Continuum of coupled Wasserstein gradient flows

Clément Cancès, Daniel Matthes, Ismael Medina, Bernhard Schmitzer

CONTINUUM OF COUPLED WASSERSTEIN GRADIENT FLOWS

CLÉMENT CANCÈS, DANIEL MATTHES, ISMAEL MEDINA, AND BERNHARD SCHMITZER

Abstract. We study a system of drift-diffusion PDEs for a potentially infinite number of incompressible phases, subject to a joint pointwise volume constraint. Our analysis is based on the interpretation as a collection of coupled Wasserstein gradient flows or, equivalently, as a gradient flow in the space of couplings under a 'fibered' Wasserstein distance. We prove existence of weak solutions, long-time asymptotics, and stability with respect to the mass distribution of the phases, including the discrete to continuous limit. A key step is to establish convergence of the product of pressure gradient and density, jointly over the infinite number of phases. The underlying energy functional is the objective of entropy regularized optimal transport, which allows us to interpret the model as the relaxation of the classical Angenent–Haker–Tannenbaum (AHT) scheme to the entropic setting. However, in contrast to the AHT scheme's lack of convergence guarantees, the relaxed scheme is unconditionally convergent. We conclude with numerical illustrations of the main results.

1. Introduction

1.1. **Problem setup.** Let $X \subset \mathbb{R}^d$ be the closure of a bounded and connected open set with smooth boundary, and let $Y \subset \mathbb{R}^{d'}$ be compact. Let $\mathcal{L}$ denote the Lebesgue measure on X, μ be a $\mathcal{L}$-absolutely continuous probability measure on X and ν an arbitrary probability measure on Y.

We will denote by ρ a probability measure on $X \times Y$, absolutely continuous with respect to $\mathcal{L} \otimes \nu$ and with density r in $L^1(X \times Y, \mathcal{L} \otimes \nu)$. Furthermore, let $V \in C^2(X \times Y)$ be a non-negative function and $\kappa > 0$ be a positive parameter.

Motivated on the one hand by multiphase porous media flows —see for instance [4] for a presentation of the physical models, [18] for some elements of mathematical analysis in the two-phase context, and for [31, 10, 11] for a variational reinterpretation of the multiphase flow problem (see also [30, 12, 14])—, and on the other hand by the classical Angenent-Haker-Tannenbaum (AHT) scheme for L^2 optimal transport [2] — see below for a detailed explanation — we study the following stratified gradient flow:

$$\partial_t r(t,x,y) = \kappa \Delta_X r(t,x,y) + \nabla_X \cdot \big(r(t,x,y)\, \nabla_X [V(x,y) - \Pi(x)] \big), \tag{1}$$

$$0 = \mathbf{n} \cdot \big(\kappa \nabla_X r(t,x,y) + r(t,x,y)\, \nabla_X [V(x,y) - \Pi(t,x)] \big) \quad \text{for } x \in \partial X, \tag{2}$$

$$\mu(x) = \int_Y r(t,x,y)\, \mathrm{d}\nu(y), \tag{3}$$

where ∇_X and $\nabla_X \cdot$ denote the gradient and the divergence with respect to the variable $x \in X$ only.

The pressure Π in (1) is implicitly given via the *first marginal* constraint (3), where by slight abuse the notation we use the symbol $\mu(x)$ for the (Lebesgue) density of the Lebesgue-absolutely

Date: November 21, 2024.

2020 *Mathematics Subject Classification.* 35K51, 35A15, 35A35 .

Key words and phrases. multiphase porous media flow, fibered Wasserstein gradient flow, minimizing movement scheme, Angenent—Haker—Tannenbaum scheme, entropic optimal transport.

continuous measure μ. Note that the mass at level $y \in Y$, that we denote *second marginal*, is constant in time thanks to the divergence structure of (1) and the no-flux boundary condition (2). Assuming the initial condition has second marginal ν, it will be subsequently preserved over time.

Formally, (1)–(3) is a y-wise Wasserstein gradient flow for the potential energy

$$\mathcal{E}(\rho) = \int_{X\times Y} V(x,y)\,\mathrm{d}\rho(x,y) + \kappa\mathcal{H}(\rho) + \iota_{\Gamma(\mu,\nu)}(\rho), \tag{4}$$

where $\mathcal{H}$ denotes the weighted negative entropy

$$\mathcal{H}(\rho \mid \nu) := \int_{X\times Y} (r\log r - r + 1)\,\mathrm{d}x\,\mathrm{d}\nu(y) \tag{5}$$

(we will write $\mathcal{H}(\rho)$ whenever ν is clear from the context) and $\iota_{\Gamma(\mu,\nu)}$ is the indicator function of the feasible couplings (or transport plans) between μ and ν, see (16) below for a rigorous definition.

1.2. **First motivation: multiphasic flows.** (1)–(3) arises naturally as a generalization of multiphase models such as those studied in [11, 13] to the case of a (possibly) infinite number of species. Indeed, Equations (1)–(3) boil down to such models when the space Y is finite, or equivalently if the measure ν is made of a finite number of Dirac masses:

$$\nu = \sum_{i=1}^{k} m_i \delta_{y_i}, \qquad \text{with } (y_i)_{i=1}^k \subset Y \text{ and } \sum_{i=1}^{k} m_i = 1. \tag{6}$$

The measure μ then corresponds to the porosity of the material, i.e. the local ratio of void space available for the fluid in the solid porous structure, within which the multiphase and incompressible fluid is flowing. The function $V(\cdot, y_i)$ then denotes the intrinsic potential energy of the i-th species. Condition (3) is then interpreted as the constraint that the whole porous space is occupied by the fluid mixture.

In our setting, κ is a given positive constant that encodes the intensity of the term $\kappa\mathcal{H}(\rho)$, which could be interpreted as a capillary energy term. The limiting case of vanishing capillarity $\kappa = 0$ is known to be highly complex. Even in the seemingly simple two-phase case, the well-posedness of the problem remains widely open. In the one-dimensional case $d = 1$, the problem reduces to some Burgers' type equation [31]. It is now well understood that in this particularly simple setting, the JKO scheme selects the unique entropy solution in Kruzkov's sense [31, 26]. In larger dimension, tools from convex integration can even be used to construct multiple weak solutions [20, 36, 24]. Moreover when the space dimension d is greater or equal to two, there is no available criterion to charaterize the solution captured by the JKO scheme, and up to our knowledge there is also no general existence theorem unless the external potential is identically equal to zero [37].

Let us expand on the physical interpretation of (1)-(3). Formally, (1) can be written as:

$$\partial_t r_t + \nabla_X \cdot (r_t v_t) = 0, \qquad \text{with} \quad v_t(x,y) := \nabla_X[\Pi_t(x) - V(x,y) - \kappa\log r_t(x,y)]. \tag{7}$$

Integrating (7) with respect to the Y-marginal ν yields:

$$0 = \partial_t\mu = \partial_t\left(\int_Y r_t(x,y)\,\mathrm{d}\nu(y)\right) = -\nabla_X\cdot\left(\int_Y r_t(x,y)v_t(x,y)\,\mathrm{d}\nu(y)\right). \tag{8}$$

Equation (8) shows that the X-projection of the momentum field $\rho_t v_t$ results in an X-divergence-free velocity field, therefore preserving the pointwise X-marginal constraint. In fact, the following decomposition (closely related to the Helmholtz decomposition) holds:

$$-\nabla_X[V(x,y) + \kappa\log r_t(x,y)] = -\nabla\Pi_t(x) + v_t(x,y). \tag{9}$$

The left-hand side corresponds to the steepest descend direction for (4) once one removes the marginal constraint. Thus, the velocity field $v_t(x, y)$ encodes the divergence-free part of the unconstrained velocity field, with $\nabla\Pi_t(x)$ claiming the gradient component. Whenever the left-hand side can be written as a sheer gradient of a function on X, the velocity field v_t becomes identically zero, resulting in a stationary point of (1)-(3). Intuitively, v_t will progressively cancel the curl component of the left-hand side until such a stationary point is attained. This will become clearer when we explore the connections to the AHT scheme [2] in Section 1.3.

For future reference, let us conclude by stating the elliptic equation satisfied by Π_t, which is a direct consequence of (8) plus the marginal constraints:

$$\nabla \cdot [\mu(x)\nabla\Pi_t(x)] = \kappa\Delta\mu(x) + \nabla \cdot \left[\int_Y \nabla_X V(x, y) r(t, x, y)\, \mathrm{d}\nu(y)\right]. \tag{10}$$

1.3. **Second motivation: Angenent-Haker-Tannenbaum scheme.** The field of computational optimal transport offers an alternative interpretation for the evolution equations (1)–(3). From this perspective, the energy (4) is the entropic transport score of the coupling ρ between μ and ν for the cost function V and regularization strength κ. This entropic objective is a common approximation to the classical, unregularized optimal transport problem (corresponding to $\kappa = 0$), which has gained increased popularity thanks to the availability of efficient solvers, most notably the Sinkhorn algorithm [21, 33] and the connection to stochastic optimal control problems [17, 27, 29]. From this perspective, the evolution equations (1)–(3) address the solution of the entropic transport problem by evolving the density of the transport plan in the direction of steepest descent of the entropic score $\mathcal{E}$, conditioned to keeping the marginals feasible.

