial_t\rho + \partial_\xi\nu = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}\times(a,b))
\tag{1.27}
$$

cannot be absolutely continuous with respect to ρ and therefore cannot admit the decomposition $\nu = \rho v$ for some $v \in L^2(-1,1;\rho)$ (except for the trivial case $\dot u^\pm \equiv 0$).

Recalling that the total mass $m = \frac{1}{2}(u^- + u^+)$ is conserved and therefore $\dot u^- = -\dot u^+$, equation (1.27) has the unique solution

$$
\nu = w\mathcal{L}^2|_{(-1,1)\times(a,b)}, \quad w(\xi,t) = \frac{1}{2}\dot u^+(t) \quad \text{for } \xi \in (-1,1),\ t \in (a,b),
\tag{1.28}
$$

trivially extended to 0 outside $[-1,1]$.

As we show below, $\mathcal{J}_0(\rho; a, b)$ has the form

$$
\begin{aligned}
&\mathcal{J}_0(\rho; a,b) := \int_0^T M(w(t), u^\pm(t))\,\mathrm{d}t\\
&\qquad \text{if } \rho(t) = \frac{1}{2}\sum u^\pm(t)\delta_{\pm 1}, \quad u^\pm \in AC(a,b;\mathbb{R}), \quad \text{with } w = \frac{1}{2}\dot u^- = -\frac{1}{2}\dot u^+,
\end{aligned}
\tag{1.29}
$$

where the function $M : \mathbb{R}\times[0,\infty)^2 \to [0,+\infty]$ is given by

$$
M(w,u^\pm) := \inf\left\{\int_{-\kappa}^{\kappa} \frac{w^2}{2u(s)} + \frac{u'(s)^2}{2u(s)}\,\mathrm{d}s \ :\ u \in H^1(-\kappa,\kappa),\ u(\pm\kappa) = u^\pm\right\}, \quad \kappa := \frac{1}{k}.
\tag{1.30}
$$

This functional $\mathcal{J}_0$ satisfies the admissibility criterion (1.16). Indeed, along any admissible curve $\rho(t)$,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_0^{\mathsf{free}}(\rho(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\sum_\pm u^\pm(t)\log u^\pm(t) = \frac{1}{2}\sum_\pm(\log u^\pm(t)+1)\dot u^\pm(t)\\
&= \frac{1}{2}[\log u^+(t) - \log u^-(t)]\,\dot u^+(t),
\end{aligned}
$$

so that the admissibility condition (1.16) is equivalent to

$$
(\log u^+ - \log u^-)w \le M(w,u^\pm) \quad \text{for every } w \in \mathbb{R},\ u^\pm > 0.
\tag{1.31}
$$

In Theorem 5.2 we prove this inequality, implying that the limiting action $\mathcal{A}_0$,

$$\mathcal{A}_0(\rho) := \mathcal{E}_0^{\text{free}}(\rho(T)) - \mathcal{E}_0^{\text{free}}(\rho(0)) + \mathcal{J}_0(\rho), \tag{1.32}$$

satisfies $\mathcal{A}_0(\rho) \geq 0$ for all ρ.

It is now natural to ask which curves ρ satisfy the equation $\mathcal{A}_0(\rho; a, b) = 0$. This equation implies equality in (1.31), which suggests defining the 'contact set' [MRS09]

$$\mathcal{C} := \{\ (u^\pm, w) \mid M(w, u^\pm) + (\log u^+ - \log u^-)w = 0\ \}.$$

A consequence of a second inequality proved in Theorem 5.2 is

$$(u^\pm, w) \in \mathcal{C} \quad \Longleftrightarrow \quad w = \frac{k}{2}(u^+ - u^-).$$

This implies that any ρ satisfying $\mathcal{A}_0(\rho; a, b) = 0$ also solves the limiting equation (1.4).

1.8. **Recovering a gradient flow.** Finally, one might ask whether it is possible to find a 'true' gradient structure, i.e. an alternative functional $\mathcal{A}_0$ that does have the dual structure as in (1.11) or (1.14). For this we need to find a dissipation potential $\psi_0(w; u^\pm)$ such that the associated contact set is equal to $\mathcal{C}$, i.e. such that

$$\mathcal{C} = \{\ (w, u^\pm)\ :\ \psi_0(w; u^\pm) + \psi_0^*(-\mathrm{D}\mathcal{E}_0(u^\pm); u^\pm) + \langle w, \mathrm{D}\mathcal{E}_0(u^\pm)\rangle = 0\ \}.$$

Using the two-sided estimate of Theorem 5.2 for M we find that a natural choice for ψ_0 is

$$\psi_0(w; u^\pm) = \frac{2}{k}\frac{\log u^+ - \log u^-}{u^+ - u^-}w^2$$

which gives the desired result (1.4).

1.9. **Structure of the paper.** In Section 2 we introduce a rescaling of space that desingularizes one of the terms in $\mathcal{J}_\varepsilon$. This rescaling allows us to prove, in Section 3, the compactness of a sequence ρ_ε with bounded initial energy $\mathcal{E}_\varepsilon(\rho(0))$ and bounded $\mathcal{J}_\varepsilon(\rho_\varepsilon)$ in a number of topologies. Sections 4 and 6 give the two parts of the Gamma-convergence result, the lower bound and the recovery sequence. Before constructing the recovery sequence we investigate in Section 5 the function M in some detail. These are the central mathematical results of the paper.

In Section 7 we place the results of this paper in the context of large-deviation principles for systems of Brownian particles, and comment on the various connections. In Section 8 we discuss various aspects of the results and their proof and comment on possible generalizations. Finally, in Section 9 we draw parallels between this work and an independent study of the same question by Herrmann and Niethammer [HN11].

Summary of notation

$\overset{*}{\rightharpoonup}$	weak convergence in duality with continuous functions	
$\mathsf{CE}(\cdot,\cdot;\cdot)$	pairs (ρ,ν) satisfying the continuity equation	(1.21)
$\mathsf{d}_W^2(\cdot,\cdot)$	Wasserstein distance of order 2	(1.22)
$\mathcal{E}_\varepsilon$, $\mathcal{J}_\varepsilon$, and $\mathcal{A}_\varepsilon$	(Section 1) general energy, dissipation fuctional, and action	
$\mathcal{E}_\varepsilon$, $\mathcal{J}_\varepsilon$, and $\mathcal{A}_\varepsilon$	(Sections 2–9) Wasserstein energy, dissipation fuctional, and action, i.e the same as $\mathcal{E}_\varepsilon^{\mathrm{free}}$, $\mathcal{J}_\varepsilon^{\mathrm{Wass}}$, and $\mathcal{A}_\varepsilon^{\mathrm{Wass}}$	(1.20), (1.24), (1.25)
$\hat{\mathcal{E}}_\varepsilon$, $\hat{\mathcal{J}}_\varepsilon$, and $\hat{\mathcal{A}}_\varepsilon$	$\mathcal{E}_\varepsilon$, $\mathcal{J}_\varepsilon$, and $\mathcal{A}_\varepsilon$ written in terms of $\hat\rho$	(2.8), (2.12), (2.13)
$\mathcal{E}_0$, $\mathcal{J}_0$, and $\mathcal{A}_0$	limit energy, dissipation, and action	(1.26), (1.29), (1.32)
γ_ε, $\hat\gamma_\varepsilon$	invariant measure (γ_ε) and its push-forward under $\hat s_\varepsilon$	(1.2) and (2.5)
g_ε, $\hat g_\varepsilon$	Lebesgue densities of γ_ε and $\hat\gamma_\varepsilon$	(2.1), (2.6)
H	'enthalpy function', potential for the Brownian particle	page 3
$k = 1/\kappa$	reaction parameter	(1.5)
$M(\cdot,\cdot)$	argument of the integral in $\mathcal{J}_0$	(1.30)
$\hat s_\varepsilon$	transform from ξ to s, inverse of $\hat\xi_\varepsilon$	(2.3)
τ_ε	time rescaling	(2.2)
u_ε	density $\mathrm{d}\rho_\varepsilon/\mathrm{d}\gamma_\varepsilon$	(1.6)
$\hat u_\varepsilon$	transform of u_ε, $\hat u_\varepsilon = u_\varepsilon \circ \xi_\varepsilon$	(2.7)
$\hat\xi_\varepsilon$	transform from s to ξ, inverse of $\hat s_\varepsilon$	(2.3)
Z_ε	normalization constant of γ_ε	(1.2)

2. Rescaling

From here on we write $\mathcal{E}_\varepsilon$, $\mathcal{J}_\varepsilon$, and $\mathcal{A}_\varepsilon$ for $\mathcal{E}_\varepsilon^{\mathrm{free}}$, $\mathcal{J}_\varepsilon^{\mathrm{Wass}}$, and $\mathcal{A}_\varepsilon^{\mathrm{Wass}}$, since we will only be using the Wasserstein framework. Since for most of the discussion the interval (a,b) will be fixed to $(0,T)$, we will also write $\mathcal{J}(\rho)$ for $\mathcal{J}(\rho;0,T)$ etcetera.

A central step in the analysis of this paper is a rescaling of the domain which stretches the region around $\xi = 0$. This converts the functions u_ε, which have steep gradients around $\xi = 0$ (see Figure 3), into functions $\hat u_\varepsilon$ of the new variable s that will have a more regular behaviour.

We call g_ε the Lebesgue density of γ_ε, namely

$$g_\varepsilon(\xi) := Z_\varepsilon^{-1}\mathrm{e}^{-H(\xi)/\varepsilon}, \text{ and we set } \kappa := \frac{1}{k} = \frac{\sqrt{|H''(0)|\,H''(1)}}{\pi} \quad \text{(as in (1.5))}. \tag{2.1}$$

We now make the choice of τ_ε precise:

$$\tau_\varepsilon := \frac{1}{2\kappa}\int_{-1}^{1}\frac{\mathrm{d}\xi}{g_\varepsilon(\xi)}. \tag{2.2}$$

An application of Watson's Lemma gives the asymptotic estimate

$$\tau_\varepsilon \Big/ \varepsilon \mathrm{e}^{1/\varepsilon} \quad \overset{\varepsilon\to 0}{\longrightarrow} \quad 1.$$

Using that g_ε is even (since H is even), we introduce the smooth increasing diffeomorphism $\xi \mapsto s = \hat s_\varepsilon(\xi)$,

$$\hat s_\varepsilon : [-1,1] \to [-\kappa,\kappa], \quad \hat s_\varepsilon(\xi) := \int_0^\xi \frac{1}{\tau_\varepsilon g_\varepsilon(\eta)}\,\mathrm{d}\eta, \quad \text{with inverse} \quad \hat\xi_\varepsilon := \hat s_\varepsilon^{-1} : [-\kappa,\kappa] \to [-1,1]. \tag{2.3}$$

Note that $\hat{\xi}_\varepsilon$ satisfies

$$\frac{\mathrm{d}}{\mathrm{d}s}\hat{\xi}_\varepsilon(s) = \tau_\varepsilon g_\varepsilon(\hat{\xi}_\varepsilon(s)), \quad \hat{\xi}_\varepsilon(-\kappa) = -1, \text{ and } \hat{\xi}_\varepsilon(\kappa) = 1. \tag{2.4}$$

With this change of variables we call $\mathrm{S} := [-\kappa, \kappa]$ the domain of the variable s and we set

$$\hat{\gamma}_\varepsilon := (\hat{s}_\varepsilon)_\# \gamma_\varepsilon, \quad \hat{\rho}_\varepsilon := (\hat{s}_\varepsilon)_\# \rho_\varepsilon. \tag{2.5}$$

Observe that the Lebesgue density $\hat{g}_\varepsilon$ of $\hat{\gamma}_\varepsilon$ satisfies

$$\hat{g}_\varepsilon(\hat{s}_\varepsilon(\xi))\frac{\mathrm{d}}{\mathrm{d}\xi}\hat{s}_\varepsilon(\xi) = g_\varepsilon(\xi) \quad \text{so that} \quad \hat{g}_\varepsilon(s) = \tau_\varepsilon g_\varepsilon^2(\hat{\xi}_\varepsilon(s)), \tag{2.6}$$

and the transformed measure $\hat{\rho}_\varepsilon$ satisfies

$$\hat{\rho}_\varepsilon = \hat{u}_\varepsilon \hat{\gamma}_\varepsilon, \quad \hat{u}_\varepsilon := u_\varepsilon \circ \hat{\xi}_\varepsilon. \tag{2.7}$$

In particular, we can easily transport the entropy functional to the new setting

$$\hat{\mathcal{E}}_\varepsilon(\hat{\rho}) := \int_{\mathrm{S}} \hat{u}(s) \log \hat{u}(s)\, \hat{\gamma}_\varepsilon(\mathrm{d}s), \quad \text{so that} \quad \hat{\mathcal{E}}_\varepsilon(\hat{\rho}) = \mathcal{E}_\varepsilon(\rho) \quad \text{if } \hat{\rho} = (\hat{s}_\varepsilon)_\# \rho. \tag{2.8}$$

If $(\rho, \rho v) \in \mathsf{CE}(a, b; \Xi)$ then the couple $(\hat{\rho}, \hat{\rho}\hat{v})$ with $\hat{v}(s) = v(\hat{\xi}(s))/(\tau_\varepsilon g_\varepsilon(\hat{\xi}_\varepsilon(s)))$ belongs to $\mathsf{CE}(a, b; \mathrm{S})$ and satisfies the continuity equation

$$\partial_t \hat{\rho} + \partial_s(\hat{\rho}\, \hat{v}) = 0 \quad \text{in } \mathcal{D}'((0, T) \times \mathbb{R}); \tag{2.9}$$

in fact, since $v(\hat{\xi}(s)) = \hat{v}(s)\hat{\xi}_\varepsilon'(s)$, for every $\hat{\phi} = \phi \circ \hat{\xi} \in C_{\mathrm{c}}^\infty((0, T) \times [0, 1])$ we have

$$\begin{aligned}
0 &= \int_0^T \int_{-1}^1 \Big(\partial_t \phi + v\, \partial_\xi \phi\Big)\, \rho_t(\mathrm{d}\xi)\, \mathrm{d}t = \int_0^T \int_{\mathrm{S}} \Big((\partial_t \phi) \circ \hat{\xi}_\varepsilon + (v \circ \hat{\xi}_\varepsilon)\, (\partial_\xi \phi) \circ \hat{\xi}_\varepsilon\Big)\, \hat{\rho}_t(\mathrm{d}s)\, \mathrm{d}t \\
&= \int_0^T \int_{\mathrm{S}} \Big(\partial_t \hat{\phi} + \hat{\xi}_\varepsilon'\, \hat{v}\, (\partial_\xi \phi) \circ \hat{\xi}_\varepsilon\Big)\, \hat{\rho}_t(\mathrm{d}s)\, \mathrm{d}t = \int_0^T \int_{\mathrm{S}} \Big(\partial_t \hat{\phi} + \hat{v}\, \partial_s \hat{\phi}\Big)\, \hat{\rho}_t(\mathrm{d}s)\, \mathrm{d}t.
\end{aligned}$$