A similar philosophy is central to the Angenent-Haker-Tannenbaum (AHT) scheme [2, 35], a classical algorithm for unregularized optimal transport (i.e. (4) for $\kappa = 0$) and V strictly convex. In this regime the minimizers of $\mathcal{E}$ are very sparse —when μ has a density, each $x \in X$ sends mass to a single $y \in Y$ [7, 25] — which motivates the parametrization of the current plan ρ as a *transport map*. Hence, the AHT scheme considers a minimizing flow for $\mathcal{E}$ in the space of transport maps, progressively removing the curl component of an initial transport map T_0 until only the gradient component remains. Formally, for $T_0 : X \to Y$ verifying $\nu = T_{0\sharp}\mu$, the AHT evolution is driven by:

$$T_t = T_0 \circ s_t^{-1} \qquad \text{with} \qquad s_0 = \mathrm{id}_X, \qquad \partial_t s_t = v_t \circ s_t, \qquad \nabla \cdot (\mu v_t) = 0 \tag{11}$$

where v_t is a velocity field whose divergence constraint ensures $(s_t)_\sharp\mu = \mu$, and therefore —if such s_t exists— T_t stays a feasible transport map between μ and ν at all times $t \geq 0$. v_t is chosen among the feasible velocity fields to realize the steepest descent for the transport cost

$$E(T) = \int_X V(x, T(x))\mu(x)\, \mathrm{d}x, \qquad \partial_t E(T_t) = \int_X \langle \nabla_X V(x, T_t(x)), v_t(x)\rangle \mu(x)\, \mathrm{d}x$$

with respect to some metric of choice. In [2] the velocity field is chosen as $v_t(x) = u_t(x)/\mu(x)$, where $u_t(x)$ is the $L^2(X)$ projection of $x \mapsto -\nabla_X V(x, T_t(x))$ onto the space of divergence-free velocity fields. If one chooses instead the $L^2(X, \mu)$-metric, the steepest descent is given by setting v_t to the projection of the vector field $x \mapsto -\nabla_X V(x, T_t(x))$ onto the subspace satisfying $\nabla \cdot (\mu v_t) = 0$ in $L^2(X, \mu)$. This projection is given by

$$v_t(x) := \nabla\Pi(x) - \nabla_X V(x, T_t(x)), \quad \text{with } \nabla \cdot [\mu(x)\nabla\Pi(x)] = \nabla \cdot [\mu(x)\nabla_X V(x, T_t(x))] \tag{12}$$

Leaving aside the (challenging) problem of existence of solutions to (11)-(12), the similarities between (7) and (12) are striking. Indeed, Π satisfies both in (10) and in (12) the same elliptic equation, once we (formally) allow in the former the choice $\kappa = 0$ and that ρ_t is given by a Monge

plan $\rho_t = (\mathrm{id}, T_t(x))_\sharp \mu$. Moreover, the AHT velocity field $v_t(x) = \nabla\Pi(x) - \nabla_X V(x, T_t(x))$ is also analogous to that in (7), once we consider that in the AHT only points of the form $(x, T_t(x))$ carry mass, and therefore the velocity field can be given as a function of x instead of (x, y). For these reasons one may additionally interpret (1)–(3) as a relaxation of the AHT scheme to entropic optimal transport, which inevitably requires the relaxation from transport maps to transport plans.

Leaving aside the well-posedness issues of the AHT scheme — which largely align with those outlined in Section 1.2 for the limit of vanishing κ in multiphase flows—, the main drawback of the AHT scheme is that it is unable to escape certain suboptimal configurations. Even though, by construction, the score $E(T)$ is non-increasing along solutions of (11)-(12), the stationary points are all the feasible maps T that verify that $x \mapsto \nabla_X V(x, T(x))$ is curl-free. Unfortunately this characterization does not suffice to grant optimality, and may in some cases include not only the best but also the *worst* transport map, as shown by the following example:

$$X = Y = [-1, 1]^d, \quad \mu = \nu = \tfrac{1}{2^d}\mathcal{L}, \quad V(x, y) = |x - y|^2, \quad T(x) := -x, \quad \nabla_X V(x, T(x)) = 4x.$$

In this case the optimal transport map would correspond instead to $T^* : x \mapsto x$, and T thus corresponds to the map *maximizing* the total transport cost. However, since $\nabla_X V(x, T(x))$ is curl-free, the velocity field v_t given by (12) is identically zero, and the trajectory T_t with initial condition $T_0 = T$ remains stationary in an utterly suboptimal configuration.

Except for special cases [8, 35], convergence of the AHT scheme to the optimal transport map thus remains an open question. Up until now, the question of whether a suitable relaxation can make the AHT scheme unconditionally convergent has remained open as well. We expect that regularization by diffusion for $\kappa > 0$ helps the algorithm to escape from such configurations.

1.4. **Outline.**

Existence of weak solutions. We dedicate Section 2 to recall the relevant optimal transport background. Under suitable regularity assumptions, in Section 3 we build an approximating sequence for the gradient flow of the energy $\mathcal{E}$ by means of a minimizing movement scheme and discuss the relevant a-priori estimates. Section 4 shows convergence of such approximating sequences to a continuous limit trajectory ρ with values on $\mathcal{P}(X \times Y)$ and a pressure Π in $L^2_{loc}(\mathbb{R}_{\geq 0}; H^1(X))$. (ρ, Π) satisfy the system of partial differential equations:

$$\int_0^\infty \int_{X\times Y} \partial_t \psi \,\mathrm{d}\rho = \int_0^\infty \int_{X\times Y} \left(\nabla_X \psi \cdot \nabla_X [V - \Pi] - \kappa \Delta_X \psi\right) \mathrm{d}\rho - \int_{X\times Y} \psi(0) \,\mathrm{d}\rho^0 \tag{13}$$

for all $\psi \in C_c^\infty(\mathbb{R}_{\geq 0} \times X \times Y)$ with $\nabla_X \psi \cdot \mathbf{n} \equiv 0$ on ∂X, and

$$\int_X \nabla\xi \cdot \nabla\Pi(t) \,\mu \,\mathrm{d}x = \int_X \nabla\xi \cdot \left[\int_Y \nabla_X V(x, y) r(t, x, y) \,\mathrm{d}\nu(y)\right] \mathrm{d}x - \kappa \int_X \mu \Delta\xi \,\mathrm{d}x \tag{14}$$

for all $\xi \in C^\infty(X)$ with $\nabla\xi \cdot \mathbf{n} = 0$ on ∂X and a.e. $t \in \mathbb{R}_{\geq 0}$, where r is the density of ρ w.r.t $\mathcal{L} \otimes \nu$.

Stability of solutions. Section 5 shows stability of weak solutions under convergence of the marginals μ and ν and of the initial condition ρ^0. This relates the continuous ν limit to previous multiphasic flow studies such as [30, 12, 14], identifying our dynamics as the limit of the multiphasic case when the mass of each phase ν is not a discrete measure supported on a finite set but rather a general probability measure.

Asymptotic convergence to the minimizer of $\mathcal{E}$. In Section 6 we show that weak solutions converge to the unique minimizer of the energy $\mathcal{E}$ as t goes to infinity. The relevance of this result stems

from the non-convexity of $\mathcal{E}$ in the fibered topology of the minimizing movement scheme. We also discuss implications to computational optimal transport.

The regime $\kappa = 0$. In Section 7 we discuss the $\kappa = 0$ regime, and whether the results of Section 6 may apply to this limit. As explained in Section 1.2, existence of solutions remains an open problem even in for the two-phase case, since the only source of regularity (the entropic term) in (1)–(3) is lost. However, from the computational optimal transport perspective this regime is extremely interesting, as it would correspond to a relaxation of the AHT scheme to the space of transport *plans*, while staying in the unregularized transport regime.

We show that, even though we cannot provide existence of solutions in the general case, the $\kappa = 0$ regime shares with the AHT scheme the existence of suboptimal stationary configurations. For instance, for $V(x, y) = |x-y|^2$, all initial conditions of the form $\rho^0 = (\mathrm{id}, \nabla\phi)_\sharp \mu$ with $\phi \in C^2(X)$ turn out to be stationary. The intuition is that, for τ sufficiently small, ρ^0 becomes a stationary point of the minimizing movement scheme, and therefore also a stationary constant weak solution. We conclude that a relaxation to the space of transport plans does not improve the convergence properties of the AHT scheme significantly, and that some kind of regularization is a necessary component for an unconditionally convergent variant of the AHT algorithm.

Numerical simulations. Finally, Section 8 exemplifies the stability and convergence results with numerical simulations of the minimizing movement scheme. Experimentally we observe linear convergence to the energy $\mathcal{E}$, which may open the door to novel numerical methods for the solution of entropic optimal transport in geometric domains.

1.5. **General hypotheses.**

- We assume $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^{d'}$ to be compact sets. X is further assumed to have connected interior and a smooth boundary.
- We will employ the symbol μ for both a probability measure on X and its density with respect to $\mathrm{d}x$; the context should resolve any ambiguity. We will also assume that μ is bounded away from zero and from above, and that it has *finite Fisher information*:

$$\frac{1}{4}\|\nabla \log \mu\|^2_{L^2(X,\mu)} = \int_X |\nabla \sqrt{\mu}|^2 \,\mathrm{d}x < \infty. \tag{15}$$

  Note that the boundedness away from zero and from above of μ makes the $L^2(X)$ and $L^2(X, \mu)$ norms (resp. $H^1(X)$ and $H^1(X, \mu)$ norms) equivalent, so we will often use them interchangeably. Finiteness of these norms for the marginal μ is needed to prove "horizontal" regularity of the solution r in H^1.
- We denote by ν a probability measure on Y.
- We use the symbol ρ to denote a coupling between μ and ν. Its Radon-Nykodym derivative with respect to $\mathcal{L} \otimes \nu$, which (when it exists) is an element of $L^1(X \times Y, \mathcal{L} \otimes \nu)$, is denoted by r. We will use Lebesgue as the X-reference measure instead of μ because it simplifies the change of variables formula. Integration of a measurable function f on X with respect to the Lebesgue measure is simply denoted by $\int_X f \,\mathrm{d}x$.
- V will be a positive $C^2(X \times Y)$ function, and κ will represent a positive parameter.
- Finally, we assume that the initial condition ρ^0 has finite energy $\mathcal{E}(\rho^0)$.

2. Preliminaries

In the following chapters we will leverage the theory of optimal transportation to build an approximation sequence for the solutions of (1)–(3). For a thorough exposition of optimal transport we refer to [35, 38].

For $Z \subset \mathbb{R}^d$ a compact set, denote by $\mathcal{M}(Z)$ the set of Radon measures on Z, by $\mathcal{M}_+(Z)$ the set of positive Radon measures on Z and by $\mathcal{P}(Z)$ the set of probability measures on Z. For $Z_1 \subset \mathbb{R}^d, Z_2 \subset \mathbb{R}^{d'}$ compact sets and $\rho_1 \in \mathcal{P}(Z_1), \rho_2 \in \mathcal{P}(Z_2)$ two probability measures, $\Gamma(\rho_1, \rho_2)$ denotes the set of *couplings* or *transport plans* between ρ_1 and ρ_2:

$$\Gamma(\rho_1, \rho_2) := \{\gamma \in \mathcal{P}(Z_1 \times Z_2) \;\mid\; \mathrm{P}_1\gamma = \rho_1, \quad \mathrm{P}_2\gamma = \rho_2\} \tag{16}$$

where the maps P_1 and P_2 denote the projections of measures on $Z_1 \times Z_2$ to its marginals, i.e.

$$(\mathrm{P}_1\gamma)(S_1) := \gamma(S_1 \times Z_2) \qquad \text{and} \qquad (\mathrm{P}_2\gamma)(S_2) := \gamma(Z_1 \times S_2)$$

for $\gamma \in \mathcal{P}(Z_1 \times Z_2)$, $S_1 \subset Z_1$, $S_2 \subset Z_2$ measurable.

The optimal transport problem between ρ_1 and ρ_2 with a continuous cost function $c : Z_1 \times Z_2 \to \mathbb{R}_{\geq 0}$ is then given by

$$\min_{\gamma \in \Gamma(\rho_1, \rho_2)} \int_{Z_1 \times Z_2} c(z_1, z_2) \,\mathrm{d}\gamma(z_1, z_2). \tag{17}$$

For compactly supported marginals and a lower-semicontinuous cost function c, Problem (17) admits a minimizer.

When $d = d'$ and $c(z_1, z_2) = |z_1 - z_2|^p$ with $p \in [1, \infty)$, the minimal score of (17) is (the p-th power of) the L^p *Wasserstein distance* (denoted by $\mathbf{W}_p$) between ρ_1 and ρ_2. With no subscript, $\mathbf{W}$ will be understood to denote the $\mathbf{W}_2$ distance. When $z_2 \mapsto c(z_1, z_2)$ is strictly convex for all $z_1 \in Z_1$ (as is the case for the L^p distance with $p > 1$) and ρ_1 is Lebesgue-absolutely continuous, the unique minimizer of (17) is of the form $\gamma = (\mathrm{id}, T)_\sharp \rho_1$ [7, 25], and T solves the so-called *Monge problem*:

$$\min_{T_\sharp \rho_1 = \rho_2} \int_{Z1} c(z, T(z)) \,\mathrm{d}\rho_1(z). \tag{18}$$

The optimal transport problem (17) admits the dual formulation:

$$\sup_{\Phi(z_1) + \Psi(z_2) \leq c(z_1, z_2)} \int_{Z_1} \Phi(z_1) \,\mathrm{d}\rho_1(z_1) + \int_{Z_2} \Psi(z_2) \,\mathrm{d}\rho_2(z_2) \tag{19}$$

where the supremum runs over continuous functions from Z_1 (resp. Z_2) to $\mathbb{R}$. If c is uniformly continuous maximizers for (19) exist.

When $Z_1 = Z_2 =: Z$, the L^p-Wasserstein distance for $p = 1$ takes a particularly convenient form:

$$\mathbf{W}_1(\rho_1, \rho_2) = \sup_{\mathrm{Lip}(\Phi) \leq 1} \int_Z \Phi(z) \,\mathrm{d}(\rho_1 - \rho_2)(z), \tag{20}$$

which is known as the *Kantorovich-Rubinstein formula*. On the other hand, for $p = 2$ it is customary to rewrite the dual relation (19) as:

$$\frac{1}{2}\mathbf{W}(\rho_1, \rho_2)^2 = \sup_{\Phi(z_1) + \Psi(z_2) \leq \frac{1}{2}|z_1 - z_2|^2} \left(\int_{Z_1} \Phi(z_1) \,\mathrm{d}\rho_1(z_1) + \int_{Z_2} \Psi(z_2) \,\mathrm{d}\rho_2(z_2) \right). \tag{21}$$

This re-scaling allows to identify the optimal transport map T as $T = \mathrm{id} - \nabla\Phi$ (ρ_1-almost everywhere). Besides, $\mathbf{W}_1$ and $\mathbf{W}_2$ satisfy an order relation: for γ_1 and γ_2 optimal transport plans between ρ_1 and ρ_2 for $\mathbf{W}_1$ and $\mathbf{W}_2$ respectively, it holds:

$$\begin{aligned}\mathbf{W}_1(\rho_0,\rho_1) &= \int_{Z^2} |z-z'|\,\mathrm{d}\gamma_1(z,z') \leq \int_{Z^2} |z-z'|\,\mathrm{d}\gamma_2(z,z') \\ &\leq \left(\int_{Z^2} |z-z'|^2\,\mathrm{d}\gamma_2(z,z')\right)^{1/2} = \mathbf{W}_2(\rho_0,\rho_1),\end{aligned} \tag{22}$$

where the first inequality comes from the optimality of γ_1 and in the second we use Jensen's inequality.

A notion of 'fiber-wise' Wasserstein distance will be central to our study. Let $Z = X \times Y$ with $X \subset \mathbb{R}^d$ and $Y \subset \mathbb{R}^{d'}$ compacts and let $\nu \in \mathcal{P}(Y)$. For $\rho \in \mathcal{P}(X \times Y)$ with $\mathrm{P}_2\rho = \nu$, define the *disintegration of ρ with respect to its second marginal* as the measurable selection of probability measures $(\rho_y)_{y\in Y}$ satisfying:

$$\int_{X\times Y} \psi(x,y)\,\mathrm{d}\rho(x,y) = \int_Y \left[\int_X \psi(x,y)\,\mathrm{d}\rho_y(x)\right]\,\mathrm{d}\nu(y) \tag{23}$$

for all $\psi \in C(X \times Y)$. ρ_y is uniquely defined ν-almost everywhere. Then for $\rho_1, \rho_2 \in \mathcal{P}(X \times Y)$ with common second marginal, i.e. $\mathrm{P}_2\rho_1 = \mathrm{P}_2\rho_2 = \nu$, define the *fiber-wise Wasserstein distance* $\mathbf{W}_F$ by

$$\mathbf{W}_F(\rho_1,\rho_2)^2 := \int_Y \mathbf{W}(\rho_{1,y},\rho_{2,y})^2\,\mathrm{d}\nu(y) \tag{24}$$

$\mathbf{W}_F$ is a metric by [32], and for our case of study admits the following reformulations:

Lemma 1. *If ρ_1 or ρ_2 is $\mathcal{L}\otimes\nu$-absolutely continuous, the fiber-wise Wasserstein distance $\mathbf{W}_F$ admits the following equivalent formulations:*

$$\mathbf{W}_F(\rho_1,\rho_2)^2 = \inf_{\substack{\gamma\in\mathcal{M}_+(X\times X\times Y)\\ \mathrm{P}_{13}\gamma=\rho_1,\,\mathrm{P}_{23}\gamma=\rho_2}} \int_{X\times X\times Y} |x-x'|^2\,\mathrm{d}\gamma(x,x',y) \tag{25}$$

$$= \inf_{\substack{\tilde\gamma\in\Gamma(\rho_1,\rho_2)\\ y=y'\ \tilde\gamma\text{-a.e.}}} \int_{X\times Y\times X\times Y} |(x,y)-(x',y')|^2\,\mathrm{d}\tilde\gamma(x,y,x',y'), \tag{26}$$

Remark 1. Even though the different order in the variables of γ and $\tilde\gamma$ may appear confusing at first, it answers to the different interpretation we attach to each formulation. Equation (25) should be interpreted in the sense that the $\mathbf{W}$-optimal transport plan in each fiber (that we may denote as γ_y) can be wrapped into a joint measure γ. On the other hand, (26) exposes the relation between $\mathbf{W}_F$ and the Wasserstein distance on $\mathcal{P}(X \times Y)$, being the former a more constrained version of the latter (and therefore yielding a higher score).

Remark 2. We make the density assumption in Lemma 1 for simplicity to avoid issues with measurability. Alternatively, one could employ a measurable selection theorem.

Proof of Lemma 1. For any feasible plan γ in (25) one can define a feasible $\tilde\gamma$ in (26) with the same score as the pushforward of γ by the map:

$$(x,x',y) \mapsto (x,y,x',y). \tag{27}$$

Viceversa, for any feasible $\tilde\gamma$ in (26), its pushforward by

$$(x,y,x',y') \mapsto (x,x',y). \tag{28}$$

yields a feasible plan for (25). For each $y \in Y$ let $\rho_{1,y}$ and $\rho_{2,y}$ denote respectively the y-disintegration of ρ_1 and ρ_2 w.r.t ν. Since either ρ_1 or ρ_2 have a density, the L^2-optimal transport plan between $\rho_{1,y}$ and $\rho_{2,y}$ is unique and a weak* continuous function of the marginals, and hence measurable. Let us denote such optimal plan by γ_y^*. Then one can define a candidate γ^* for (25) by the measure acting as:

$$\int_{X\times X\times Y} \phi(x,x',y)\,\mathrm{d}\gamma^*(x,x',y) := \int_Y \left[\int_{X\times X} \phi(x,x',y)\,\mathrm{d}\gamma_y^*(x,x')\right]\,\mathrm{d}\nu(y). \tag{29}$$

One can easily check that with this choice the right hand side of (25) equals the left hand side; therefore:

$$\mathbf{W}_F(\rho_1,\rho_2)^2 \geq \inf_{\substack{\gamma\,\in\,\mathcal{M}_+(X\times X\times Y)\\ \mathrm{P}_{13}\gamma=\rho_1,\,\mathrm{P}_{23}\gamma=\rho_2}} \int |x-x'|^2\,\mathrm{d}\gamma(x,x',y). \tag{30}$$

For the reversed inequality, note that for any γ feasible for the right hand side of (25), it holds $\mathrm{P}_3\gamma = \nu$, and the disintegration of γ with respect to ν (denoted again by γ_y) provides a feasible transport plan between $\rho_{1,y}$ and $\rho_{2,y}$. This means that, for ν-a.e. $y \in Y$,

$$\mathbf{W}(\rho_{1,y},\rho_{2,y})^2 \leq \int_{X\times X} |x-x'|^2\,\mathrm{d}\gamma_y(x,x'), \tag{31}$$

so that integrating in Y and taking the infimum yields the remaining inequality. □

Unfortunately, the metric $\mathbf{W}_F$ does not turn the space of probability measures on $X \times Y$ into a compact metric space. The reason is that when ν does not have finite support, it is possible to build a sequence of measures on $X \times Y$ that are increasingly oscillating in the horizontal direction, so that the resulting sequence has no Cauchy subsequence (cf. [32, Example 3.15]). As a consequence, the initial step of extracting a cluster point of the discrete trajectories of a gradient flow —which for the $\mathbf{W}$ metric is straighforward —becomes an obstacle in the case of the $\mathbf{W}_F$ metric.