Setting $\hat{w} := \hat{v}\, \hat{u}\, \hat{g}_\varepsilon$ we also have

$$\frac{1}{2\tau_\varepsilon} \int_\Xi v^2\, \mathrm{d}\rho = \frac{1}{2\tau_\varepsilon} \int_{\mathrm{S}} \hat{v}^2\, \tau_\varepsilon^2 g_\varepsilon^2(\hat{\xi}_\varepsilon)\, \mathrm{d}\hat{\rho} = \frac{1}{2} \int_{\mathrm{S}} \hat{v}^2\, \hat{g}_\varepsilon^2 \hat{u}\, \mathrm{d}s = \frac{1}{2} \int_{\mathrm{S}} \frac{\hat{w}^2}{\hat{u}}\, \mathrm{d}s. \tag{2.10}$$

Since

$$\partial_s \sqrt{\hat{u}} = (\partial_\xi \sqrt{u} \circ \hat{\xi}_\varepsilon)\, \hat{\xi}_\varepsilon' = \tau_\varepsilon\, (\partial_\xi \sqrt{u} \circ \hat{\xi}_\varepsilon)\, g_\varepsilon \circ \hat{\xi}_\varepsilon$$

we also get

$$2\tau_\varepsilon \int_\Xi \Big|\partial_\xi \sqrt{u}\Big|^2 \mathrm{d}\gamma_\varepsilon = 2 \int_{\mathrm{S}} \Big|\partial_s \sqrt{\hat{u}}\Big|^2 \frac{1}{\tau_\varepsilon g_\varepsilon^2(\hat{\xi}_\varepsilon)}\, \mathrm{d}\hat{\gamma}_\varepsilon = 2 \int_{\mathrm{S}} \Big|\partial_s \sqrt{\hat{u}}\Big|^2 \mathrm{d}s. \tag{2.11}$$

Combining (2.9), (2.10), and (2.11), we now define the functional

$$\hat{\mathcal{J}}_\varepsilon(\hat{\rho}; 0, T) := \int_0^T \Big(\frac{1}{2} \int_{\mathrm{S}} \frac{\hat{w}^2}{\hat{u}}\, \mathrm{d}s + 2 \int_{\mathrm{S}} \Big|\partial_s \sqrt{\hat{u}}\Big|^2 \mathrm{d}s\Big)\, \mathrm{d}t, \quad \hat{u} = \frac{\mathrm{d}\hat{\rho}}{\mathrm{d}\hat{\gamma}_\varepsilon}, \quad \hat{w} = \hat{u}\, \hat{v}\, \hat{g}_\varepsilon, \tag{2.12}$$

and $(\hat{\rho}, \hat{\rho}\hat{v}) = (\hat{\rho}, \hat{w}\mathcal{L}^2) \in \mathsf{CE}(a, b; \mathrm{S})$. This calculation shows that

$$\begin{aligned}
&\mathcal{J}_\varepsilon(\rho; 0, T) = \hat{\mathcal{J}}_\varepsilon(\hat{\rho}; 0, T), \quad \text{and} \\
&\mathcal{A}_\varepsilon(\rho; 0, T) = \hat{A}_\varepsilon(\hat{\rho}; 0, T) = \hat{\mathcal{E}}_\varepsilon(\hat{\rho}(b)) - \hat{\mathcal{E}}_\varepsilon(\hat{\rho}(a)) + \hat{\mathcal{J}}_\varepsilon(\hat{\rho}; 0, T).
\end{aligned} \tag{2.13}$$

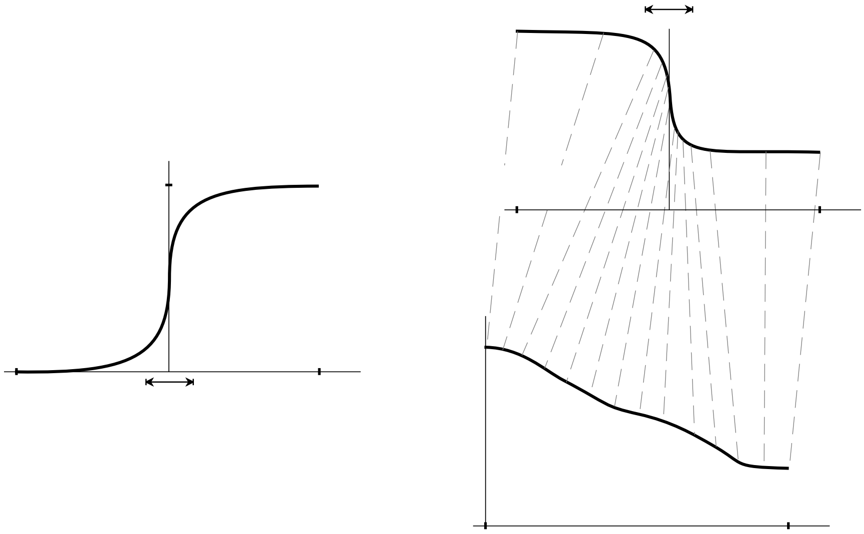

FIGURE 4. The transformations from ξ to s and from $u_\varepsilon(\xi)$ to $\hat{u}_\varepsilon(s)$. The left-hand graph shows the bijection between $\xi \in [-1,1]$ and $s \in [0,1]$. The right-hand graphs illustrate how the transformation (2.7) expands the region around $\xi = 0$ and converts the function u_ε with a near-discontinuity around $\xi = 0$ into a function $\hat{u}_\varepsilon$ that has a slope of order $O(1)$.

Remark 2.1. The desingularizing effect of the transformation from u_ε to $\hat{u}_\varepsilon$ can best be recognized in the last term in (2.12). In terms of $\hat{u}_\varepsilon$, this term is the H^1-seminorm of $\sqrt{\hat{u}_\varepsilon}$, and indeed boundedness of $\hat{\mathcal{J}}_\varepsilon$ implies boundedness of $\sqrt{\hat{u}_\varepsilon}$ in $L^2(0,T;H^1(\mathrm{S}))$ (see the proof of Theorem 3.2). Compare this with the corresponding term in the non-transformed version (2.11), where the vanishing of $\tau_\varepsilon\gamma_\varepsilon$ close to $\xi = 0$ allows for large gradients at that point.

As an independent way of viewing the effect of the transformation, the equation (1.7) for u_ε transforms into the equation

$$\hat{g}_\varepsilon \, \partial_t \hat{u}_\varepsilon = \partial_{ss} \hat{u}_\varepsilon$$

for $\hat{u}_\varepsilon$. Here the structure of the term $\partial_s^2 \hat{u}_\varepsilon$ (specifically, the lack of singular terms *inside* the derivatives) is related to the better behaviour of $\hat{u}_\varepsilon$. □

3. COMPACTNESS

The main results of this section, Theorems 3.1 and 3.2, describe compactness properties of sequences ρ_ε, and their transformed versions $\hat{\rho}_\varepsilon$, for which the initial energy $\mathcal{E}_\varepsilon(\rho_\varepsilon(0)) = \hat{\mathcal{E}}_\varepsilon(\hat{\rho}_\varepsilon(0))$ and $\mathcal{J}_\varepsilon(\rho_\varepsilon) = \hat{\mathcal{J}}_\varepsilon(\hat{\rho}_\varepsilon)$ are bounded.

Let us first comment on what one might expect. For ρ_ε and $\hat{\rho}_\varepsilon$ the limit objects are measures ρ_0 and $\hat{\rho}_0$ concentrated in $\{-1,1\}$ and $\{-\kappa,\kappa\}$. The existence of converging subsequences is a simple consequence of the bounded total variation of the measures and the bounded domain of definition. However, this convergence alone does not contain enough information for the lower bound result of Theorem 4.1.

The key to obtaining more detailed convergence statements lies in using the objects that appear in $\mathcal{J}_\varepsilon$ and $\hat{\mathcal{J}}_\varepsilon$, which are w_ε, $\hat{w}_\varepsilon$, $\partial_\xi u_\varepsilon$, and $\partial_s \hat{u}_\varepsilon$. Boundedness of $\mathcal{J}_\varepsilon(\rho_\varepsilon)$ implies, using the definitions (1.24) and (2.12), that there exists a constant $C > 0$ such that

$$\frac{1}{\tau_\varepsilon}\int_0^T\int_\Xi \frac{w_\varepsilon^2}{u_\varepsilon}\,\mathrm{d}\xi\mathrm{d}t \le C, \qquad \tau_\varepsilon\int_0^T\int_\Xi \frac{\partial_\xi u_\varepsilon^2}{u_\varepsilon}\,\gamma_\varepsilon(\mathrm{d}\xi)\mathrm{d}t \le C, \tag{3.1}$$

$$\int_0^T\int_{\mathrm{S}} \frac{\hat{w}_\varepsilon^2}{\hat{u}_\varepsilon}\,\mathrm{d}s\mathrm{d}t \le C, \qquad \int_0^T\int_{\mathrm{S}} \frac{\partial_s \hat{u}_\varepsilon^2}{\hat{u}_\varepsilon}\,\mathrm{d}s\mathrm{d}t \le C. \tag{3.2}$$

If the sequence $\hat{u}_\varepsilon$ also happens to be bounded in L^∞, then the bounds (3.2) imply weak compactness of $\hat{w}_\varepsilon$ and $\partial_s\hat{u}_\varepsilon$ in $L^2((0,T)\times \mathrm{S})$. Also, since $\tau_\varepsilon\gamma_\varepsilon$ is unbounded in any set not containing $\xi = 0$, (3.1) suggests that u_ε should become constant in $[-1,0)$ and $(0,1]$. In Theorem 3.1, where we make this additional assumption of boundedness in L^∞, we show that the remarks above indeed are true. Moreover, we shall see that we can recover the canonical decomposition $\rho_0 = \frac{1}{2}u^-\delta_{-1} + \frac{1}{2}u^+\delta_1$ by taking the limit of the traces of the densities u_ε at $\xi = \pm 1$, and similarly for $\hat{\rho}_0$.

When $\hat{u}_\varepsilon$ is not assumed to be bounded in L^∞, singular behaviour is possible that violates the L^∞ bound but influences neither the boundedness of energy and dissipation nor the limit object ρ_0. We treat this case in Theorem 3.2.

Theorem 3.1 (Compactness under uniform L^∞ bounds). *Let $\rho_\varepsilon = u_\varepsilon\gamma_\varepsilon \in C([0,T];\mathcal{Z}^{meas})$ satisfy, for suitable constants $m, C > 0$ and for all $\varepsilon > 0$, the estimates*

$$\rho_\varepsilon(t,\Xi) = m \text{ for all } t\in[0,T], \quad \mathcal{E}_\varepsilon(\rho_\varepsilon(0)) \le C, \quad \mathcal{J}_\varepsilon(\rho_\varepsilon) \le C, \quad \text{and} \quad \|u_\varepsilon\|_\infty \le C. \tag{3.3}$$

Then there exists a subsequence (not relabeled) and a limit $\rho_0 = u_0\gamma_0 \in C([0,T];\mathcal{Z}^{meas})$ such that the following hold:

1. *$\rho_\varepsilon(t) \xrightarrow{*} \rho_0(t)$ in $\mathcal{M}(\Xi)$ for every $t\in[0,T]$;*
2. *The spatial traces $u_\varepsilon(\cdot,\pm1)$ are well-defined and converge strongly to $u_0^\pm(\cdot) = u_0(\cdot,\pm1)$ in $L^1(0,T)$;*
3. *For all $0<\delta<1$ the function u_ε converges uniformly to the limiting trace values $u_0^\pm$ in $L^1(0,T;L^\infty(-1,-\delta))$ and $L^1(0,T;L^\infty(\delta,1))$.*

Let $\hat{u}_\varepsilon$ be the transformed sequence and let $\hat{w}_\varepsilon$ be given as in (2.12). *Then there exist limits $\hat{u}_0 \in L^\infty((0,T)\times \mathrm{S}) \cap L^2(0,T;W^{1,2}(\mathrm{S}))$ and $\hat{w}_0 \in L^2((0,T)\times\mathrm{S})$ such that*

4. *$\hat{u}_\varepsilon \xrightarrow{*} \hat{u}_0$ in $L^\infty((0,T)\times\mathrm{S})$, and $\partial_s\hat{u}_\varepsilon \rightharpoonup \partial_s\hat{u}_0$, $\hat{w}_\varepsilon \rightharpoonup \hat{w}_0$ in $L^2((0,T)\times\mathrm{S})$;*
5. *the traces $\hat{u}_0^\pm$ of $\hat{u}_0(\cdot,s)$ at $s=\pm\kappa$ coincide with the traces of u_0, i.e. they satisfy $\hat{u}_0^\pm = u_0^\pm$ in $(0,T)$;*
6. *$\hat{\rho}_\varepsilon \xrightarrow{*} \hat{\rho}_0(t) = \frac{1}{2}(u_0^-(t)\delta_{-\kappa} + u_0^+(t)\delta_\kappa)$ in $\mathcal{M}(\mathrm{S})$ for every $t\in[0,T]$;*
7. *$\hat{w}_0(t,\cdot)$ is constant in S for a.e. $t\in(0,T)$ and satisfies $\hat{w}_0(t,\cdot) = \frac{1}{2}\dot{u}_0^+(t)$ a.e. in $(0,T)$. The couple $(\hat{\rho}_0,\hat{\nu}_0)$, $\hat{\nu}_0 = \hat{w}_0\mathcal{L}^2|_{(0,T)\times\mathrm{S}}$ satisfies the continuity equation*

$$\partial_t\hat{\rho}_0 + \partial_s\hat{\nu}_0 = 0 \quad \text{in } \mathcal{D}'((0,T)\times\mathbb{R}). \tag{3.4}$$

We will see in Theorem 5.4 that in the special case of *solutions* of (1.1), which satisfy $\hat{\mathcal{A}}_\varepsilon(\hat{\rho}_\varepsilon) = 0$ rather than $\hat{\mathcal{A}}_\varepsilon(\hat{\rho}_\varepsilon) \le C$, the limit object $\hat{u}_0$ is a linear interpolation of the values at $s = \pm\kappa$.