Nevertheless, $\mathbf{W}_F$ bounds the $\mathbf{W}$ metric on the product space $X \times Y$ (see Remark 1 and [32, Proposition 3.10], and therefore we will still be able to extract cluster points with respect to the standard Wasserstein distance on $X \times Y$. Adding this remark to (22) yields:

Lemma 2. *For ρ_1, ρ_2 probability measures on $X \times Y$ with identical Y-marginal it holds:*

$$\mathbf{W}_1(\rho_1,\rho_2) \leq \mathbf{W}_2(\rho_1,\rho_2) \leq \mathbf{W}_F(\rho_1,\rho_2). \tag{32}$$

We will write $\rho \ll \mathcal{L}\otimes\nu$ for a coupling $\rho \in \Gamma(\mu,\nu)$ that is absolutely continuous with respect to the measure $\mathcal{L}\otimes\nu$. This grants the existence of an L^1 density or *Radon-Nykodim derivative*, that we will typically denote by r. The following Lemma collects some useful calculus rules for absolutely continuous couplings:

Lemma 3. *Let $\rho \in \Gamma(\mu,\nu)$ with $\mu \ll \mathrm{d}x$ and $\rho \ll \mathcal{L}\otimes\nu$. Denote by r the density of ρ with respect to $\mathcal{L}\otimes\nu$. Then:*

i) For a.e. $x \in X$ it holds:

$$\int_Y r(x,y)\,\mathrm{d}\nu(y) = \mu(x), \tag{33}$$

and the disintegration of ρ w.r.t. its first marginal at x is given by the measure $\frac{r(x,\cdot)}{\mu(x)}\nu$.

ii) For ν-a.e. $y \in Y$ it holds:

$$\int_X r(x,y)\,\mathrm{d}x = 1, \tag{34}$$

and the disintegration of ρ w.r.t. its second marginal at y has a density given by $r(\cdot, y)$.

Proof. i) For a continuous test function $\xi \in C(X)$,

$$\begin{aligned}\int_X \xi(x)\left[\int_Y r(x,y)\,\mathrm{d}\nu(y)\right]\mathrm{d}x &= \int_{X\times Y} \xi(x) r(x,y)\,\mathrm{d}x\,\mathrm{d}\nu(y) = \int_{X\times Y}\xi(x)\,\mathrm{d}\rho(x,y)\\ &= \int_X \xi(x)\,\mathrm{d}\mu(x) = \int_X \xi(x)\mu(x)\,\mathrm{d}x,\end{aligned}$$

and it is straightforward to verify that the disintegration of ρ with respect to μ at $x \in X$ is given by $\frac{r(x,\cdot)}{\mu(x)}\nu$; indeed, for any $\psi \in C(X\times Y)$,

$$\begin{aligned}\int_X\left[\int_Y \psi(x,y)\frac{r(x,y)}{\mu(x)}\,\mathrm{d}\nu(y)\right]\mathrm{d}\mu(x) &= \int_X\left[\int_Y \psi(x,y)\frac{r(x,y)}{\mu(x)}\,\mathrm{d}\nu(y)\right]\mu(x)\,\mathrm{d}x\\ &= \int_{X\times Y}\psi(x,y) r(x,y)\,\mathrm{d}x\,\mathrm{d}\nu(y) = \int_{X\times Y}\psi(x,y)\,\mathrm{d}\rho(x,y).\end{aligned}$$

An analogous computation yields ii). □

Remark 3. In view of Lemma 3, if $\rho_1, \rho_2 \in \Gamma(\mu,\nu)$ are absolutely continuous w.r.t. $\mathcal{L}\otimes\nu$ with respective densities r_1 and r_2, their fibered Wasserstein distance may be denoted as:

$$\mathbf{W}_F(\rho_1,\rho_2)^2 := \int_Y \mathbf{W}(r_1(\cdot,y), r_2(\cdot,y))^2\,\mathrm{d}\nu(y). \tag{35}$$

The following lemma collects some useful facts about the structure of the minimizers of the energy $\mathcal{E}$; the proof can be found in [6, Prop 2.5, (iii)] and references therein. The assumptions are not minimal, but adequate for our matter of study.

Lemma 4 (Minimizers of $\mathcal{E}$)**.** *Assume that $V \in C(X\times Y)$, and let $\mu \ll \mathcal{L}$. Then:*

(1) $\mathcal{E}$ has a unique minimizer $\rho_* \in \Gamma(\mu,\nu)$.

(2) There exist measurable $\Pi_* : X \to \mathbb{R}$ and $\Psi_* : Y \to \mathbb{R}$ such that

$$\rho_* = \exp\left(\frac{\Pi_* \oplus \Psi_* - V}{\kappa}\right)\mathcal{L}\otimes\nu \tag{36}$$

(3) Conversely, if $\Pi : X \to \mathbb{R}$ and $\Psi : Y \to \mathbb{R}$ are such that $\rho := \exp\left(\frac{\Pi\oplus\Psi - V}{\kappa}\right)\mathcal{L}\otimes\nu$ belongs to $\Gamma(\mu,\nu)$, then ρ is the unique minimizer of $\mathcal{E}$.

During most of this article the marginals μ and ν will be fixed, and therefore the energy (4) and entropy (5) are defined without ambiguity. However, for some results (mainly Lemma 13) we will need to consider sequences $(\mu_n)_n$ and $(\nu_n)_n$ of marginals and respective couplings $\rho_n \in \Gamma(\mu_n,\nu_n)$. In that case it becomes practical to make the dependence of the energy in the marginals explicit, and hence define:

$$\mathcal{E}(\rho \mid \mu,\nu) := \int_{X\times Y} V(x,y)\,\mathrm{d}\rho(x,y) + \kappa\mathcal{H}(\rho\mid\nu) + \iota_{\Gamma(\mu,\nu)}(\rho). \tag{37}$$

The following Lemma will prove useful:

Lemma 5. *The marginal-dependent energy* (37) *is lower-semicontinuous under joint weak* convergence of its arguments.*

Proof. Let $(\mu_n)_n$ be a sequence in $\mathcal{P}(X)$ converging weak* to μ, and analogously $(\nu_n)_n$ be a sequence in $\mathcal{P}(Y)$ converging weak* to ν. Let $\rho_n \in \mathcal{P}(X \times Y)$ for all n converge to $\rho \in \mathcal{P}(X \times Y)$.

$\mathcal{E}(\rho \mid \mu, \nu)$ is a sum of three contributions. The first one is simply integration with respect to a fixed continuous function so it is continuous with respect to weak* convergence of ρ_n. The second entropic contribution is lower-semicontinuous with respect to joint weak* convergence of its first and second argument [1], as is the case here. For the convex indicator function note the following: if $\rho_n \notin \Gamma(\mu_n, \nu_n)$ for all n, then $\mathcal{E}(\rho_n \mid \mu_n, \nu_n) = +\infty$ and there is nothing to prove. If however $\rho_n \in \Gamma(\mu_n, \nu_n)$ along some subsequence, then also $\rho \in \Gamma(\mu, \nu)$ by weak* continuity of the marginal projection operators. Therefore, the last contribution is also jointly weak* lower-semicontinuous with respect to its arguments. □

3. Discrete minimizing movements

We will construct an approximate solution via minimizing movements. To that end, fix a time step $\tau > 0$ and define, for any given $\bar{\rho} \in \Gamma(\mu, \nu)$,

$$\mathcal{E}_\tau\big(\rho \mid \bar{\rho}\big) := \frac{1}{2\tau}\mathbf{W}_F(\rho, \bar{\rho})^2 + \mathcal{E}(\rho). \tag{38}$$

Starting from $\rho_\tau^0 := \rho^0$, define ρ_τ^n for $n = 1, 2, \dots$ inductively as minimizer of the respective $\mathcal{E}_\tau(\cdot \mid \rho_\tau^{n-1})$. We will denote by r_τ^n the density of ρ_τ^n with respect to $\mathcal{L} \otimes \nu$.