Proof. We divide the proof in a few steps; we will denote by C various constants which are independent of ε.

Step 1: Entropy estimates. *There exists a constant $C>0$ such that*

$$\mathcal{E}_\varepsilon(\rho_\varepsilon(t)) = \hat{\mathcal{E}}_\varepsilon(\hat{\rho}_\varepsilon(t)) \le C \quad \text{for every } t\in[0,T]; \tag{3.5}$$

in particular, for any subsets $A \Subset (-1,1)$ and $\hat A \Subset (-\kappa,\kappa)$ we have

$$\lim_{\varepsilon\to 0}\sup_t \rho_\varepsilon(t,A)=0,\quad \lim_{\varepsilon\to 0}\sup_t \hat\rho_\varepsilon(t,\hat A)=0. \tag{3.6}$$

It is sufficient to prove (3.5) for the unrescaled measures $\rho(t)$. First we note that $\mathcal{E}_\varepsilon$ is nonnegative; denoting by $\check\rho(t) := \rho(T-t)$ the time-reversed curve, since $\mathcal{J}_\varepsilon(\check\rho) = \mathcal{J}_\varepsilon(\rho)$, the bounds (3.3) and (1.16) imply that $\mathcal{E}_\varepsilon(\rho_\varepsilon(t)) \le C$ for every $t\in[0,T]$.

Property (3.6) follows from (3.5) and the fact that $\lim_{\varepsilon\downarrow 0}\gamma_\varepsilon(A) = \lim_{\varepsilon\downarrow 0}\hat\gamma_\varepsilon(\hat A)=0$. Considering e.g. the case of $A\Subset(-1,1)$, by using first the inequality $r\log r\ge -\mathrm{e}^{-1}$ and then Jensen inequality, we get for every $A\Subset(-1,1)$

$$\frac{1}{\mathrm{e}}\gamma_\varepsilon((-1,1)\setminus A)+\mathcal{E}_\varepsilon(\rho_\varepsilon(t))+m\log m \ge \int_A u_\varepsilon(t,\xi)\log u_\varepsilon(t,\xi)\,\gamma_\varepsilon(\mathrm{d}\xi) \ge \rho_\varepsilon(t,A)\log\Big(\frac{\rho_\varepsilon(t,A)}{\gamma_\varepsilon(A)}\Big), \tag{3.7}$$

so that $\gamma_\varepsilon(A)\to 0$ implies $\sup_t \rho_\varepsilon(t,A)\to 0$ as $\varepsilon\to 0$.

Step 2: Estimates on $\hat u_\varepsilon, \hat w_\varepsilon$. *There exists a constant $C>0$ such that*

$$\|\hat u_\varepsilon\|_\infty \le C \quad\text{and}\quad \int_0^T\int_{\mathrm{S}} \hat w_\varepsilon(t,s)^2\,\mathrm{d}s\mathrm{d}t \le C. \tag{3.8}$$

The first bound derives from assumption (3.3), and the second follows easily from (3.2) and the L^∞-bound on $\hat u_\varepsilon$. We state these here to contrast with the corresponding, weaker, versions in the proof of Theorem 3.2.

Step 3: Pointwise weak convergence of $\hat\rho_\varepsilon(t)$ (statement 6): *there exists a sequence $\varepsilon_n\downarrow 0$ and a limit $\hat\rho_0(t)\ll\hat\gamma_0$ such that $\hat\rho_{\varepsilon_n}(t)\overset{*}{\rightharpoonup}\hat\rho_0(t)$ for every $t\in[0,T]$.*

Starting from the continuity equation we have for every $0\le t_0<t_1\le T$ and $\varphi\in C^1([0,1])$,

$$\int_{\mathrm{S}}\varphi\,\mathrm{d}\hat\rho_\varepsilon(t_1) - \int_{\mathrm{S}}\varphi\,\mathrm{d}\hat\rho_\varepsilon(t_0) = \int_{t_0}^{t_1}\int_{\mathrm{S}} \partial_s\varphi\;\hat w_\varepsilon\,\mathrm{d}s\mathrm{d}t.$$

Recalling the definition of the L^1-Wasserstein distance d_{W_1} [AGS05], we find

$$\begin{aligned}\mathsf{d}_{W_1}(\rho_\varepsilon(t_1),\rho_\varepsilon(t_0)) &:= \sup\Big\{\int_{\mathrm{S}}\varphi\,\mathrm{d}\hat\rho_\varepsilon(t_1) - \int_{\mathrm{S}}\varphi\,\mathrm{d}\hat\rho_\varepsilon(t_0) : \varphi\in C^1([0,1]),\ |\partial_s\varphi|\le 1\Big\}\\ &\le \int_{t_0}^{t_1}\int_{\mathrm{S}}|\hat w_\varepsilon(t,s)|\,\mathrm{d}s\mathrm{d}t \le \sqrt{2\kappa T}\Big(\int_0^T\int_{\mathrm{S}}\hat w_\varepsilon(t,s)^2\,\mathrm{d}s\mathrm{d}t\Big)^{1/2}.\end{aligned}$$

It follows by (3.8) that the curves $t\mapsto\hat\rho_{\varepsilon,t}$ have uniformly bounded total variation in the space $\mathcal{M}(\mathrm{S})$ endowed with the L^1-Wasserstein distance; since the total mass is m, the claim follows by Helly's compactness theorem.

Step 4. Weak convergence of $\rho_\varepsilon(t)$ (statement 1). *Writing the limit $\hat\rho_0$ of the previous step as $\hat\rho_0(t) = \frac12 u_0^-(t)\delta_{-\kappa}+\frac12 u_0^+(t)\delta_\kappa$, we have for every $t\in[0,T]$,*

$$\rho_{\varepsilon_n}(t)\overset{*}{\rightharpoonup}\rho_0(t) = \frac12 u_0^-(t)\delta_{-1}+\frac12 u_0^+\delta_1. \tag{3.9}$$

Let us fix $t\in[0,T]$ and let us consider a subsequence ε_n' of ε_n along which $\rho_{\varepsilon_n'}(t)$ converges weakly to some $\tilde\rho\in\mathcal{M}(\Xi)$. By the result of Step 1, we know that $\tilde\rho = \frac12\tilde u^-\delta_{-1}+\frac12\tilde u^+\delta_1$ for some $\tilde u^\pm$: if we show that $\tilde u^\pm = u_0^\pm(t)$ we have proved the thesis. Considering $\tilde u^-$ and taking a function $\hat\varphi\in C(\mathrm{S})$, we know that $\varphi_\varepsilon := \hat\varphi \circ s_\varepsilon$ converges pointwise to $\varphi_0(\xi) = \hat\varphi \circ s_0$, where $s_0(\xi) = \operatorname{sign}(\xi)$ (with $s_0(0) = 0$) and the convergence is uniform on compact subsets of $[-1,1] \setminus \{0\}$. It then follows that

$$\frac{1}{2}\sum_{\pm} \hat u_0^{\pm}(t)\hat\varphi(\pm\kappa) = \lim_{n\to\infty}\int_{\mathrm{S}} \hat\varphi(s)\,\hat\rho_{\varepsilon_n'}(t,\mathrm{d}s) = \lim_{n\to\infty}\int_1^1 \varphi_{\varepsilon_n'}(\xi)\,\rho_{\varepsilon_n'}(t,\mathrm{d}\xi) = \int_{-1}^1 \varphi_0(\xi)\,\tilde\rho(\mathrm{d}\xi)$$
$$= \frac{1}{2}\sum_{\pm} \tilde u^{\pm}\hat\varphi(\pm\kappa).$$

Since $\hat\varphi$ is arbitrary, we obtain (3.9).

This shows the intuitive result that the densities of u_0 and $\hat u_0$ (with respect to γ_0 and $\hat\gamma_0$) are the same; we call them both $u_0^{\pm}$. It mirrors the fact that the traces of u_ε (in $\xi = \pm 1$) and of $\hat u_\varepsilon$ (in $s = \pm\kappa$) are also the same.

Step 5. Convergence of u_ε (statements 2 and 3): *The traces $u_{\varepsilon_n}^{\pm} = u_{\varepsilon_n}(\,\cdot\,, \pm 1)$ strongly converge in $L^1(0,T)$ to the limits $u_0^{\pm}$ defined in the previous step. In addition, setting $\omega_\delta^{\pm} := \pm(\delta, 1)$ we have*

$$\lim_{n\to\infty}\int_0^T \sup_{\xi\in\omega_\delta^{\pm}} |u_{\varepsilon_n}(t,\xi) - u_0^{\pm}(t)|\,\mathrm{d}t = 0 \quad \text{for every } 0<\delta<1. \tag{3.10}$$

Let us first observe that the quantities

$$\bar u_{\varepsilon_n,\delta}^{\pm}(t) := \frac{\rho_{\varepsilon_n,t}(\omega_\delta^{\pm})}{\gamma_\varepsilon(\omega_\delta^{\pm})} = \frac{1}{\gamma_\varepsilon(\omega_\delta^{\pm})}\int_{\omega_\delta^{\pm}} u_{\varepsilon_n}(t,\xi)\,\mathrm{d}\gamma_\varepsilon(\xi), \quad 0<\delta<1,$$

are uniformly bounded and converge pointwise to $u_0^{\pm}(t)$ for every $t \in (0,T)$ by Step 4. By (3.1) and the boundedness of u_ε we have

$$\lim_{\varepsilon\downarrow 0}\int_0^T |\theta_\varepsilon^{\pm}|^2(t)\,\mathrm{d}t = 0, \quad \text{where} \quad \theta_\varepsilon^{\pm}(t) := \sup_{\xi,\eta\in\omega_\delta^{\pm}} |u_\varepsilon(t,\xi) - u_\varepsilon(t,\eta)| \le \Big(\delta\int_{\omega_\delta^{\pm}} |\partial_\xi u_\varepsilon|^2\,\mathrm{d}\xi\Big)^{1/2}.$$

We then calculate for $\xi,\eta \in \omega_\delta^{\pm}$,

$$|u_\varepsilon(t,\xi) - \bar u_{\varepsilon_n,\delta}^{\pm}(t)| \le \frac{1}{\gamma_\varepsilon(\omega_\delta^{\pm})}\int_{\omega_\delta^{\pm}} |u_\varepsilon(t,\xi) - u_\varepsilon(t,\eta)|\,\gamma_\varepsilon(\mathrm{d}\eta) \le \frac{\theta_\varepsilon^{\pm}(t)}{\gamma_\varepsilon(\omega_\delta^{\pm})}.$$

Recalling that $\gamma_\varepsilon(\omega_\delta^{\pm}) \to 1/2$ as $\varepsilon \to 0$, we thus obtain

$$\lim_{n\to\infty}\int_0^T \sup_{\xi\in\omega_\delta^{\pm}} |u_{\varepsilon_n}(\xi,t) - \bar u_{\varepsilon_n,\delta}^{\pm}(t)|\,\mathrm{d}t = 0$$

which in particular yields (3.10) and the convergence of the traces, since $\bar u_{\varepsilon_n,\delta}^{\pm}$ strongly converge to $u_0^{\pm}$ in every $L^p(0,T)$, $p<+\infty$.

Step 6. Compactness and limits (statement 4). Given the estimates (3.8) and (3.2), this follows from standard results.

Step 7. Identification of the traces of $\hat u_0$ (statement 5). Since the trace operator Tr is weakly continuous from $H^1(\mathrm{S})$ to $L^2(\{-\kappa,\kappa\}) \simeq \mathbb{R}^2$, the weak convergence of $\hat u_{\varepsilon_n}$ in $H^1(\mathrm{S})$ implies that its traces $\hat u_{\varepsilon_n}(\cdot,\pm\kappa) = u_{\varepsilon_n}(\cdot,\pm 1)$ converge weakly in $L^2(0,T;\mathbb{R}^2)$ to $\mathrm{Tr}\,\hat u_0$. Since $u_{\varepsilon_n}(\cdot,\pm 1)$ converges strongly to $u_0^{\pm}$ in $L^1(0,T)$ (Step 5), it follows that $\mathrm{Tr}\,\hat u_0 = u_0^{\pm}$.

Step 8: The continuity equation and the structure of $\hat{w}_0$ (statement 7). Passing to the limit in the continuity equation (2.9) and using the previous convergence result we immediately find (3.4). Since $\hat\rho_0$ is supported in $[0,T]\times\{-\kappa,\kappa\}$, we obtain that $\partial_s\hat\nu_0=0$ in $[0,T]\times(-\kappa,\kappa)$, so that w_0 depends only on t.

Choosing a test function of the form $\varphi(t,s)=\psi(t)\zeta(s)$ with $\psi\in C_{\mathrm{c}}^\infty(0,T)$ and $\zeta\in C_{\mathrm{c}}^\infty(\mathbb{R})$, $\zeta\equiv 1$ on a neighborhood of κ and $\zeta\equiv 0$ on $(-\infty,0]$, we obtain

$$\frac12\int_0^T\dot\psi(t)\hat u_0^+(t)\,\mathrm{d}t=\int_0^T\int_{\mathbb{R}}\dot\psi(t)\zeta(s)\,\hat\rho_0(t,\mathrm{d}s)\,\mathrm{d}t=-\int_0^T\int_{\mathrm{S}}\psi(t)\zeta'(s)\hat w_0(t)\,\mathrm{d}s\,\mathrm{d}t.$$

Since $\int_{\mathrm{S}}\zeta'(s)\,\mathrm{d}s=\zeta(\kappa)=1$, we conclude that $\hat w_0$ is the distributional derivative of $\frac12\hat u_0^+$. This concludes the proof of Theorem 3.1. $\square$

We now discuss the case where $\hat u_\varepsilon$ is not assumed to be bounded in L^∞. A simple example shows how $\hat u_\varepsilon$ may become singular without affecting any of the relevant limit processes. Take any sequence $\hat\rho_\varepsilon$ with bounded $\hat{\mathcal{E}}_\varepsilon(\hat\rho_\varepsilon(0))$ and $\hat{\mathcal{J}}_\varepsilon(\hat\rho_\varepsilon)$, and let $\hat u_\varepsilon=\mathrm{d}\hat\rho_\varepsilon/\mathrm{d}\hat\gamma_\varepsilon$ be bounded from above and away from zero as well. Fix two nonnegative functions $\varphi\in C_c^1((-\kappa,\kappa))$ and $\psi\in C_c^1(\mathbb{R})$, fix $0<t_0<T$, and define

$$\widetilde\rho_\varepsilon(t,s):=\hat\rho_\varepsilon(t,s)+\frac{1}{\sqrt\varepsilon}\psi\Big(\frac{t-t_0}{\varepsilon}\Big)\varphi(s)\hat\gamma_\varepsilon(s).$$

Note that since the additional term blows up polynomially, while $\hat\gamma_\varepsilon$ converges to zero exponentially fast on $\operatorname{supp}\varphi$, the limits of $\widetilde\rho_\varepsilon$ and $\hat\rho_\varepsilon$ are the same. For the same reason the perturbed $\widetilde w_\varepsilon$, satisfying $\partial_t\widetilde\rho_\varepsilon+\partial_s\widetilde w_\varepsilon=0$, only differs from $\hat w_\varepsilon$ by an exponentially small amount. Therefore

$$\limsup_{\varepsilon\to0}\int_0^T\int_{\mathrm{S}}\frac{\widetilde w_\varepsilon^2}{\widetilde u_\varepsilon}\,\mathrm{d}s\mathrm{d}t=\limsup_{\varepsilon\to0}\int_0^T\int_{\mathrm{S}}\frac{\hat w_\varepsilon^2}{\hat u_\varepsilon}\,\mathrm{d}s\mathrm{d}t<\infty.$$

We also estimate

$$\int_0^T\int_{\mathrm{S}}\frac{\partial_s\widetilde u_\varepsilon^2}{\widetilde u_\varepsilon}\,\mathrm{d}s\mathrm{d}t\le 2\int_0^T\int_{\mathrm{S}}\frac{\partial_s\hat u_\varepsilon^2}{\hat u_\varepsilon}\,\mathrm{d}s\mathrm{d}t+\frac2\varepsilon\int_0^T\psi^2\Big(\frac{t-t_0}{\varepsilon}\Big)\int_{\mathrm{S}}\frac{\varphi'^2}{\hat u_\varepsilon}\,\mathrm{d}s\mathrm{d}t.$$

The first term of this estimate is bounded by assumption, and the second is bounded by the scaling in ε and the assumption that $\hat u_\varepsilon$ is uniformly bounded away from zero.

This example shows that the assumptions of bounded initial energy and bounded $\mathcal{J}$ do not rule out singular behaviour of the sequence $\hat u_\varepsilon$ *between* $-\kappa$ and κ. The example also suggests what form this singular behaviour might take: that of a singular measure in time (called $\lambda^\perp$ below), but with bounded total variation. This is exactly what we prove in the following theorem.

Theorem 3.2 (Compactness, the general case). *Let $\rho_\varepsilon=u_\varepsilon\gamma_\varepsilon\in C([0,T];\mathcal{Z}^{meas})$ satisfy, for suitable constants $m,C>0$ and for all $\varepsilon>0$, the estimates*

$$\rho_\varepsilon(t,\Xi)=m \text{ for all } t\in[0,T],\quad \mathcal{E}_\varepsilon(\rho_\varepsilon(a))\le C,\quad \text{and } \mathcal{J}_\varepsilon(\rho_\varepsilon)\le C.$$

Then there exists a subsequence (not relabeled) and a limit $\rho_0=u_0\gamma_0\in C([0,T];\mathcal{Z}^{meas})$ such that the following hold:

1. *$\rho_\varepsilon(t)\xrightarrow{*}\rho_0(t)$ in $\mathcal{M}(\Xi)$ for every $t\in[0,T]$;*
2. *The spatial traces $u_\varepsilon(\cdot,\pm1)$ are well-defined and converge strongly to $u_0^\pm(\cdot)=u_0(\cdot,\pm1)$ in $L^1(0,T)$;*
3. *For all $0<\delta<1$ the function u_ε converges uniformly to the limiting trace values $u_0^\pm$ in $L^1(0,T;L^\infty(-1,-\delta))$ and $L^1(0,T;L^\infty(\delta,1))$.*

Let $\hat{u}_\varepsilon$ be the transformed sequence and let $\hat{w}_\varepsilon$ be given as in (2.12). *Then there exist limit functions $\hat{u}_0 \in L^1(0,T;W^{1,1}(-\kappa,\kappa))$, $\hat{w}_0 \in L^1((0,T)\times \mathrm{S})$, a reference singular measure $\lambda^\perp \in \mathcal{M}([0,T])$ with $\lambda^\perp \perp \mathcal{L}^1$, and a function $\hat{m}_0 \in L^\infty_{\Lambda^\perp}([0,T]\times\mathrm{S})$ with $\partial_s \hat{m}_0 \in L^2_{\Lambda^\perp}([0,T]\times\mathrm{S})$, where $\Lambda^\perp = \lambda^\perp \otimes \mathcal{L}^1|_{\mathrm{S}} \in \mathcal{M}([0,T]\times\mathrm{S})$, such that*

4 $\hat{u}_\varepsilon \xrightarrow{*} \hat{u}_0 + \hat{m}_0\Lambda^\perp$, $\partial_s\hat{u}_\varepsilon \xrightarrow{*} \partial_s\hat{u}_0 + \partial_s\hat{m}_0\Lambda^\perp$, *and* $\hat{w}_\varepsilon \xrightarrow{*} \hat{w}_0$ *in the duality with* $C([0,T]\times\mathrm{S})$*;*
5 *the traces $\hat{u}_0^\pm$ of $\hat{u}_0(\cdot,s)$ at $s=\pm\kappa$ coincide with the traces of u_0, i.e. they satisfy $\hat{u}_0^\pm = u_0^\pm$ a.e. in $(0,T)$; the traces $\hat{m}_0^\pm(t)$ of $\hat{m}_0(t,\cdot)$ vanish for $\lambda^\perp$-a.e. $t\in[0,T]$;*
6 $\hat{\rho}_\varepsilon \xrightarrow{*} \hat{\rho}_0(t) = \frac12(u_0^-(t)\delta_{-\kappa} + u_0^+(t)\delta_\kappa)$ *in $\mathcal{M}(\mathrm{S})$ for every $t\in[0,T]$;*
7 *$\hat{w}_0(t,\cdot)$ is constant in S for a.e. $t\in(0,T)$ and satisfies $\hat{w}_0(t,\cdot) = \frac12\dot{u}^+(t)$ a.e. in $(0,T)$. The couple $(\hat{\rho}_0,\hat{\nu}_0)$, $\hat{\nu}_0 = \hat{w}_0\mathcal{L}^2|_{(0,T)\times\mathrm{S}}$ satisfies the continuity equation*

$$\partial_t\hat{\rho}_0 + \partial_s\hat{\nu}_0 = 0 \quad \textit{in } \mathcal{D}'((0,T)\times\mathbb{R}). \tag{3.11}$$

Remark 3.3. Parts 1–3, 6, and 7 are the same as in Theorem 3.1. The main difference lies in the structure of the limits of $\hat{u}_\varepsilon$ and $\hat{w}_\varepsilon$ (part 4) and therefore the identification of the traces of $\hat{u}_0$ and $\hat{m}_0$ (part 5). □

Proof. Some of the steps are the same as in the case of Theorem 3.1; for those we only give the statement. For the others we detail the differences.

Step 1: Entropy estimates. *There exists a constant $C>0$ such that*

$$\mathcal{E}_\varepsilon(\rho_\varepsilon(t)) = \hat{\mathcal{E}}_\varepsilon(\hat{\rho}_\varepsilon(t)) \le C \quad \textit{for every } t\in[0,T];$$

in particular, for any subsets $A \Subset (-1,1)$ and $\hat{A} \Subset (-\kappa,\kappa)$ we have

$$\lim_{\varepsilon\to0}\sup_t \rho_\varepsilon(t,A) = 0, \quad \lim_{\varepsilon\to0}\sup_t \hat{\rho}_\varepsilon(t,\hat{A}) = 0.$$

Step 2: Estimates on $\hat{u}_\varepsilon, \hat{w}_\varepsilon$. *There exists a constant $C>0$ such that*

$$\int_0^T \sup_{s\in\mathrm{S}} |\hat{u}_\varepsilon(t,s)|\,\mathrm{d}t \le C, \quad \int_0^T\int_{\mathrm{S}} |\hat{w}_\varepsilon(t,s)|\,\mathrm{d}s\,\mathrm{d}t \le C. \tag{3.12}$$

These bounds are weaker than the L^∞-bound on $\hat{u}_\varepsilon$ and the L^2-bound on $\hat{w}_\varepsilon$ of Theorem 3.1. Let us set $\hat{p}_\varepsilon := \sqrt{\hat{u}_\varepsilon}$: since $\hat{p}_\varepsilon \in L^2(0,T;W^{1,2}(-\kappa,\kappa))$ its traces at $s=\pm\kappa$ are well defined and belong to $L^2(0,T)$. We set $\hat{\theta}_\varepsilon(t) := \sup_{r,s\in\mathrm{S}} |\hat{p}_\varepsilon(t,r) - \hat{p}_\varepsilon(t,s)|$. Standard estimates yield

$$\hat{\theta}_\varepsilon(t) \le \Big(2\kappa\int_{\mathrm{S}} |\partial_s\hat{p}_\varepsilon|^2\,\mathrm{d}s\Big)^{1/2}, \quad \int_0^T \hat{\theta}_\varepsilon^2(t)\,\mathrm{d}t \le \hat{\mathcal{J}}_\varepsilon(\hat{\rho}_\varepsilon;0,T) \le C.$$

Moreover

$$\Big(\int_{\mathrm{S}} \hat{p}_\varepsilon\,\mathrm{d}\hat{\gamma}_\varepsilon\Big)^2 \le \int_{\mathrm{S}} \hat{p}_\varepsilon^2\,\mathrm{d}\hat{\gamma}_\varepsilon = m,$$

and

$$\hat{p}_\varepsilon(t,s) \le \int_{\mathrm{S}} \hat{p}_\varepsilon(r)\,\mathrm{d}\hat{\gamma}_\varepsilon(r) + \hat{\theta}_\varepsilon(t) \le \sqrt{m} + \hat{\theta}_\varepsilon(t) \quad \text{for every } s\in\mathrm{S},$$

and therefore

$$\sup_{s\in\mathrm{S}} \hat{u}_\varepsilon(t,s) \le 2m + 2\hat{\theta}_\varepsilon^2(t).$$

The second estimate of (3.12) then follows from

$$\int_{\mathrm{S}} |\hat w_\varepsilon(t,s)|\,\mathrm{d}s \le \Big(\int_{\mathrm{S}} \frac{|\hat w_\varepsilon(t,s)|^2}{\hat u_\varepsilon}\,\mathrm{d}s\Big)^{1/2}\Big(\int_{\mathrm{S}} \hat u_\varepsilon(t,s)\,\mathrm{d}s\Big)^{1/2}.$$

Step 3: Pointwise weak convergence of $\hat\rho_\varepsilon(t)$ (statement 6): *there exists a sequence $\varepsilon_n \downarrow 0$ and a limit $\hat\rho_0(t)$ such that $\hat\rho_{\varepsilon_n}(t) \overset{*}{\rightharpoonup} \hat\rho_0(t)$ for every $t\in[0,T]$, and $\hat\rho_0(t) \ll \hat\gamma_0$.*

Step 4. Weak convergence of $\rho_\varepsilon(t)$ (statement 1). *Writing the limit $\hat\rho_0$ of the previous step as $\hat\rho_0(t) = \frac12 u_0^-(t)\delta_{-\kappa} + \frac12 u_0^+(t)\delta_\kappa$, we have for every $t\in[0,T]$,*

$$\rho_{\varepsilon_n}(t) \overset{*}{\rightharpoonup} \rho_0(t) = \frac12 u_0^-(t)\delta_{-1} + \frac12 u_0^+\delta_1.$$

Step 5. Strong convergence of traces (statements 2 and 3): *the traces $u^\pm_{\varepsilon_n} = u_{\varepsilon_n}(\,\cdot\,,\pm1)$ strongly converge in $L^1(0,T)$ to the limits $u_0^\pm$ defined in the previous step. In addition, setting $\omega^\pm_\delta := \pm(\delta,1)$ we have*

$$\lim_{n\to\infty}\int_0^T \sup_{\xi\in\omega^\pm_\delta} |u_{\varepsilon_n}(t,\xi) - u_0^\pm(t)|\,\mathrm{d}t = 0 \quad \text{for every } 0<\delta<1.$$

The proof of this step is similar to that of Theorem 3.2, but uses instead the estimate on $p_\varepsilon := \sqrt{u_\varepsilon}$,

$$\lim_{\varepsilon\downarrow0}\int_0^T |\theta^\pm_\varepsilon|^2(t)\,\mathrm{d}t = 0, \quad \theta^\pm_\varepsilon(t) := \sup_{\xi,\eta\in\omega^\pm_\delta} |p_\varepsilon(\xi,t) - p_\varepsilon(\eta,t)| \le \Big(\delta\int_{\omega^\pm_\delta} |\partial_\xi p_\varepsilon|^2\,\mathrm{d}\xi\Big)^{1/2}.$$

Step 6. Compactness and limits (statement 4). Because of the lack of an L^∞ bound, from here on the proof differs significantly from that of Theorem 3.1. Let us set $\ell_\varepsilon(t) := 1+\sup_{s\in\mathrm{S}}|\hat u_\varepsilon(t,s)|$. Up to extracting a suitable subsequence $\varepsilon\to0$ (without changing notation) we can assume that there exist weak limits $\lambda\in\mathcal{M}([0,T])$ and $\hat\mu_0,\hat\nu_0,\hat\varsigma_0,\hat\sigma_0\in\mathcal{M}([0,T]\times\mathrm{S})$, such that (identifying functions with the corresponding measures)

$$\ell_\varepsilon \overset{*}{\longrightharpoonup} \lambda, \quad \hat u_\varepsilon \overset{*}{\longrightharpoonup} \hat\mu_0, \quad \hat w_\varepsilon \overset{*}{\longrightharpoonup} \hat\nu_0, \quad \partial_s\hat u_\varepsilon \overset{*}{\longrightharpoonup} \hat\varsigma_0, \quad \text{and} \quad |\partial_s\hat u_\varepsilon| \overset{*}{\longrightharpoonup} \hat\sigma_0. \tag{3.13}$$