Lemma 6. *The minimizing movement scheme* (38) *is well-defined and yields a unique minimizer. Moreover, $r_\tau^n > 0$ for $\mu \otimes \nu$-almost every $(x, y) \in X \times Y$ and each $n \in \mathbb{N}$.*

Proof. We will show existence of minimizers of (38) by constructing an auxiliary convex problem. First let us show that the minimizing movement step (38) can be equivalently written as:

$$\rho^n = \mathrm{P}_{13}\gamma^n, \quad \text{with}$$
$$\gamma^n = \operatorname*{argmin}_{\substack{\gamma \in \mathcal{M}_+(X \times X \times Y) \\ \mathrm{P}_1\gamma = \mu,\, \mathrm{P}_{23}\gamma = \rho^{n-1}}} \int_{X\times X\times Y} \left(\frac{|x - x'|^2}{2\tau} + V(x, y) \right) \mathrm{d}\gamma(x, x', y) + \kappa\mathcal{H}(\mathrm{P}_{13}\gamma). \tag{39}$$

Intuitively, γ represents a collection of transport plans on $(\gamma_y)_y \subset \mathcal{M}_+(X \times X)$ indexed by the variable Y, the constraint $\mathrm{P}_{23}\gamma = \rho^{n-1}$ grants that the second marginal of γ_y is ρ_y^{n-1}, and the constraint $\mathrm{P}_1\gamma$ ensures that the new iterate ρ^n has the right first marginal prescribed by μ; the second marginal is automatically satisfied since no transport happens along the Y-axis and it is therefore inherited from ρ^{n-1}. Rewriting (38) in this form will allow us to use the direct method of calculus of variations on the space $\mathcal{M}_+(X \times X \times Y)$, simplifying the analysis.

The proof of equivalence is very similar to that in Lemma 1. First note that any feasible candidate ρ in (38) must be absolutely continuous with respect to $\mathcal{L} \otimes \nu$, otherwise the entropy term would yield an infinite value. Likewise, any feasible plan γ in (39) must have a $\mathcal{L}\otimes\nu$-absolutely continuous projection $\mathrm{P}_{13}\gamma$. Now, for ρ a feasible candidate for (38), and ν-a.e. $y \in Y$, define γ_y as the $\mathbf{W}$-optimal transport plan between the y-disintegration of ρ and ρ^{n-1} w.r.t ν (denoted respectively by ρ_y and ρ_y^{n-1}). By absolute continuity of ρ, Brenier's theorem grants uniqueness of such optimal transport plan. Then one can define a candidate γ for (39) by the measure acting as:

$$\int_{X\times X\times Y} \phi(x, x', y)\, \mathrm{d}\gamma(x, x', y) := \int_Y \left[\int_{X\times X} \phi(x, x', y)\, \mathrm{d}\gamma_y(x, x') \right] \mathrm{d}\nu(y). \tag{40}$$

One can easily check that the value $\mathcal{E}_\tau(\rho \mid \rho^{n-1})$ is equal to that of the primal objective (39) applied to γ. Conversely, if γ is a feasible candidate for the primal problem (39), its score is necessary larger or equal than $\mathcal{E}_\tau(\mathrm{P}_{13}\gamma \mid \rho^{n-1})$, since the V and $\mathcal{H}$ terms yield the same value, and the transport part is optimal in (38). We conclude that the minimization of $\mathcal{E}_\tau(\cdot \mid \rho^{n-1})$ is equivalent to the convex problem (39) (i.e. they share the same optimal values, and if there are minimizers for one of them we showed how to construct a minimizer for the other).

Next, all the terms constituting the functional in (39) are lower-semicontinuous in γ. By the marginal constraints any admissible γ lies in $\mathcal{P}(X \times X \times Y)$, which is pre-compact under the weak* topology (by compactness of X and Y). Therefore the direct method in the calculus of variations yields existence of a minimizer γ^n. Besides, the strict convexity of the entropy $\mathcal{H}$ guarantees that $\rho^n := \mathrm{P}_{13}\gamma^n$ is unique, which yields a unique minimizer for $\mathcal{E}_\tau(\cdot \mid \rho^{n-1})$.

To show positivity of the newly obtained ρ^n, let $\varepsilon \in (0,1)$ and consider the feasible coupling $\rho_\varepsilon := (1-\varepsilon)\rho^n + \varepsilon\mu \otimes \nu$. We will show that if ρ^n were to vanish in a set $A \subset X \times Y$ with $(\mu \otimes \nu)(A) > 0$ then there would exist a value of $\varepsilon > 0$ such that $\mathcal{E}_\tau(\rho_\varepsilon \mid \rho^{n-1}) < \mathcal{E}_\tau(\rho^n \mid \rho^{n-1})$, which would contradict the optimality of ρ^n. To that end let us bound the upper derivative of $\mathcal{E}_\tau(\rho_\varepsilon \mid \rho^{n-1})$ with respect to ε (denoted by $\overline{D}_\varepsilon \mathcal{E}_\tau(\rho_\varepsilon \mid \rho^{n-1})$) at $\varepsilon = 0$. First, denote by γ (resp. $\gamma_\otimes$) a feasible transport plan between ρ^n (resp. $\mu \otimes \nu$) and ρ^{n-1} for problem (25). Then one can build a $\mathbf{W}_F$-feasible transport plan between ρ_ε and ρ^{n-1} as $\gamma_\varepsilon := (1-\varepsilon)\gamma + \varepsilon\gamma_\otimes$. Consequently:

$$\mathbf{W}_F(\rho_\varepsilon, \rho^{n-1})^2 \leq (1-\varepsilon)\mathbf{W}_F(\rho^n, \rho^{n-1})^2 + \varepsilon(\mathrm{diam}X)^2 \leq \mathbf{W}_F(\rho^n, \rho^{n-1})^2 + \varepsilon(\mathrm{diam}X)^2, \tag{41}$$

so $\overline{D}_\varepsilon \mathbf{W}_F(\rho_\varepsilon, \rho^{n-1})^2\big|_{\varepsilon=0} \leq (\mathrm{diam}X)^2$. For the potential term, by linearity:

$$\int_{X\times Y} V \,\mathrm{d}\rho_\varepsilon = (1-\varepsilon)\int_{X\times Y} V \,\mathrm{d}\rho^n + \varepsilon \int_{X\times Y} V \,\mathrm{d}(\mu\otimes\nu) \leq \int_{X\times Y} V \,\mathrm{d}\rho^n + \varepsilon\|V\|_\infty, \tag{42}$$

and thus $\overline{D}_\varepsilon \langle V, \rho_\varepsilon\rangle\big|_{\varepsilon=0} \leq \|V\|_\infty$. Finally, for the entropic term, let $\mathcal{H}(\rho) = \int_{X\times Y} f(r)\,\mathrm{d}x\,\mathrm{d}\nu(y)$, with $f(s) := s\log s - s + 1$. f is a convex function with derivative $f'(s) = \log s$. Since f is convex, the finite difference $(f(s+h) - f(s))/h$ converges monotonously to $f'(s)$ as $h \to 0$. This allows us to use the monotone convergence theorem in the following calculation

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{H}(\rho_\varepsilon)\big|_{\varepsilon=0} &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(\mathcal{H}(\rho_\varepsilon) - \mathcal{H}(\rho)) = \lim_{\varepsilon\to 0}\int_{X\times Y}\frac{1}{\varepsilon}[f(r^n + \varepsilon(\mu - r^n)) - f(r^n)]\,\mathrm{d}x\,\mathrm{d}\nu(y)\\
&= \int_{X\times Y}\log(r^n(x,y))[\mu(x) - r^n(x,y)]\,\mathrm{d}x\,\mathrm{d}\nu(y)\\
&= \int_{X\times Y}\log(r^n(x,y))\,\mathrm{d}(\mu\otimes\nu)(x,y) - \mathcal{H}(\rho^n) + \int_{X\times Y}(1 - r^n(x,y))\,\mathrm{d}x\,\mathrm{d}\nu(y)\\
&\leq \int_{X\times Y}\log(r^n(x,y))\,\mathrm{d}(\mu\otimes\nu)(x,y) + |X| - 1.
\end{aligned}$$

where the last step holds by the positivity of the entropy. Thus, defing $C = (\mathrm{diam}X)^2/(2\tau) + \|V\|_\infty + |X| - 1 < \infty$ we obtain

$$\overline{D}_\varepsilon \mathcal{E}_\tau(\rho_\varepsilon \mid \rho^{n-1})\big|_{\varepsilon=0} \leq C + \int_{X\times Y}\log(r^n(x,y))\,\mathrm{d}(\mu\otimes\nu)(x,y), \tag{43}$$

which by optimality of ρ^n implies that r^n must be positive $(\mu\otimes\nu)$-almost everywhere. □

Lemma 7. *The iterates satisfy*

$$\mathbf{W}(\rho_\tau^m, \rho_\tau^n) \leq \mathbf{W}_F(\rho_\tau^m, \rho_\tau^n) \leq C\sqrt{\tau\,|m-n|} \tag{44}$$

with $C = \sqrt{2\mathcal{E}(\rho^0)}$.