Since $\hat u_\varepsilon(t,s)\le\ell_\varepsilon(t)$ we have $\hat\mu_0\le\Lambda := \lambda\otimes\mathcal{L}^1|_{\mathrm{S}}$, so that $\hat\mu_0 = \hat m_0\Lambda$ for a suitable bounded Borel function $\hat m_0\in L^\infty_\Lambda([0,T]\times\mathrm{S})$. Since

$$\int_0^T\int_{\mathrm{S}}\Big(\frac{\hat w_\varepsilon^2}{\hat u_\varepsilon} + \frac{\partial_s\hat u_\varepsilon^2}{\hat u_\varepsilon}\Big)\,\mathrm{d}s\,\mathrm{d}t \le C,$$

it follows (see Lemma 3.5 below) that in the limit $\hat\nu_0\ll\hat\mu_0$ and $|\hat\varsigma_0|\le\hat\sigma_0\ll\hat\mu_0$. In particular $\hat\nu_0 = \hat n_0\Lambda$ and $\hat\varsigma_0 = \hat g_0\,\Lambda$ with $\hat n_0,\hat g_0\in L^\infty_\Lambda([0,T]\times\mathrm{S})$. Since $\partial_s\hat\mu_0 = \hat\varsigma_0$, we easily have for every couple of test functions $\psi\in C^\infty([0,T]),\varphi\in C^\infty_{\mathrm{c}}(-\kappa,\kappa)$,

$$\iint_{(0,T)\times\mathrm{S}} \psi(t)\varphi(s)\,\hat g_0(t,s)\,\mathrm{d}s\,\lambda(\mathrm{d}t) = -\iint_{(0,T)\times\mathrm{S}} \psi(t)\varphi'(s)\,\hat m_0(t,s)\,\mathrm{d}s\,\lambda(\mathrm{d}t).$$

Since ψ is arbitrary, we deduce that $\partial_s\hat m_0(t,\cdot) = \hat g_0(t,\cdot)$ in $L^\infty(\mathrm{S})$ for λ-a.e. $t\in[0,T]$. We also deduce that

$$\int_0^T\int_{\mathrm{S}}\Big(\frac{\hat n_0^2}{\hat m_0} + \frac{\hat g_0^2}{\hat m_0}\Big)\,\mathrm{d}s\,\lambda(\mathrm{d}t) \le C. \tag{3.14}$$

The measure $\lambda\in\mathcal{M}([0,T])$ can be decomposed as $\lambda = \ell\mathcal{L}^1 + \lambda^\perp$ with $\ell\in L^1(0,T)$ and $\lambda^\perp\perp\mathcal{L}^1$, and similarly $\Lambda = \ell\mathcal{L}^1\otimes\mathcal{L}^1|_{\mathrm{S}} + \Lambda^\perp$ with $\Lambda^\perp = \lambda^\perp\otimes\mathcal{L}^1|_{\mathrm{S}}$. We set $\hat u_0 := \hat m_0\ell$ and $\hat w_0 := \hat n_0\,\ell$, so that the limits in (3.13) can be decomposed as $\hat\mu_0 = \hat u_0 + \hat m_0\Lambda^\perp$, $\varsigma_0 = \partial_s\hat u_0 + \partial_s\hat u_0\Lambda^\perp$, and $\hat\nu_0 = \hat w_0 + \hat n_0\Lambda^\perp$. In Step 8 below we show that the last term, $\hat n_0\Lambda^\perp$, vanishes.

Step 7. $\hat{\mu}_0$ and $\hat{u}_0$ have equal traces (statement 5). Let us consider, e.g., the case of $-\kappa$ and take nonnegative test functions $\psi\in C([0,T])$ with $\sup_{t\in[0,T]}|\psi(t)|\le 1$, and $\varphi\in C([-\kappa,\kappa])$ with support in $[-\kappa,0)$ and integral 1, so that $\Phi(s) := \int_s^\kappa \varphi(r)\,\mathrm{d}r$ is decreasing, supported in $[-\kappa,0)$, and satisfies $\Phi(-\kappa) = 1$. We also set $\varphi_\delta(s) := \delta^{-1}\varphi(-\kappa+\delta^{-1}(s+\kappa))$, $\Phi_\delta(s) := \Phi(-\kappa+\delta^{-1}(s+\kappa))$, $\varphi_\delta(s) = -\Phi'_\delta(s)$. Denoting by Ω the product $[0,T]\times\mathrm{S}$, we have

$$\Big|\int_0^T \psi(t)u_0^-(t)\,\mathrm{d}t - \iint_\Omega \psi(t)\varphi_\delta(s)\,\hat\mu_0(\mathrm{d}s\mathrm{d}t)\Big| \le \int_0^T \big|u_0^-(t) - \hat u_\varepsilon^-(t,-\kappa)\big|\,\mathrm{d}t$$
$$+\iint_\Omega \psi(t)\varphi_\delta(s)|\hat u_\varepsilon^-(t,-\kappa) - \hat u_\varepsilon(t,s)|\,\mathrm{d}s\,\mathrm{d}t + \Big|\iint_\Omega \psi(t)\varphi_\delta(s)\hat u_\varepsilon(t,s)\,\mathrm{d}s\,\mathrm{d}t - \iint_\Omega \psi(t)\varphi_\delta(s)\,\hat\mu_0(\mathrm{d}s\mathrm{d}t)\Big|.$$

Passing to the limit as $\varepsilon\to 0$ the first and third terms vanish; concerning the second term, we have

$$\begin{aligned}\iint_\Omega \psi(t)\varphi_\delta(s)|\hat u_\varepsilon(t,-\kappa) - \hat u_\varepsilon(t,s)|\,\mathrm{d}s\,\mathrm{d}t &\le \iint_\Omega \psi(t)\varphi_\delta(s)\Big(\int_{-\kappa}^s |\partial_s\hat u_\varepsilon(t,r)|\,\mathrm{d}r\Big)\,\mathrm{d}s\,\mathrm{d}t\\ &= \int_0^T \psi(t)\Big(\int_{-\kappa}^\kappa \Phi_\delta(s)|\partial_s\hat u_\varepsilon(t,s)|\,\mathrm{d}s\Big)\,\mathrm{d}t.\end{aligned}$$

Combining these inequalities and passing to the limit as $\varepsilon\to 0$ we get

$$\Big|\int_0^T \psi(t)u_0^-(t)\,\mathrm{d}t - \iint_\Omega \psi(t)\varphi_\delta(s)\,\hat\mu_0(\mathrm{d}s\mathrm{d}t)\Big| \le \iint_\Omega \psi(t)\Phi_\delta(s)\,\hat\sigma_0(\mathrm{d}s\mathrm{d}t),$$

so that, applying Lebesgue's dominated convergence theorem with the fact that $\Phi_\delta(s)\to 0$ for $s>-\kappa$ and $0\le\Phi_\delta\le 1$, we obtain

$$\lim_{\delta\downarrow 0}\Big|\int_0^T \psi(t)u_0^-(t)\,\mathrm{d}t - \iint_\Omega \psi(t)\varphi_\delta(s)\,\hat\mu_0(\mathrm{d}s\mathrm{d}t)\Big| \le \iint_{[0,T]\times\{-\kappa\}} \psi(t)\,\hat\sigma_0(\mathrm{d}s\mathrm{d}t) = 0,$$

since $\hat\sigma_0 \ll \Lambda$.

On the other hand, recalling that $\hat\mu_0 = \hat m_0\,\Lambda$ and $\hat m_0\in L^1_\lambda(0,T;W^{1,1}(\mathrm{S}))$, an analogous argument yields for $\hat m_0^-(t) := \hat m_0(t,-\kappa)$,

$$\begin{aligned}\Big|\int_0^T \psi(t)\hat m_0^-(t)\,\lambda(\mathrm{d}t) - \iint_\Omega \psi(t)\varphi_\delta(s)\,\hat\mu_0(\mathrm{d}s\mathrm{d}t)\Big| &\le \iint_\Omega \psi(t)\varphi_\delta(s)\big|\hat m_0^-(t) - \hat m_0(t,s)\big|\,\mathrm{d}s\,\lambda(\mathrm{d}t)\\ &\le \iint_\Omega \Phi_\delta(s)\big|\hat g_0(t,s)\big|\,\mathrm{d}s\,\lambda(\mathrm{d}t) \ \xrightarrow{\delta\downarrow 0}\ 0.\end{aligned}$$

Since ψ is arbitrary, we conclude that

$$\hat m_0^\pm\,\lambda = u_0^\pm\mathcal{L}^1. \tag{3.15}$$

Step 8: Passing to the limit in the continuity equation (statement 7). This step is the same as in the proof of Theorem 3.1.

Conclusion: Vanishing of the singular part of $\hat{\nu}_0$, i.e. $\hat{w}_0\lambda^\perp = 0$. From (3.15) it follows that $\hat m_0^\pm(t) = 0$ for $\lambda^\perp$-a.e. $t\in[0,T]$. On the other hand, (3.14) yields for $\lambda^\perp$-a.e. $t\in[0,T]$ and for every $\eta>0$ and $\hat s(t)$

with $\hat{m}_0(\hat{s}(t), t) > 0$,

$$
\begin{aligned}
+\infty > \frac{1}{2}\int_{\mathrm{S}}\Big(\frac{\hat{w}_0(t)^2}{\hat{m}_0(t,s)} + \frac{\hat{g}_0(t,s)^2}{\hat{m}_0(t,s)}\Big)\,\mathrm{d}s &\geq \frac{1}{2}\int_{\mathrm{S}}\Big(\frac{\hat{w}_0(t)^2}{\eta + \hat{m}_0(t,s)} + \frac{\partial_s \hat{m}_0(t,s)^2}{\eta + \hat{m}_0(t,s)}\Big)\,\mathrm{d}s \\
\geq |\hat{w}_0(t)| \int_{\mathrm{S}} |\partial_s \log(\eta + \hat{m}_0(t,s))|\,\mathrm{d}s &\geq 2|\hat{w}_0(t)|\Big|\log\eta - \log(\eta + \hat{m}_0(t,\hat{s}(t)))\Big|.
\end{aligned}
$$

Since $\eta > 0$ is arbitrary, we conclude that $\hat{w}_0(t) = 0$ $\lambda^{\perp}$-a.e. $t \in [0,T]$. □

The Lemma below is similar to many other duality results (see e.g. [AFP00, §2.6] or [AGS05, Lemma 9.4.4]) and seems to have some wider usefulness. We state it in $\mathbb{R}^d$ for generality.

Lemma 3.4. *Let $\Omega \subset \mathbb{R}^d$. For $\mu \in \mathcal{M}(\Omega)$ and $\nu \in \mathcal{M}(\Omega;\mathbb{R}^d)$,*

$$
\frac{1}{2}\int_{\Omega}\Big|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\Big|^2\,\mathrm{d}\mu = \sup\bigg\{\int_{\Omega}\big[\,a\,\mathrm{d}\mu + b\cdot\mathrm{d}\nu\big] : \ a \in C_b(\Omega),\ b \in C_b(\Omega;\mathbb{R}^d),\ a + \frac{|b|^2}{2} \leq 0\bigg\}. \tag{3.16}
$$

In particular, if the right-hand side is finite, then $\nu \ll \mu$ and $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\Omega)$.

Proof. We write $\mathcal{F}(\nu|\mu)$ for the left-hand side, and $\mathcal{F}'(\nu|\mu)$ for the right-hand side. We first show that $\mathcal{F}'(\nu|\mu) \leq \mathcal{F}(\nu|\mu)$. If ν is not absolutely continuous with respect to μ, then $\mathcal{F}(\nu|\mu) = \infty$, and there is nothing to prove; if $\nu \ll \mu$, then we can write $\nu = f\mu$. For all a and b continuous, bounded, and satisfying $a + |b|^2/2 \leq 0$, we have

$$
\int_{\Omega}\big[a\,\mathrm{d}\mu + b\cdot\mathrm{d}\nu\big] = \int_{\Omega}\big[a + b\cdot f\big]\,d\mu \leq \int_{\Omega}\Big[a + \frac{|b|^2}{2} + \frac{|f|^2}{2}\Big]\,d\mu \leq \int_{\Omega}\frac{|f|^2}{2}\,d\mu = \mathcal{F}(\nu|\mu).
$$

To prove the opposite inequality, we assume that $\mathcal{F}'(\nu|\mu) < \infty$, and first show that $\nu \ll \mu$. Suppose not; then there exists a Borel set $A \subset \Omega$ such that $\mu(A) = 0$ and $\nu(A) \neq 0$. Take $c > 0$, set $a = -c\chi_A$, and define a sequence $a_n \in C_b(\Omega)$ such that $a_n \uparrow a$. Then $\int a_n\,\mathrm{d}\mu \to 0$ as $n \to \infty$. On the other hand, setting $b_n := \sqrt{-2a_n}\nu(A)/|\nu(A)|$, we have $\int b_n\cdot\mathrm{d}\nu \to \sqrt{2c}\,|\nu(A)| > 0$. Since c is arbitrary, this violates the finiteness of $\mathcal{F}'(\nu|\mu)$, and therefore $\nu \ll \mu$.

Again writing $\nu = f\mu$, with $f \in L^2(\mu)^d$, we now choose $b_n \in C_b(\Omega)$ such that $b_n \to f$ in $L^2(\mu)^d$, so that $\int b_n\cdot f\,\mathrm{d}\mu \to \int |f|^2\,\mathrm{d}\mu = 2\mathcal{F}(\nu|\mu)$. Setting $a_n := -|b_n|^2/2$ we have $a_n \to -|f|^2/2$ in $L^1(\mu)$, and therefore $\int a_n\,\mathrm{d}\mu \to -\mathcal{F}(\nu|\mu)$. The result follows. □

The above dual characterization (3.16) of the property $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\Omega)$ will now be used to characterize the limits in Step 6 of the above proof.

Lemma 3.5. *If $u_n \overset{*}{\rightharpoonup} \mu$ and $w_n \overset{*}{\rightharpoonup} \nu$, and*

$$
\sup_n \int_{\Omega}\frac{|w_n|^2}{u_n}\,\mathrm{d}x =: C < \infty,
$$

then $\nu \ll \mu$ with $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\Omega)$ and

$$
\int_{\Omega}\Big|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\Big|^2\,\mathrm{d}\mu \leq \liminf_{n\to\infty}\int_{\Omega}\frac{|w_n|^2}{u_n}. \tag{3.17}
$$

Proof. For each pair (a,b) as in the right-hand side of (3.16) we have

$$C \geq \int_\Omega \frac{|w_n|^2}{u_n}\,\mathrm{d}x \geq \int_\Omega \big[a u_n + b\cdot w_n\big]\,\mathrm{d}x \to \int_\Omega \big[a\,\mathrm{d}\mu + b\cdot \mathrm{d}\nu\big].$$

Thus, the hypothesis of Lemma 3.4 is satisfied and $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L^2_\mu(\Omega)$ follows.