Proof. (44) follows from the classical JKO energy estimates applied to (38): choosing ρ_τ^{n-1} as a competitor in (38), one gets:

$$\mathbf{W}_F(\rho_\tau^n, \rho_\tau^{n-1})^2 \leq 2\tau(\mathcal{E}(\rho_\tau^{n-1}) - \mathcal{E}(\rho_\tau^n)). \tag{45}$$

W.l.o.g. let $n < m$. Then the triangle inequality for the fiber metric $\mathbf{W}_F$ followed by Jensen's inequality yields:

$$\mathbf{W}_F(\rho_\tau^m, \rho_\tau^n)^2 \leq \left(\sum_{k=n+1}^{m} \mathbf{W}_F(\rho_\tau^k, \rho_\tau^{k-1})\right)^2 \leq (m-n)\sum_{k=n+1}^{m} \mathbf{W}_F(\rho_\tau^k, \rho_\tau^{k-1})^2.$$

Finally, using the energy estimate (45):

$$\leq 2(m-n)\tau(\mathcal{E}(\rho_\tau^n) - \mathcal{E}(\rho_\tau^m)) \leq 2\mathcal{E}(\rho^0)\tau(m-n),$$

and the bound in $\mathbf{W}$ follows from Lemma 2, non-negativity of $\mathcal{E}$, and the finiteness of the initial energy. □

Lemma 8. *For each $n = 1, 2, \ldots$, there is a map $\Pi_\tau^n \in H^1(X)$ such that, for ν-a.e. $y \in Y$,*

$$\rho_{\tau,y}^{n-1} = \left(\mathrm{id}_X + \tau\nabla_X\left[V(\cdot, y) - \Pi_\tau^n + \kappa \log r_\tau^n(\cdot, y)\right]\right)_\sharp \rho_{\tau,y}^n. \tag{46}$$

Π_τ^n can be chosen to have zero mean with respect to μ for all n and τ. Besides, it holds

$$\int_{X\times Y} |\nabla_X(V - \Pi_\tau^n + \kappa \log r_\tau^n)|^2 \,\mathrm{d}\rho_\tau^n = \frac{\mathbf{W}_F(\rho^n, \rho^{n-1})^2}{\tau^2}. \tag{47}$$

In particular, for any $0 \leq k_1 < k_2$,

$$\mathcal{E}(\rho_\tau^{k_1}) + \sum_{n=k_1+1}^{k_2} \tau \int_{X\times Y} |\nabla_X(V - \Pi_\tau^n + \kappa \log r_\tau^n)|^2 \,\mathrm{d}\rho_\tau^n \leq \mathcal{E}(\rho^0). \tag{48}$$

Moreover, ρ_τ^n is bounded away from zero and from above with a uniform bound depending on κ and τ but not on n. Furthermore, for each $N \in \mathbb{N}$:

$$\tau\sum_{n=1}^{N}\left(\kappa^2 \int_{X\times Y} |\nabla_X \sqrt{r_\tau^n}|^2 \,\mathrm{d}x\,\mathrm{d}\nu(y) + \int_X \left|\nabla \Pi_\tau^n\right|^2 \mathrm{d}\mu\right) \leq 4\mathcal{E}(\rho^0) + C\tau N, \tag{49}$$

with C a constant depending only on $\|\nabla_X V\|_\infty$, κ, and the Fisher information of μ.

Proof. In this proof we will drop the subscript τ for better readability.

We extend the method developed in [35]. Let $\Phi : X \times Y \to \mathbb{R}$ be a measurable function such that, for ν-a.e. $y \in Y$, the map $x \mapsto \Phi(x, y)$ is the optimal dual potential in (21) between ρ_y^n and ρ_y^{n-1}, i.e.,

$$(\mathrm{id}_X - \nabla_X \Phi(\cdot, y))_\sharp \rho_y^n = \rho_y^{n-1} \tag{50}$$

for those y. For ν-a.e. $y \in Y$, the density $r^n(\cdot, y)$ is positive a.e. on X by Lemma 6, hence $\Phi(\cdot, y)$ is Lipschitz continuous for those y, with a y-uniform Lipschitz constant given by $\mathrm{diam}(X)$. Further, $\Phi(\cdot, y)$ is unique up to an additive constant, again thanks to positivity of ρ^n and the connectedness

of X; we assume that $\Phi(\cdot, y)$ is normalized to have zero average (that normalization will be changed at the end of the proof). Finally, as outlined in Section 2, optimality implies that

$$\frac{1}{2}\mathbf{W}\big(r^n(\cdot,y), r^{n-1}(\cdot,y)\big)^2 = \frac{1}{2}\int_X \big|\nabla_X \Phi(x,y)\big|^2 r^n(x,y)\,\mathrm{d}x, \tag{51}$$

$$= \int_X \big[\Phi(x,y)r^n(x,y) + \Psi(x,y)r^{n-1}(x,y)\big]\,\mathrm{d}x, \tag{52}$$

for ν-a.e. $y \in Y$, where $\Psi(\cdot, y)$ is the corresponding Kantorovich potential (cf. (21)).

Now let $\theta \in L^\infty(X \times Y, \mathcal{L} \otimes \nu)$ be a perturbation satisfying

$$\int_X \theta(x,y) r^n(x,y)\,\mathrm{d}x = 0 \quad \text{for } \nu\text{-a.e. } y \in Y, \qquad \int_Y \theta(x,y) r^n(x,y)\,\mathrm{d}\nu(y) = 0 \quad \text{for a.e. } x \in X. \tag{53}$$

Then $\rho_\varepsilon := (1 + \varepsilon\theta)\rho^n \in \Gamma(\mu, \nu)$, at least for sufficiently small $\varepsilon \geq 0$, i.e., $\varepsilon < 1/\|\theta\|_{L^\infty}$, since (53) guarantees the conservation of the marginals μ and ν. Notice further that the corresponding density $r_\varepsilon(\cdot, y)$ is positive for ν-a.e. $y \in Y$. Let Φ_ε and Ψ_ε be such that $\Phi_\varepsilon(\cdot, y)$ is the Kantorovich potential from $r_\varepsilon(\cdot, y)$ to $r^{n-1}(\cdot, y)$ for ν-a.e. $y \in Y$, made unique by normalization to zero average, and $\Psi_\varepsilon(\cdot, y)$ is the respective c-transform. Again by positivity of the densities, and by boundedness of X, we may conclude that $\Phi_\varepsilon(\cdot, y) \to \Phi(\cdot, y)$ uniformly as $\varepsilon \to 0$ for ν-a.e. $y \in Y$ [24, Theorem 1.52]. By optimality of ρ^n in $\mathcal{E}(\cdot \mid \rho^{n-1})$, and of Φ and Ψ for the transport, we then have

$$\begin{aligned}
&\frac{1}{\tau}\int_{X\times Y} \big[\Phi_\varepsilon(x,y) r_\varepsilon(x,y) + \Psi_\varepsilon(x,y) r^{n-1}(x,y)\big]\,\mathrm{d}x\,\mathrm{d}\nu(y) + \mathcal{E}(\rho_\varepsilon) \\
&\geq \frac{1}{\tau}\int_{X\times Y} \big[\Phi(x,y) r^n(x,y) + \Psi(x,y) r^{n-1}(x,y)\big]\,\mathrm{d}x\,\mathrm{d}\nu(y) + \mathcal{E}(\rho^n) \\
&\geq \frac{1}{\tau}\int_{X\times Y} \big[\Phi_\varepsilon(x,y) r^n(x,y) + \Psi_\varepsilon(x,y) r^{n-1}(x,y)\big]\,\mathrm{d}x\,\mathrm{d}\nu(y) + \mathcal{E}(\rho^n).
\end{aligned}$$

After subtraction and division by $\varepsilon > 0$, it follows that

$$\begin{aligned}
\frac{1}{\tau}\int_{X\times Y} \Phi_\varepsilon(x,y)(r^n\theta)(x,y)\,\mathrm{d}x\,\mathrm{d}\nu(y) + \int_{X\times Y} V(x,y)(r^n\theta)(x,y)\,\mathrm{d}x\,\mathrm{d}\nu(y) \\
+ \frac{\kappa}{\varepsilon}\int_{X\times Y} \big(r_\varepsilon \log r_\varepsilon - r_\varepsilon - r^n \log r^n + r^n\big)\,\mathrm{d}x\,\mathrm{d}\nu(y) \geq 0.
\end{aligned}$$

Passing to $\varepsilon \to 0$, recalling that $\Phi_\varepsilon(\cdot, y) \to \Phi(\cdot, y)$ uniformly and Φ is uniformly bounded by boundedness of the domain, we can extract the limit of the first two terms by means of dominated convergence. For the third term, since $s \mapsto f(s) = s\log s - s + 1$ is a differentiable convex function with $f'(s) = \log s$, it holds pointwise that $f(r_\varepsilon) - f(r^n) \geq \varepsilon\theta r^n \log r^n$, which in the $\varepsilon \to 0$ limit yields:

$$\int_{X\times Y} \left[\frac{1}{\tau}\Phi(x,y) + V(x,y) + \kappa \log r^n(x,y)\right]\theta(x,y)\,\mathrm{d}\rho^n(x,y) \geq 0.$$

Given that $\theta \in L^\infty(X \times Y, \mathcal{L} \otimes \nu)$ is arbitrary subject to (53), we conclude that there are functions $P \in L^1(X, \mu)$ and $Q \in L^1(Y, \nu)$ such that

$$\frac{1}{\tau}\Phi(x,y) + V(x,y) + \kappa \log r^n(x,y) = P(x) + Q(y); \tag{54}$$

see Lemma 17 from the Appendix for details. Since Q only shifts the level of the potentials $\Phi(\cdot, y)$ —which does not affect the x-gradient—, (46) and (47) follow in view of (50) with $\Pi^n := P$. P can

be chosen to have zero mean with respect to μ by adding an additional global constant offset to Φ. Moreover, summing (47) in time and using (45) yields (48).

Let us now show that r^n is uniformly bounded from above and away from zero, with a bound that does not depend on the iteration number n. The function $\Phi/\tau + V$ is uniformly Lipschitz in x by (50), with Lipschitz constant $L := \operatorname{diam}(X)/\tau + \|\nabla V\|_\infty$, so (54) implies that $\kappa \log r^n - \Pi^n$ must be uniformly L-Lipschitz in x as well. To deduce the individual regularity of each r^n and Π^n, define first the auxiliary function $\sigma := r^n e^{-\Pi^n/\kappa}$. Then $\kappa \log \sigma = \kappa \log r^n - \Pi^n$, so $\kappa \log \sigma$ is L-Lipschitz in x. Now note that, since $|\nabla_X(\kappa \log r^n - \Pi^n)|$ is uniformly bounded by L, after averaging in Y the Lipschitz bound cannot be worse. Therefore we find that μ-almost everywhere in X:

$$\begin{aligned} L\mu &\geq \left| \int_Y \nabla_X(\kappa \log r^n - \Pi^n) r^n \,\mathrm{d}\nu(y) \right| = \left| \kappa \int_Y (\nabla_X \log \sigma) \sigma e^{\Pi^n/\kappa} \,\mathrm{d}\nu(y) \right| = \left| \kappa \int_Y (\nabla_X \sigma) e^{\Pi^n/\kappa} \,\mathrm{d}\nu(y) \right| \\ &= \left| \kappa \nabla_X \left(\int_Y \sigma \,\mathrm{d}\nu(y) \right) e^{\Pi^n/\kappa} \right| = \left| \kappa \nabla \left(\mu e^{-\Pi^n/\kappa} \right) e^{\Pi^n/\kappa} \right| = \mu |\nabla(\kappa \log \mu - \Pi^n)|. \end{aligned}$$

This shows that the function $\Pi^n - \kappa \log \mu$ is also L-Lipschitz, and since by assumption $\nabla \log \mu$ is an L^2 function, $\nabla \Pi^n$ is in L^2 as well.