Moreover, choosing a pair (a,b) in (3.16) that approximates the left-hand side in (3.17), we also obtain the desired estimate (3.17). ☐

4. Lower bound

Theorem 4.1 (Lower bound)**.** *Under the same conditions as in Theorems 3.1 or 3.2 let us assume, without loss of generality, that $\rho_\varepsilon(t) \xrightarrow{*} \rho_0(t)$ for every $t \in [0,T]$. Then*

$$\mathcal{J}_0(\rho_0) \leq \liminf_{\varepsilon\to 0} \mathcal{J}_\varepsilon(\rho_\varepsilon) \quad \textit{and} \quad \mathcal{E}_0(\rho_0(t)) \leq \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(\rho_\varepsilon(t)) \quad \textit{for every } t\in[0,T]. \tag{4.1}$$

Proof. The lower semicontinuity of the entropy functionals under weak convergence is well known, see e.g. [AGS05, Lemma 9.4.3] or [ASZ09, Lemma 6.2].

Turning to $\mathcal{J}_\varepsilon$, we can suppose by Theorem 3.2 (which contains Theorem 3.1) that

$$\hat u_\varepsilon \xrightarrow{*} \hat\mu_0 = \hat m_0\Lambda, \qquad \partial_s \hat u_\varepsilon \xrightarrow{*} \hat\varsigma_0 = \hat g_0\Lambda, \qquad \text{and} \qquad \hat w_\varepsilon \xrightarrow{*} \hat\nu_0 = \hat n_0\Lambda.$$

Setting $\hat u_0 = \hat m_0\ell$ as in the proof of Theorem 3.2, we also have $\hat g_0\ell = \partial_s\hat u_0$. By (3.17) we then have

$$\frac12\int_0^T\int_{\mathrm{S}} \Big[\Big(\frac{\mathrm{d}\hat\nu_0}{\mathrm{d}\hat\mu_0}\Big)^2 + \Big(\frac{\mathrm{d}\hat\varsigma_0}{\mathrm{d}\hat\mu_0}\Big)^2\Big]\,\mathrm{d}\hat\mu_0 \leq \liminf_{\varepsilon\to0}\mathcal{J}_\varepsilon(\rho_\varepsilon).$$

We now discard the singular part $\hat m_0\Lambda^\perp$ of $\hat\mu_0$ and again write $\hat w_0 := \hat n_0\ell$, by which we find

$$\frac12\int_0^T\int_{\mathrm{S}}\Big(\frac{\hat w_0^2}{\hat u_0} + \frac{\partial_s\hat u_0^2}{\hat u_0}\Big)\,\mathrm{d}s\,\mathrm{d}t \leq \frac12\int_0^T\int_{\mathrm{S}}\Big[\Big(\frac{\mathrm{d}\hat\nu_0}{\mathrm{d}\hat\mu_0}\Big)^2 + \Big(\frac{\mathrm{d}\hat\varsigma_0}{\mathrm{d}\hat\mu_0}\Big)^2\Big]\,\hat u_0\,\mathrm{d}s\mathrm{d}t \leq \liminf_{\varepsilon\to0}\mathcal{J}_\varepsilon(\rho_\varepsilon).$$

Recalling that the traces of $\hat u_0$ at $s=\pm\kappa$ coincide with $u_0^\pm$, and that $\hat w_0(t,s) = \hat w_0(t)$ is constant with respect to s with $\hat w_0 = \frac12\dot u^+$, we see that for a.e. t the integrand in the left-hand side of the previous inequality satisfies

$$M(\hat w_0(t), u_0^\pm(t)) \leq \frac12\int_{\mathrm{S}}\Big(\frac{\hat w_0(t)^2}{\hat u_0(t,s)} + \frac{\partial_s\hat u_0(t,s)^2}{\hat u_0(t,s)}\Big)\,\mathrm{d}s.$$

This implies the lower bound on $\mathcal{J}_\varepsilon$ and concludes the proof of Theorem 4.1. ☐

5. The minimization problem defining M and interpolation

The minimization problem defining M is

$$M(w,u^\pm) := \inf_u \Big\{ \frac12\int_{\mathrm{S}}\Big[\frac{w^2}{u(s)} + \frac{u'(s)^2}{u(s)}\Big]\,\mathrm{d}s \ : \quad u(\pm\kappa) = u^\pm \Big\}. \tag{5.1}$$

This minimization problem gives rise to a natural *interpolation* of the boundary values $u^\pm$, which we study in the following theorem.

Theorem 5.1. *Let $u[w, u^{\pm}](\cdot)$ be the solution of the minimization problem $M(w, u^{\pm})$. Then the mapping*

$$(w, u^{\pm}) \mapsto u[w, u^{\pm}]$$

is well-defined and continuous from $\mathbb{R} \times (0, \infty)^2$ into $C^2(\mathrm{S})$. The function $(w, u^{\pm}) \mapsto M(w, u^{\pm})$ is convex, smooth away from $u^{\pm} = 0$, minimal at $w = 0$ and $u^+ = u^-$, and satisfies $M(w, u^{\pm}) = M(w, u^{\mp})$.

If $u^{\pm} \in C^2([0, T]; [\delta, \infty))$ for some $\delta > 0$, then the function

$$(t, s) \mapsto u\Big[\frac{1}{2}\dot{u}^+(t), u^{\pm}(t)\Big](s) \tag{5.2}$$

is an element of $C^1([0, T] \times \mathrm{S})$.

Proof. By the transformation $z = \sqrt{u}$ we can rewrite the minimization problem (1.30) as

$$\inf_z \int_{\mathrm{S}} \Big[\frac{w^2}{2z(s)^2} + 2z'(s)^2\Big] \,\mathrm{d}s \ : \quad z(\pm\kappa) = \sqrt{u^{\pm}}.$$

The corresponding stationarity equation is

$$-z'' - \frac{w^2}{4z^3} = 0, \qquad z(\pm\kappa) = \sqrt{u^{\pm}}, \tag{5.3}$$

which implies that any solution z is concave and therefore $z \geq \min \sqrt{u^{\pm}}$, or $u \geq \min u^{\pm}$.

Since $u^{\pm} > 0$, the existence and uniqueness of the solution u of (5.1), or equivalently of the solution z of (5.3), are classical, and the continuity follows from classical results for the continuous dependence of the solutions of elliptic problems on parameters. Similarly, if $u^{\pm}$ is a function $u^{\pm} \in C^2([0, T])^2$ and bounded away from zero, then the solution $u[\dot{u}^+(\cdot)/2, u^{\pm}(\cdot)]$ is C^1 on its domain (note that one degree of differentiation in time is lost since $\dot{u}^+$ appears as a parameter in the equation for z).

The symmetry and minimality properties of M are immediate. To prove the convexity of M, take $(w_1, u_1^{\pm})$ and $(w_2, u_2^{\pm})$ with $M(w_1, u_1^{\pm})$, $M(w_2, u_2^{\pm}) < \infty$, $\lambda \in [0, 1]$, and let u_1 and u_2 be the corresponding minimizers. Since $(u, w) \mapsto w^2/u$ is convex, it follows that

$$\frac{(\lambda w_2 + (1-\lambda)w_1)^2}{8(\lambda u_2(s) + (1-\lambda)u_1(s))} \leq \lambda \frac{w_2^2}{8u_2(s)} + (1-\lambda)\frac{w_1^2}{8u_1(s)},$$

with a similar inequality for the second term in (5.1). Since $\lambda u_2 + (1-\lambda)u_1$ is admissible for $M(\lambda w_2 + (1-\lambda)w_1, \lambda u_2^{\pm} + (1-\lambda)u_1^{\pm})$, we then have

$$\begin{aligned}
&M(\lambda w_2 + (1-\lambda)w_1, \lambda u_2^{\pm} + (1-\lambda)u_1^{\pm}) \leq \\
&\qquad \leq \int_0^1 \Big[\frac{(\lambda w_2 + (1-\lambda)w_1)^2}{8(\lambda u_2(s) + (1-\lambda)u_1(s)} + \frac{(\lambda u_2' + (1-\lambda)u_1')^2}{2(\lambda u_2(s) + (1-\lambda)u_1(s))}\Big] \, ds \\
&\qquad \leq \lambda \int_0^1 \Big[\frac{w_2^2}{8u_2(s)} + \frac{u_2'(s)^2}{2u_2(s)}\Big] \, ds + (1-\lambda) \int_0^1 \Big[\frac{w_1^2}{8u_1(s)} + \frac{u_1'(s)^2}{2u_1(s)}\Big] \, ds \\
&\qquad = \lambda M(w_2, u_2^{\pm}) + (1-\lambda) M(w_1, u_1^{\pm}).
\end{aligned}$$

This concludes the proof of Theorem 5.1. □

As indicated at the end of the introduction, we can find good lower and upper bounds on the integrand M, which are given in the following theorem.

Theorem 5.2. *For all $u^\pm > 0$ and all $w \in \mathbb{R}$ we have the estimate*

$$w(\log u^+ - \log u^-) \le M(w; u^\pm) \le \frac{\log u^+ - \log u^-}{4\kappa(u^+ - u^-)}\big(4\kappa^2 w^2 + (u^+ - u^-)^2\big), \tag{5.4}$$

where both inequalities are equalities if and only if $w = (u^+ - u^-)/2\kappa$. In this case the minimizer u in the definition (5.1) *of $M(w; u^\pm)$ is the affine interpolation $u(s) = (\kappa + s)u^+/2\kappa + (\kappa - s)u^-/2\kappa$.*

Remark 5.3. Note that the left-hand side of (5.4) can be interpreted as $2\langle \mathcal{E}'(u^\pm), w\rangle$ (see Section 1.7), implying that $M(w; u^\pm) \ge 2\langle \mathcal{E}'(u^\pm), w\rangle$ and therefore $\mathcal{J}_0(u^\pm) \ge 0$ for all $u^\pm$. □

Proof. We define the functional $J(w; u) = \frac{1}{2}\int_{\mathrm{S}} \frac{1}{u}(w^2 + u'^2)\, ds$ such that M is obtained by minimizing $J(w; u)$ over all u satisfying the boundary conditions $u(\pm\kappa) = u^\pm$.

The lower estimate follows by neglecting the nonnegative term in

$$J(w; u) \;=\; \int_{\mathrm{S}} \frac{1}{2u}\big(w - u'\big)^2 + w\int_{\mathrm{S}} \frac{u'}{u} \;\ge\; w(\log u(\kappa) - \log u(-\kappa))$$

and using the boundary conditions. We also see that equality holds if and only if $u' \equiv v/2$, which implies $v = 4(m{-}a)$.

The upper estimate is obtained by testing with the affine function $u(s) = (\kappa + s)u^+/2\kappa + (\kappa - s)u^-/2\kappa$. Obviously, the lower estimate and the upper estimate coincide for $w = (u^+ - u^-)/2\kappa$. Hence, the result is proved. □

The fact that optimality occurs at affine functions also gives a characterization of the limit $\hat{u}_0$ of a sequence of *solutions* $\hat{u}_\varepsilon$:

Theorem 5.4. *Let ρ_ε be a sequence of solutions of* (1.1) *such that $\mathcal{E}_\varepsilon(\rho_\varepsilon(0))$ converges as $\varepsilon \to 0$. Then the assertions of Theorem 3.2 hold, and in addition $\hat{u}_0$ is affine in s:*

$$\text{for almost all } t, s, \qquad \hat{u}_0(t, s) = \frac{\kappa + s}{2\kappa}\hat{u}_0(t, \kappa) + \frac{\kappa - s}{2\kappa}\hat{u}_0(t, -\kappa).$$

Proof. The transformed solutions $\hat{u}_\varepsilon$ satisfy the equation

$$\hat{g}_\varepsilon \partial_t \hat{u}_\varepsilon = \partial_{ss}\hat{u}_\varepsilon.$$

The density $\hat{g}_\varepsilon$ concentrates on to the boundary points $s = \pm\kappa$, implying that in the interior of the interval S the equation formally reduces to $0 = \partial_{ss}\hat{u}_\varepsilon$. Using classical methods for partial differential equations one can convert this observation into a proof that the limit $\hat{u}_0$ is affine for each t.

Instead we prefer to stay within the realm of the gradient-flow structure. Since the ρ_ε are solutions, $\mathcal{A}_\varepsilon(\rho_\varepsilon) = 0$; by Theorem 4.1 and the assumption of convergence of the initial energies $\mathcal{E}_\varepsilon(\rho_\varepsilon(0))$, we have $\mathcal{A}_0(\rho_0) \le 0$. Since $\mathcal{A}_0$ satisfies condition (1.16), $\mathcal{A}_0(\rho_0) = 0$. This implies that $\hat{u}_0$ is a minimizer of M for almost all t, and by Theorem 5.2 it is therefore affine for almost all t. □

6. Recovery sequence

Theorem 6.1 (Recovery sequence)**.** *Let $u^\pm \in AC(0, T; \mathbb{R})$ be such that $\mathcal{J}_0(u^\pm) < \infty$. Then there exists a sequence $\hat{u}_\varepsilon \in C^1([0, T] \times \mathrm{S})$ such that $\hat{u}_\varepsilon(\cdot, \pm\kappa) \to u^\pm$ in $L^1(0, T)$, $\hat{\mathcal{E}}_\varepsilon(\hat{u}_\varepsilon(0)) \to \mathcal{E}_0(u^\pm(0))$, $\hat{\mathcal{E}}_\varepsilon(\hat{u}_\varepsilon(T)) \to \mathcal{E}_0(u^\pm(T))$, and $\hat{\mathcal{J}}_\varepsilon(\hat{u}_\varepsilon) \to \mathcal{J}_0(u^\pm)$.*

Remark 6.2. By this result the sequence $\hat{u}_\varepsilon$ and its other forms u_ε, ρ_ε, and $\hat{\rho}_\varepsilon$ converge in the different senses provided by Theorem 3.1. □

Proof. By a diagonal argument, and using the lower bound (4.1), it is sufficient to prove the following approximation result: given $\delta > 0$, there exists a sequence $(\hat{u}^\delta_\varepsilon)_{\varepsilon>0}$ with $\hat{u}^\delta_\varepsilon \hat{\gamma}_\varepsilon \in C^1([0,T] \times \mathrm{S})$ and

$$\limsup_{\varepsilon\to 0}\ \max\Big\{ \|\hat{u}^\delta_\varepsilon(\cdot, \pm\kappa) - u^\pm\|_{L^1(0,T)^2},\ \big|\hat{\mathcal{E}}_\varepsilon(\hat{u}^\delta_\varepsilon(T)) - \mathcal{E}_0(u^\pm(T))\big|, \big|\hat{\mathcal{E}}_\varepsilon(\hat{u}^\delta_\varepsilon(0)) - \mathcal{E}_0(u^\pm(0))\big| \Big\} \le \delta, \tag{6.1}$$

and

$$\limsup_{\varepsilon\to 0} \hat{\mathcal{J}}_\varepsilon(\hat{u}^\delta_\varepsilon) - \mathcal{J}_0(u^\pm) \le \delta. \tag{6.2}$$

We now prove this approximation result in several steps. First note that $u^\pm \in W^{1,1}(0,T)$, and that the finiteness of $\mathcal{J}_0(u^\pm)$ implies that $u^+ + u^-$ is constant in time (say $2m$) and therefore $0 \le u^\pm \le 2m$ and $\dot{u}^+ = -\dot{u}^-$. To simplify we only specify the value u^- at $-\kappa$, and consider the corresponding value u^+ at $+\kappa$ as defined by the condition of constant mass.

We first approximate u^- by a function that is bounded away from zero and from $2m$. We do this by setting $y^-_\eta := m + (1-\eta)(u^- - m)$, for some small η; as $\eta \to 0$, $y^-_\eta \to u^-$ in $W^{1,1}(0,T)$. The function y^-_η is bounded away from 0 and $2m$; the convexity of M and the fact that it vanishes when $w = 0$ and $u^+ = u^-$ (Theorem 5.1) imply that for almost all t, $M(\dot{y}^+_\eta(t)/2; y^\pm_\eta(t))$ is decreasing in η, and that $M(\dot{y}^+_\eta(t)/2; y^\pm_\eta(t)) \uparrow M(\dot{u}^+(t)/2; u^\pm(t))$ as $\eta \downarrow 0$. This implies that

$$\int_0^T M\Big(\frac{1}{2}\dot{y}_\eta(t)^+; y^\pm_\eta(t)\Big)\,\mathrm{d}t \longrightarrow \int_0^T M\Big(\frac{1}{2}\dot{u}^+(t); u^\pm(t)\Big)\,\mathrm{d}t \qquad \text{as } \eta \to 0.$$

Similarly $\mathcal{E}_0(y^\pm_\eta(T)) \to \mathcal{E}_0(u^\pm(T))$ and $\mathcal{E}_0(y^\pm_\eta(0)) \to \mathcal{E}_0(u^\pm(0))$, implying that for given $\delta > 0$ we may choose $\eta > 0$ such that

$$\max\Big\{ \|y^\pm_\eta - u^\pm\|_{L^1(0,T)^2},\ \big|\mathcal{E}_0(y^\pm_\eta(T)) - \mathcal{E}_0(u^\pm(T))\big|, \\ \big|\mathcal{E}_0(y^\pm_\eta(0)) - \mathcal{E}_0(u^\pm(0))\big|,\ \big|\mathcal{J}_0(y^\pm_\eta) - \mathcal{J}_0(u^\pm)\big| \Big\} \le \frac{\delta}{4}. \tag{6.3}$$

We fix this number η.

The next step is to smoothen y_η. We approximate y_η in $W^{1,1}(0,T)$ by convolution to give a $\widetilde{y} \in C^2([0,T])$, while preserving the pointwise upper and lower bounds. Because M is convex, it follows that

$$\mathcal{J}_0(\widetilde{y}^\pm) \le \mathcal{J}_0(y^\pm_\eta), \tag{6.4}$$

and we can choose $\widetilde{y}$ such that

$$\max\Big\{ \|\widetilde{y}^\pm - y^\pm_\eta\|_{L^1(0,T)^2},\ \big|\mathcal{E}_0(\widetilde{y}^\pm(T)) - \mathcal{E}_0(y^\pm_\eta(T))\big|,\ \big|\mathcal{E}_0(\widetilde{y}^\pm(0)) - \mathcal{E}_0(y^\pm_\eta(0))\big|, \Big\} \le \frac{\delta}{4}. \tag{6.5}$$

We now interpolate $\widetilde{y}$ by Theorem 5.1 without changing notation; note that then $\widetilde{y} \in C^1([0,T] \times \mathrm{S})$. Since $\widetilde{y}$ is fixed and C^1, it follows that as $\varepsilon \to 0$, the corresponding function $\hat{w}_\varepsilon$, defined by $\partial_t(\widetilde{y}\hat{\gamma}_\varepsilon) + \partial s \hat{w}_\varepsilon = 0$, is uniformly bounded and satisfies

$$\forall s \in [-\kappa, \kappa), \qquad \hat{w}_\varepsilon(t,s) = \int_{-\kappa}^s \dot{\widetilde{y}}(t,\sigma)\,\hat{\gamma}_\varepsilon(\mathrm{d}\sigma) \to \frac{1}{2}\dot{\widetilde{y}}(t,-\kappa) = \frac{1}{2}\dot{\widetilde{y}}^-(t).$$

Therefore

$$\lim_{\varepsilon\to 0}\int_0^T\int_{\mathrm{S}}\frac{\hat w_\varepsilon^2}{\widetilde y}\,\mathrm{d}s\mathrm{d}t=\int_0^T\int_{\mathrm{S}}\frac{w^2}{\widetilde y}\,\mathrm{d}s\mathrm{d}t\qquad\text{with } w(t)=\frac12\dot{\widetilde y}^{\,-}(t),$$

and we have for sufficiently small $\varepsilon>0$ that

$$|\hat{\mathcal{J}}_\varepsilon(\widetilde y)-\mathcal{J}_0(\widetilde y^{\pm})|\le\frac{\delta}{4}. \tag{6.6}$$

Similarly, since $\hat\gamma_\varepsilon\overset{*}{\rightharpoonup}\hat\gamma_0$, we have for sufficiently small $\varepsilon>0$ that

$$\max\Big\{\,\big|\hat{\mathcal{E}}_\varepsilon(\widetilde y(T))-\mathcal{E}_0(\widetilde y^{\pm}(T))\big|,\ \big|\hat{\mathcal{E}}_\varepsilon(\widetilde y(0))-\mathcal{E}_0(\widetilde y^{\pm}(0))\big|\,\Big\}\le\frac{\delta}{4}. \tag{6.7}$$

The final step is to approximate $\tilde y$ by a function of the right mass. Since $\widetilde y\in C^2([0,T]\times\mathrm{S})$, the mass discrepancy

$$\widetilde m(t):=\int_{\mathrm{S}}\widetilde y(t,s)\,\hat\gamma_\varepsilon(\mathrm{d}s)-\int_{\mathrm{S}}\widetilde y(t,s)\,\hat\gamma_0(\mathrm{d}s)$$

converges to zero uniformly on $[0,T]$. Setting

$$\hat u_\varepsilon^\delta(t,s):=\widetilde y(t,s)-\widetilde m(t),$$

we find that for sufficiently small ε

$$\begin{aligned}\max\Big\{\,&\|\hat u_\varepsilon^\delta(\cdot,\pm\kappa)-\widetilde y^{\pm}\|_{L^1(0,T)^2},\ \big|\hat{\mathcal{E}}_\varepsilon(\hat u_\varepsilon^\delta(T))-\hat{\mathcal{E}}_\varepsilon(\widetilde y(T))\big|,\\ &\big|\hat{\mathcal{E}}_\varepsilon(\hat u_\varepsilon^\delta(0))-\hat{\mathcal{E}}_\varepsilon(\widetilde y(0))\big|,\ \big|\hat{\mathcal{J}}_\varepsilon(\hat u_\varepsilon^\delta)-\hat{\mathcal{J}}_\varepsilon(\widetilde y)\big|\,\Big\}\le\frac{\delta}{4}.\end{aligned} \tag{6.8}$$

The claims (6.1) and (6.2) then follow from combining the estimates (6.3), (6.4), (6.5), (6.6), (6.7), and (6.8). □

7. Connections with stochastic particle systems

The mathematical results of this paper make important use of Definition 1.1 of a gradient-flow solution. This formulation is more than a mathematical convenience: it arises naturally when considering equation (1.1) as the deterministic limit of a stochastic system of particles. We now explain this connection and its consequences.

Fix ε for the moment. Consider a collection of n independent particles, each of which performs a Brownian motion in a potential landscape given by the energy function H/ε. Equation (1.1) is the *continuum* or *hydrodynamic limit* of this system of particles, as the number n of particles tends to infinity. One way of describing this limit is by considering the *empirical measure*

$$L_n:[0,T]\to\mathcal{M}([-1,1]),\qquad L_n(t):=\frac1n\sum_{i=1}^n\delta_{\xi_i(t)},$$

where $\xi_i(t)$ is the position in $[-1,1]$ of particle i at time t. As $n\to\infty$, with probability 1 this empirical measure converges weakly to a limit measure $\rho(t)$ at every time t, and this limit measure solves the equation (1.1).[2]

[2] We deliberately disregard the role of initial conditions for the moment; one could, for instance, choose a single delta function δ_a as the initial datum for (1.1), and have all particles start at $\xi=a$.

Given this connection, a *large-deviation* result characterizes the probability of finding the empirical measure $L_n(t)$ *far* from the solution ρ of (1.1). Such a result roughly takes the form

$$P_n(L_n \approx \tilde{\rho}) \sim \exp(-nI(\tilde{\rho})), \tag{7.1}$$

in terms of a *rate functional* I. A rigorous version of this statement can be found, for instance, in [FK06, Th. 13.37].

The surprising feature, however, is that for this system of particles, the rate functional I above is *exactly* equal to the functional $\mathcal{A}_\varepsilon$ in (1.25) (see e.g. [KO90] or [FK06, Th. 13.37]). This feature has several consequences.

1 Since $\mathcal{A}_\varepsilon = I$, the large-deviation result (7.1) gives an alternative explanation why $\mathcal{A}_\varepsilon \geq 0$ and why $\mathcal{A}_\varepsilon = 0$ implies a solution of the deterministic system. The positivity of I, and therefore of $\mathcal{A}_\varepsilon$, arises directly from the property (7.1) and the fact that probabilities are bounded by 1; and since the hydrodynamic limit is assumed with probability 1, the solution ρ of the limit equation (1.1) necessarily satisfies $\mathcal{A}_\varepsilon(\rho) = I(\rho) = 0$.[3]

2 In Section 1.2 we mentioned that there exist at least two different gradient-flow structures for equation (1.1). The fact that one of these structures arises in the large-deviation description of this stochastic system, may be interpreted to signify that this gradient-flow structure is more 'natural'—at least when we view (1.1) as arising from this specific stochastic particle system. Of course, there may well be a different stochastic system whose large-deviation behaviour is related to the structure (1.18), and there may be other arguments that favour other structures.

3 This connection provides an answer to the question, often heard, 'why does the Wasserstein metric figure in this gradient-flow structure?', since the Wasserstein dissipation arises directly from the large-deviation behaviour. However, a complete answer requires describing the large-deviation result in some detail, which would take us too far; see [ADPZ10] for a detailed discussion.

4 In the context of a large-deviation result, it is natural to consider sets of the form $\{\tilde{\rho} : I(\tilde{\rho}) \leq \delta\}$ for $\delta > 0$; these correspond to collections of 'least unlikely' states, in the sense that their probability vanishes no faster than $\mathrm{e}^{-n\delta}$. Again, this ties in with the results proved above, in which we do not assume $\mathcal{A}_\varepsilon = 0$, but only boundedness of $\mathcal{E}_\varepsilon(\rho_\varepsilon(0))$ and $\mathcal{J}_\varepsilon(\rho_\varepsilon)$.

The connection with large-deviation principles also explains the structure of M in (1.30). The well-known *contraction principle* describes how rate functions transform under projection, i.e. under loss of information. Suppose that I is a rate function describing the behaviour of a sequence of probability measures P_n on a space X, in the sense of (7.1). Let $p : X \to Y$ be a continuous map, and $Q_n := P_n \circ p^{-1}$ the corresponding probability measures describing the behaviour of the system after projection under p onto Y. Then Q_n satisfies the large-deviation principle [dH00, Th. III.20]

$$Q_n(y) \sim \exp(-nI_p(y)) \qquad \text{with} \qquad I_p(y) := \inf_{x \in X : p(x) = y} I(x).$$

The form of the function M can be understood in terms of this contraction principle. In the limit $\varepsilon = 0$, the only information about u_ε or $\hat{u}_\varepsilon$ that survives are the boundary values. Consequently the large-deviation behaviour of the system in the limit follows from the contraction principle by interpolating between the boundary values, in such a way as to minimize the functional over all missing information. The function M is the direct consequence of this.

---

[3] A careful treatment of this argument actually requires a more precise definition of (7.1) and a discussion of topology; we omit both.

8. Discussion

Passing to the limit in gradient flows. The aim of this work is to explore the potential of the Wasserstein gradient-flow structure of (1.1) for rigorous passing to the limit. For this specific system, we have succeeded to a large degree, and we comment below on the specific assumptions that we have made.

The property $\mathcal{A} = 0$ is a reformulation of the concept of a *curve of maximal slope*, which was introduced by DeGiorgi and co-authors (see e.g. [DGMT80]) as a metric-space generalization of a gradient flow. Sandier, Serfaty, and Stefanelli [SS04, Ste08] appear to be the first to explore in detail the use of this structure for passing to the limit. Serfaty [Ser09] discusses the case of metric spaces, with obvious applications for the case of the Wasserstein metric. She leaves aside the question of compactness, however, and one of the main contributions of this paper is to show that appropriate compactness 'in time' can also be obtained from the Wasserstein structure.

A related result is that of Ambrosio, Savaré, and Zambotti [ASZ09], who study entropy-Wasserstein gradient flows in a Hilbert space with respect to a weakly converging sequence of reference measures. Their approach first proves convergence of time-discrete approximations for fixed time, and then uses error bounds to prove convergence of the time-continuous solutions.