We will now show the uniform lower bound of r^n. Define $\Psi(x,y) = \kappa \log r^n(x,y) - \kappa \log \mu(x)$, which by (54) can be rewritten as:

$$\Psi(x,y) = \Pi^n(x) - \kappa \log \mu - \frac{1}{\tau}\Phi(x,y) - V(x,y). \tag{55}$$

Therefore Ψ has an x-Lipschitz constant of $2L$. On the other hand note that, by the properties of disintegrations (Lemma 3), for ν-a.e. $y \in Y$ it holds:

$$1 = \int_X r(x,y) \,\mathrm{d}x = \int_X \frac{r(x,y)}{\mu(x)} \,\mathrm{d}\mu(x) = \int_X e^{\Psi(x,y)/\kappa} \,\mathrm{d}\mu(x), \tag{56}$$

and consequently:

$$0 = \kappa \log \int_X e^{\Psi(x,y)/\kappa} \,\mathrm{d}\mu(x). \tag{57}$$

Since the log-sum-exp behaves like a weighted average, (57) is between the maximum and minimum value of Ψ; and since Ψ is continuous in x and X is connected, we conclude that for ν-a.e. $y \in Y$ there exists a point $\hat{x} \in X$ such that $\Psi(\hat{x}, y) = 0$. As a result, $|\Psi|$ is uniformly bounded by $2L \operatorname{diam} X$. We conclude that there exists a finite constant $C > 0$ depending on $\operatorname{diam}(X)$, τ, $\|\nabla V\|_\infty$, κ and the lower and upper bounds of μ (but not on n) such that $r^n = \mu \exp(\Psi/\kappa)$ is bounded as:

$$e^{-C} \leq r^n(x,y) \leq e^C. \tag{58}$$

For the estimate (49) let us further develop (47):

$$\begin{aligned} \frac{1}{\tau^2} \mathbf{W}_F(\rho^n, \rho^{n-1})^2 &= \int_{X\times Y} \left| \nabla_X(V - \Pi^n + \kappa \log r^n) \right|^2 \mathrm{d}\rho^n \\ &= \int_X \left| \nabla \Pi^n \right|^2 \mathrm{d}\mu + \int_{X\times Y} \left| \nabla_X(V + \kappa \log r^n) \right|^2 \mathrm{d}\rho^n - 2 \int_{X\times Y} \nabla \Pi^n \cdot \nabla_X(V + \kappa \log r^n) \,\mathrm{d}\rho^n. \end{aligned}$$

Rearranging terms, and noting that the second contribution is always non-negative:

$$\int_{X\times Y} \left| \nabla \Pi^n \right|^2 \mathrm{d}\mu \leq \frac{1}{\tau^2} \mathbf{W}_F(\rho^n, \rho^{n-1})^2 + 2 \int_{X\times Y} \nabla \Pi^n \cdot \nabla_X \left(V + \kappa \log r^n \right) \mathrm{d}\rho^n. \tag{59}$$

Note that for the entropic term we can write:

$$\int_{X\times Y} \nabla\Pi^n \cdot \nabla_X \log r^n \,\mathrm{d}\rho^n = \int_{X\times Y} \nabla\Pi^n \cdot \frac{\nabla_X r^n}{r^n} \,\mathrm{d}\rho^n = \int_{X\times Y} \nabla\Pi^n \cdot \nabla_X r^n \,\mathrm{d}x\,\mathrm{d}\nu(y)$$
$$= \int_X \nabla\Pi^n \cdot \nabla\mu \,\mathrm{d}x = \int_X \nabla\Pi^n \cdot \nabla \log\mu \,\mathrm{d}\mu.$$

Thus, summing (59) in time:

$$\sum_{n=1}^N \tau \int_{X\times Y} \left|\nabla\Pi^n\right|^2 \mathrm{d}\mu(x) \le \sum_{n=1}^N \frac{1}{\tau}\mathbf{W}_F(\rho^n,\rho^{n-1})^2 + 2\tau \int_{X\times Y} \nabla\Pi^n \cdot \nabla_X\left(V + \kappa\log r^n\right)\,\mathrm{d}\rho^n,$$

and using (45) for the first term:

$$\le 2\mathcal{E}(\rho^0) + \sum_{n=1}^N 2\tau \int_{X\times Y} \nabla\Pi^n \cdot \nabla_X V \,\mathrm{d}\rho^n + 2\tau\kappa \int_X \nabla\Pi^n \cdot \nabla\log\mu\,\mathrm{d}\mu$$
$$\le 2\mathcal{E}(\rho^0) + \sqrt{N\tau}\big(2\|\nabla_X V\|_\infty + 2\kappa\|\nabla\log\mu\|_{L^2(X,\mu)}\big)\left(\sum_{n=1}^N \tau\int_X \left|\nabla\Pi^n\right|^2\mathrm{d}\mu\right)^{1/2}$$
$$\le 2\mathcal{E}(\rho^0) + \frac{1}{2}N\tau\big(2\|\nabla_X V\|_\infty + 2\kappa\|\nabla\log\mu\|_{L^2(X,\mu)}\big)^2 + \frac{1}{2}\sum_{n=1}^N \tau\int_X \left|\nabla\Pi^n\right|^2\mathrm{d}\mu.$$

The bound on Π^n in (49) now follows from cancelling half of the sum in time and accounting for (15). To obtain the bound on $\sqrt{r^n}$, we expand the derivative of Φ in a different way:

$$\frac{1}{\tau}\int_{X\times Y} |\nabla_X\Phi|^2\,\mathrm{d}\rho^n = \tau\int_{X\times Y} |\nabla_X(\kappa\log r^n + V - \Pi^n)|^2\,\mathrm{d}\rho^n$$
$$\ge \tau\int_{X\times Y}\left[\frac{\kappa^2}{2}|\nabla_X\sqrt{r^n}|^2 - 2|\nabla_X V|^2 r^n - 2|\nabla\Pi^n|^2 r^n\right]\,\mathrm{d}x\,\mathrm{d}\nu(y).$$

Using the already established estimate on $\nabla\Pi^n$, the remaining part of (49) follows. □

Lemma 9. *For every $\psi \in C^\infty(\overline{X}\times Y)$ with $\mathbf{n}\cdot\nabla_X\psi(x,y) = 0$ for all $x\in\partial X$, $y\in Y$, we have that*

$$\int_{X\times Y}\psi\,\mathrm{d}\left[\frac{\rho_\tau^n - \rho_\tau^{n-1}}{\tau}\right] = -\int_{X\times Y}\nabla_X\psi\cdot\nabla_X\left[V - \Pi_\tau^n\right]\mathrm{d}\rho_\tau^n + \kappa\int_{X\times Y}\Delta_X\psi\,\mathrm{d}\rho_\tau^n + \tau\,R_\tau^n(\psi), \quad (60)$$

where the remainder is bounded as

$$\left|R_\tau^n(\psi)\right| \le \frac{1}{2}\|\psi\|_{C^2}S_\tau^n, \quad S_\tau^n := \frac{\mathbf{W}_F(\rho_\tau^n,\rho_\tau^{n-1})^2}{\tau^2}. \quad (61)$$

In particular, for every $\xi\in C^\infty(\overline{X})$ with $\mathbf{n}\cdot\nabla_X\xi(x)\equiv 0$ on ∂X,

$$\int_X \nabla_X\xi\cdot\nabla_X\Pi_\tau^n\,\mathrm{d}\mu - \int_{X\times Y}\nabla_X\xi\cdot\nabla_X V\,\mathrm{d}\rho_\tau^n + \kappa\int_{X\times Y}\Delta_X\xi\,\mathrm{d}\mu(x) = \tau\,R_\tau^n(\xi). \quad (62)$$

Proof. Integrating ρ_τ^{n-1} with respect to ψ and using (46) to perform:

$$\begin{aligned}\int_{X\times Y}\psi(x,y)\,\mathrm{d}\rho_\tau^{n-1}(x,y) &= \int_{X\times Y}\psi(x+\tau\nabla_X[V-\Pi_\tau^n+\kappa\log r_\tau^n],y)r_\tau^n(x,y)\,\mathrm{d}x\,\mathrm{d}\nu(y)\\ &= \int_{X\times Y}[\psi(x,y)+\tau\nabla_X\psi\cdot\nabla_X(V-\Pi_\tau^n+\kappa\log r_\tau^n)]r_\tau^n\,\mathrm{d}x\,\mathrm{d}\nu(y)\\ &\quad+\frac{\tau^2}{2}R_\tau^n(\psi),\end{aligned}$$

where the remainder $R_\tau^n(\psi)$ is bounded by the second-order term:

$$|R_\tau^n(\psi)|\le\|\nabla_X^2\psi\|_\infty\int_{X\times Y}|\nabla_X(V-\Pi_\tau^n+\kappa\log r_\tau^n)|^2\,\mathrm{d}\rho_\tau^n\le\|\nabla_X^2\psi\|_\infty\int_{X\times Y}\frac{\mathbf{W}_F(\rho_\tau^n,\rho_\tau^{n-1})^2}{\tau^2}.$$

Then (60) follows by collecting the finite difference in ρ_τ^n and integrating the entropic term by parts, while (62) results from choosing $\psi(x,y)=\xi(x)$. □

4. Existence of weak solutions

Throughout this section we denote by $\bar\Pi_\tau$ and $\bar\rho_\tau$ the piecewise-constant interpolations of the iterates $(\Pi_\tau^n)_n$ and $(\rho_\tau^n)_n$:

$$\bar\rho_\tau(t)=\rho_\tau^{\lfloor t/\tau\rfloor},\qquad \bar r_\tau(t)=r_\tau^{\lfloor t/\tau\rfloor},\qquad \bar\Pi_\tau(t)=\Pi_\tau^{\lfloor t/\tau\rfloor},\qquad\text{for all } t\in[0,\infty).\tag{63}$$

The main result is their convergence to weak solutions of the evolution equation (1)-(3). There is a technical aspect worth noting. Although most of the terms in the time-integrated version of (60) can be shown to converge by sheer weak* compactness of $\bar\rho_\tau$, the pairing between $\bar\Pi_\tau$ and $\bar\rho_\tau$ poses a special difficulty, since both converge only weakly. The control on the horizontal derivatives of $\bar\rho_\tau$ provided by Lemma 8 is unfortunately insufficient to extract even weak convergence of the pairing $\bar\rho_\tau\nabla\bar\Pi_\tau$.