Assumptions. The assumptions in the main theorems are the boundedness of the initial energy and of the dissipation function $\mathcal{J}_\varepsilon$, both of which are natural objects in the Wasserstein gradient flow. The relaxation of the condition $\mathcal{A}_\varepsilon = 0$ to the condition $\sup_\varepsilon \mathcal{J}_\varepsilon < \infty$ is a broadening of scope: it implies that the compactness result holds not only for solutions, with their accompanying higher-regularity properties, but for a much wider class of sequences. In addition, this class arises naturally in the context of large deviations for an underlying stochastic particle system (see Section 7).

However, a central tool is the mapping $\xi \mapsto s$, which desingularizes the diffusion term and allows for a more detailed study of the limit behaviour. This mapping is very specific for this problem, and it is an interesting question how to generalize it to singular systems described by different PDEs (e.g. higher-order [Ott98, GO01, Gla03] or nonlocal parabolic equations [CMV03, CMV06]) or more complicated geometric spatial structure.

Weak formulations and compactness. A related question arose during the work presented here: can our definition of a gradient-flow solution, Definition 1.1, be viewed as a *weak form* of the gradient-flow equation (1.8),

$$\dot{z} = -\nabla_G \mathcal{E}(z)? \tag{1.8}$$

The straightforward answer to this question seems to be negative, since traditionally weak formulations serve to reduce regularity requirements, and in both cases the function z is necessarily differentiable. Therefore shifting from (1.8) to Definition 1.1 brings no advantage on that front.

However, we argue here that a different aspect is just as important: the compactness and convergence properties of the formulation. As we have shown, solutions of a sequence of problems are compact in an appropriate way, and a subsequence converges to a limiting object that can be considered an appropriate generalization of a gradient flow (see Definition 1.1, and also the discussion in Section 7).

This ties in with the strongly related work of Herrmann and Niethammer [HN11], that we discuss separately below. One aspect that this paper and [HN11] have in common is the reformulation of a nonlinear, singular differential equation as a parameter-dependent variational problem, thereby opening the door to methods of variational calculus.

Choice of convergence. We prove Gamma-convergence of the sequence of functionals $\mathcal{J}_\varepsilon$. If one is only interested in convergence of solutions, then this is actually too strong: it suffices to prove the lower bound inequality, Theorem 4.1. We prove the recovery sequence, Theorem 6.1 nonetheless, especially since it completes the picture of the convergence on $\mathcal{J}_\varepsilon$.

Incidentally, the fact that Gamma-convergence is a natural form of convergence for large-deviation rate functionals (see Section 7) has also been recognized in the probabilistic literature [Léo07].

The micro-problem. The transformation to the new spatial variable s has the effect of blowing up the region in which the derivative of u_ε is large. The resulting function $\hat{u}_\varepsilon$ is more regular, as reflected by the H^1-bound on $\sqrt{\hat{u}_\varepsilon}$ (see Figure 4 and Remark 2.1). This blow-up argument is reminiscent of the 'cell problem' in homogenization [Hor97] or the 'inner' and 'outer layers' in singular perturbation theory [Ver05]. Similarly, the parallel convergence results of those two theorems reflect these separate behaviours at two different scales.

It is interesting to note, however, that for the lower-bound inequality one does not need to know much about the function M; actually, only its definition. Some additional information (the inequalities of Theorem 5.2) is necessary to identify solutions of $\mathcal{A}_0 = 0$ as solutions of a corresponding differential equation. Other additional information (Theorem 5.1) is necessary for the recovery sequence, Theorem 6.1.

9. Comparison with Herrmann-Niethammer [HN11]

This is not the first paper to give an answer to the question that was raised in [PSV10], *Can we prove convergence using the Wasserstein gradient flow?* In [HN11], Herrmann and Niethammer give a different (but again affirmative) answer. Here we briefly describe their approach and comment on the differences.

The authors of [HN11] build upon a solution concept for gradient flows based on an integrated form of the Rayleigh principle. This concept has been used before in nearly-finite-dimensional situations [NO01, NO10], but its application in a truly parabolic context appears to be new. We describe it here in the case of a linear space $\mathcal{Z}$; the generalization to a manifold is straightforward. Given an energy functional $\mathcal{E}$ on $\mathcal{Z}$ and a Riemannian metric G (see the introduction), it is a straightforward observation that if z is a solution of the gradient-flow equation (1.8), then its time derivative $\dot{z}(t)$ at time t is a minimizer of

$$v \mapsto \langle v, G(z(t))v\rangle + \langle \mathrm{D}\mathcal{E}(z(t)), v\rangle. \tag{9.1}$$

Inspired by this, the authors of [HN11] define an *integrated Rayleigh principle* as follows: an absolutely continuous function $z : [0,T] \to \mathcal{Z}$ satisfies this principle if its time derivative $\dot{z}$ minimizes the functional

$$\int_0^T \Big(\frac{1}{2}\langle v(t), G(z(t))v(t)\rangle + \big\langle \mathrm{D}\mathcal{E}(z(t)), v(t)\big\rangle\Big)\,\mathrm{d}t, \tag{9.2}$$

among all $v : [0,T] \to \mathcal{Z}$.

In [HN11], the authors first remark that at finite $\varepsilon > 0$, the solution u_ε of (1.7) is a minimizer of this integrated Rayleigh principle. In addition, the *a priori* estimates of [PSV10] provide appropriate compactness of the sequence u_ε. The central result is then that the integrated Rayleigh principle for the limiting function u_0 can be derived from the same principle for the solutions u_ε.

The work by Herrmann and Niethammer is interesting for various reasons. First, this solution concept merits to be considered more closely, and we make some comments on this below. Next, the authors themselves state as a drawback that the compactness that they use does not derive from the Wasserstein gradient-flow structure, but from the linear semi-group structure used in [PSV10]. They pose in turn the question whether the compactness can be derived from the Wasserstein structure. And finally, what is exactly the relationship between the solution concepts of [HN11] and of this paper, and similarly of the convergence theorems of the two papers?

There is a problem with the definition of the 'integrated Rayleigh principle' for a general function z. Take the example of a Hilbert space H and a continuous semigroup generated by a non-negative self-adjoint operator A, which solves the equation $\dot z = Az$ in H. This is a gradient flow with $\langle z, Gy\rangle = (z,y)_H$ and $\mathcal{E}(z) = \frac12 (z, Az)_H$. If $z(t) \notin D(A^{1/2})$ at $t>0$, then $\mathcal{E}(z(t)) = \infty$ and $\mathrm{D}\mathcal{E}(z(t))$ is not well defined. Another way of stating this is that the right-hand side of (9.1) is not bounded from below, and its infimum equals $-\infty$. This in turn implies that even though $\dot z$ might minimize (9.2) *for fixed* $z(\cdot)$, in the neighbourhood of that z there exist perturbations $\widetilde{z}$, arbitrary close to z, for which the infimum equals $-\infty$. Therefore the formulation (9.2) is very unstable under perturbations of z.

It is no coincidence that the expression (9.2) is closely related to the functional $\mathcal{A}$. We can write

$$\begin{aligned}\mathcal{A}(z) &= \mathcal{E}(z(T)) - \mathcal{E}(z(0)) + \int_0^T \left(\frac12 \langle \dot z, G(z)\dot z\rangle + \frac12\Big\langle -\mathrm{D}\mathcal{E}(z), G(z)^{-1}(-\mathrm{D}\mathcal{E}(z))\Big\rangle\right)\mathrm{d}t\\ &= \int_0^T \left(\frac12\langle \dot z, G(z)\dot z\rangle + \big\langle \mathrm{D}\mathcal{E}(z), \dot z\big\rangle\right)\mathrm{d}t + \frac12\int_0^T \Big\langle -\mathrm{D}\mathcal{E}(z), G(z)^{-1}(-\mathrm{D}\mathcal{E}(z))\Big\rangle\,\mathrm{d}t. \qquad (9.3)\end{aligned}$$

This form shows that Definition 1.1 with the structure (1.11) is different from the integrated Rayleigh principle above in two ways: first, the integral $\int_0^T \big\langle \mathrm{D}\mathcal{E}(z(t)), \dot z(t)\big\rangle\,\mathrm{d}t$ has been converted into the end point values $\mathcal{E}(z(T)) - \mathcal{E}(z(0))$, and secondly, the addition of the dual dissipation potential $\psi^*(-\mathrm{D}\mathcal{E}(z))$ (the second term in (9.3)) penalizes 'non-regular' values of $z(t)$.

Both of these changes appear to improve the robustness of the formulation. The addition of the dual potential has the effect of penalizing 'unfavorable' choices for the function z; and the conversion of the cross term into end point values mitigates the effect of fast oscillations. The compactness results of Theorems 3.1 and 3.2 certainly suggest that Definition 1.1 and (1.11) provide a useful basis for the analysis of these more general gradient flows.

Acknowledgements. The research of M. A. Peletier has received funding from the ITN "FIRST" of the Seventh Framework Programme of the European Community (grant agreement number 238702).

References

[ADPZ10] 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. *Arxiv preprint arxiv:1004.4076*, 2010. Accepted for publication in Communications in Mathematical Physics.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, first edition, 2000.

[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, 2005.

[ASZ09] L. Ambrosio, G. Savaré, and L. Zambotti. Existence and stability for Fokker–Planck equations with log-concave reference measure. *Probability theory and related fields*, 145(3):517–564, 2009.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84:375–393, 2000.

[BCC08] A. Blanchet, V. Calvez, and J. A. Carrillo. Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model. *SIAM J. Numer. Anal.*, 46(2):691–721, 2008.

[Bre73] H. Brezis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North Holland, 1973.

[CDF$^+$] J. A. Carrillo, M. DiFrancesco, A. Figalli, T. Laurent, and D. Slepcev. Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations. *Duke Math. J.* To appear.

[CG04] E. A. Carlen and W. Gangbo. Solution of a model Boltzmann equation via steepest descent in the 2-Wasserstein metric. *Arch. Ration. Mech. Anal.*, 172(1):21–64, 2004.

[CMV03] J. A. Carrillo, R. J. McCann, and C. Villani. Kinetic equilibration rates for granular media and related equations: Entropy dissipation and mass transportation estimates. *Revista Matematica Iberoamericana*, 19(3):971–1018, 2003.

[CMV06] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Archive for Rational Mechanics and Analysis*, 179:217–263, 2006.

[DGMT80] E. De Giorgi, A. Marino, and M. Tosques. Problems of evolution in metric spaces and maximal decreasing curve. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8)*, 68(3):180–187, 1980.

[dH00] F. den Hollander. *Large Deviations*. American Mathematical Society, Providence, RI, 2000.

[FK06] J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2006.

[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.

[Gla03] K. Glasner. A diffuse-interface approach to Hele-Shaw flow. *Nonlinearity*, 16(1):49–66, 2003.

[GO01] L. Giacomelli and F. Otto. Variational formulation for the lubrication approximation of the Hele-Shaw flow. *Calculus of Variations and Partial Differential Equations*, 13(3):377–403, 2001.

[GST09] U. Gianazza, G. Savaré, and G. Toscani. The Wasserstein gradient flow of the Fisher information and the quantum drift-diffusion equation. *Arch. Ration. Mech. Anal.*, 194(1):133–220, 2009.

[HN11] M. Herrmann and B. Niethammer. Kramers' formula for chemical reactions in the context of Wasserstein gradient flows. *Communications in Mathematical Sciences*, 9(2):623–635, 2011.

[Hor97] U. Hornung. *Homogenization and Porous Media*. Springer Verlag, 1997.

[JKO97] R. Jordan, D. Kinderlehrer, and F. Otto. Free energy and the Fokker-Planck equation. *Physica D: Nonlinear Phenomena*, 107(2-4):265–271, 1997.

[KO90] C. Kipnis and S. Olla. Large deviations from the hydrodynamical limit for a system of independent Brownian particles. *Stochastics and stochastics reports*, 33(1-2):17–25, 1990.

[Kra40] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[Léo07] C. Léonard. A large deviation approach to optimal transport. *Arxiv preprint arXiv:0710.1461*, 2007.

[MMS09] D. Matthes, R. J. McCann, and G. Savaré. A family of nonlinear fourth order equations of gradient flow type. *Comm. Partial Differential Equations*, 34(10-12):1352–1397, 2009.

[MRS08] A. Mielke, T. Roubíček, and U. Stefanelli. Γ-limits and relaxations for rate-independent evolutionary problems. *Calculus of Variations and Partial Differential Equations*, 31(3):387–416, 2008.

[MRS09] A. Mielke, R. Rossi, and G. Savaré. Modeling solutions with jumps for rate-independent systems on metric spaces. *Discrete and Continuous Dynamical Systems A*, 25(2), 2009.

[MS09] A. Mielke and U. Stefanelli. Weighted energy-dissipation functionals for gradient flows. *ESAIM Control Optim. Calc. Var.*, 2009. Publ. online. WIAS preprint 1400.

[NO01] B. Niethammer and F. Otto. Ostwald ripening: The screening length revisited. *Calculus of Variations and Partial Differential Equations*, 13(1):33–68, 2001.

[NO10] B. Niethammer and Y. Oshita. A rigorous derivation of mean-field models for diblock copolymer melts. *Calculus of Variations and Partial Differential Equations*, 2010. To appear.

[Ott98] F. Otto. Lubrication approximation with prescribed nonzero contact angle. *Communications in Partial Differential Equations*, 23(11):63–103, 1998.

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

[PSV10] M. A. Peletier, G. Savaré, and M. Veneroni. From diffusion to reaction via Gamma-convergence. *SIAM Journal on Mathematical Analysis*, 42(4):1805–1825, 2010.

[Ris84] H. Risken. The Fokker-Planck Equation, 1984.

[Sav07] G. Savaré. Gradient flows and diffusion semigroups in metric spaces under lower curvature bounds. *C. R. Math. Acad. Sci. Paris*, 345(3):151–154, 2007.

[Ser09] S. Serfaty. Gamma-convergence of gradient flows on Hilbert and metric spaces and applications. *http://www.math.nyu.edu/faculty/serfaty/gcv-erice2.pdf*, 2009.

[SS04] E. Sandier and S. Serfaty. Gamma-convergence of gradient flows with applications to Ginzburg-Landau. *Communications on Pure and Applied Mathematics*, 57(12):1627–1672, 2004.

[Ste08] U. Stefanelli. The Brezis–Ekeland principle for doubly nonlinear equations. *SIAM Journal on Control and Optimization*, 47:1615, 2008.

[Ver05] F. Verhulst. *Methods and Applications of Singular Perturbations: Boundary Layers and Multiple Timescale Dynamics*. Springer Verlag, 2005.

[Vil03] C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.