Fortunately, as we shall show in Lemma 10 below, thanks precisely to the horizontal regularity bound shown in Lemma 8, the Y-average of $\bar\rho_\tau$ against a smooth test function yields a sequence converging strongly to the corresponding average of the limiting ρ. Since the potential $\bar\Pi_\tau$ is merely a function on X, integration against this strongly convergent sequence of averages yields the desired convergence of the joint product. This intuition is made rigorous by the following Lemma:

Lemma 10. *There are functions $\rho\in C(\mathbb{R}_{\ge0};\Gamma(\mu,\nu))$ and $\Pi\in L^2_{loc}(\mathbb{R}_{\ge0};H^1(X))$ such that, in the limit $\tau\to0$, at least along a suitable sequence and for each $T\in(0,\infty)$:*

$$\bar\rho_\tau(t)\rightharpoonup\rho(t)\quad\textit{locally uniformly w.r.t. } t\in\mathbb{R}_{\ge0}\textit{ in }\mathbf{W},\tag{64}$$

$$\bar\Pi_\tau\rightharpoonup\Pi\quad\textit{weakly in } L^2([0,T];H^1(X)).\tag{65}$$

Moreover, define for any given $\zeta\in C_c^1(\mathbb{R}_{\ge0}\times X\times Y)$ the vertical averages

$$\omega_\tau(t,x):=\int_Y\zeta(t,x,y)\bar r_\tau(t,x,y)\,\mathrm{d}\nu(y),\qquad\omega(t,x):=\int_Y\zeta(t,x,y)r(t,x,y)\,\mathrm{d}\nu(y).\tag{66}$$

where $\bar r_\tau\in L^1(X\times Y,\mathcal{L}\otimes\nu)$ and $r\in L^1(X\times Y,\mathcal{L}\otimes\nu)$ denote respectively the densities of $\bar\rho_n$ and ρ with respect to $\mathcal{L}\otimes\nu$. Then, along the same sequence as above,

$$\omega_\tau\to\omega\quad\textit{strongly in } L^2([0,T]\times X),\tag{67}$$

$$\omega_\tau\nabla\bar\Pi_\tau\rightharpoonup\omega\nabla\Pi\quad\textit{weakly in } L^1([0,T]\times X).\tag{68}$$

The limit (ρ, Π) satisfies, for all $\psi \in C_c^\infty(\mathbb{R}_+ \times X \times Y)$ with $\nabla_X \psi \cdot \mathbf{n} \equiv 0$ on ∂X the weak equation:

$$\int_0^\infty \int_{X\times Y} \partial_t \psi \,\mathrm{d}\rho = \int_0^\infty \int_{X\times Y} \left[\nabla_X \psi \cdot \nabla_X [V - \Pi] - \kappa \Delta_X \psi\right] \mathrm{d}\rho - \int_{X\times Y} \psi(0)\,\mathrm{d}\rho^0. \tag{69}$$

The potential Π is given for a.e. time as the unique solution satisfying $\int_X \Pi \,\mathrm{d}\mu = 0$ of the elliptic equation:

$$\int_X \nabla\xi \cdot \nabla\Pi(t)\ \mathrm{d}\mu = \int_X \nabla\xi \cdot \left[\int_Y \nabla_X V(x,y) r(t,x,y)\,\mathrm{d}\nu(y)\right] \mathrm{d}x - \kappa \int_X \Delta\xi \,\mathrm{d}\mu \tag{70}$$

for all $\xi \in C^\infty(X)$ with $\nabla\xi \cdot \mathbf{n} = 0$ on ∂X.

Proof. For the convergence of the trajectories (64) we employ the refined version of Ascoli-Arzela in [1, Proposition 3.3.1]: we use Lemma 7 for the equicontinuity of ρ w.r.t. $\mathbf{W}$, while the pointwise compactness follows from the compactness of $(\mathcal{P}(X \times Y), \mathbf{W})$ for X and Y compact. The fact that $\mathcal{E}(\bar\rho_\tau)$ is uniformly bounded by $\mathcal{E}(\rho_0)$ imposes a uniform upper bound on $\mathcal{H}(\bar\rho_\tau)$, which by lower-semicontinuity of the energy (Lemma 5) implies that

$$\mathcal{E}(\rho(t)) \le \liminf_{\tau\to 0} \mathcal{E}(\bar\rho_\tau(t)) \le \mathcal{E}(\rho^0), \tag{71}$$

and thus $\rho(t)$ must have a density with respect to $\mathcal{L}\otimes\nu$ for all $t \ge 0$, denoted by $r(t)$. On the other hand, the weak convergence of the potentials (65) is a direct consequence of the uniform bound (49) and the weighted Poincaré inequality for functions with zero mean [22].

For (67) let us first identify the limit. For a fixed $\zeta \in C_c^1(\mathbb{R}_{\ge 0} \times X \times Y)$, integrating with respect to a test function $\psi \in C_0(\mathbb{R}_{\ge 0} \times X)$:

$$\begin{aligned}
\int_0^\infty \int_X \psi(t,x)\omega_\tau(t,x)\,\mathrm{d}x\,\mathrm{d}t &= \int_0^\infty \int_{X\times Y} \psi(t,x)\zeta(t,x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y) \\
&\to \int_0^\infty \int_{X\times Y} \psi(t,x)\zeta(t,x,y)\,\mathrm{d}\rho(t,x,y) = \int_0^\infty \int_X \psi(t,x)\left[\int_Y \zeta(t,x,y) r(t,x,y)\,\mathrm{d}\nu(y)\right]\mathrm{d}x\,\mathrm{d}t \\
&= \int_0^\infty \int_X \psi(t,x)\omega(t,x)\,\mathrm{d}x\,\mathrm{d}t
\end{aligned} \tag{72}$$

This shows at least weak* convergence of $\omega_\tau\,\mathrm{d}x\,\mathrm{d}t$ to $\omega_\tau\,\mathrm{d}x\,\mathrm{d}t$. For showing a stronger convergence, we use the version of the Aubin-Lions lemma given in [34, Theorem 2]. As Banach space, we choose $B := L^1(X)$, as properly coercive integrand, we use

$$\mathcal{F}(\omega) = \int_X |\nabla_X \omega|\,\mathrm{d}x + \|\omega\|_{L^\infty(X)},$$

and the pseudo-distance is the dual norm in $C^1(X)$, i.e.,

$$g(\omega_1,\omega_2) = \sup\left\{\int_X \xi(x)\big(\omega_1(x) - \omega_2(x)\big)\,\mathrm{d}x \,\middle|\, \xi \in C^1(X),\ \|\xi\|_{C^1(X)} \le 1\right\}.$$

Let us first show that ω_τ is uniformly bounded in $L^\infty([0,T]\times X)$: since $\rho_\tau(t) \in \Gamma(\mu,\nu)$ for a.e. $t \in \mathbb{R}_{\ge 0}$, it holds:

$$|\omega_\tau(t,x)| = \left|\int_Y \zeta(t,x,y)\bar r_\tau(t,x,y)\,\mathrm{d}\nu(y)\right| \le \|\zeta\|_\infty \int_Y \bar r_\tau(t,x,y)\,\mathrm{d}\nu(y) = \|\zeta\|_\infty \mu(x) \tag{73}$$

for a.e. $(t,x) \in \mathbb{R}_{\geq 0} \times X$, and the upper bound on μ completes the claim. Further, ω_τ inherits from $\bar\rho_\tau$ the regularity of its X-gradient:

$$\begin{aligned}
\int_0^T \left(\int_X |\nabla_X \omega_\tau| \,\mathrm{d}x \right) \mathrm{d}t &= \int_0^T \int_X \left| \nabla_X \int_Y \zeta(t,x,y) \bar r_\tau(t,x,y) \,\mathrm{d}\nu(y) \right| \mathrm{d}t \,\mathrm{d}x \\
&\leq \int_0^T \int_{X\times Y} |\nabla_X (\zeta \bar r_\tau)| \,\mathrm{d}t \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&\leq T \|\nabla_X \zeta\|_\infty + \|\zeta\|_\infty \int_0^T \int_{X\times Y} |\nabla_X \bar r_\tau| \,\mathrm{d}t \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&= T \|\nabla_X \zeta\|_\infty + \|\zeta\|_\infty \int_0^T \int_{X\times Y} 2\sqrt{\bar r_\tau} \nabla_X (\sqrt{\bar r_\tau}) \,\mathrm{d}t \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&\leq T \|\nabla_X \zeta\|_\infty + 2\|\zeta\|_\infty \sqrt{T} \left(\int_0^T \int_{X\times Y} \left| \nabla_X \sqrt{\bar r_\tau} \right|^2 \mathrm{d}t \,\mathrm{d}x \,\mathrm{d}\nu(y) \right)^{1/2}.
\end{aligned}$$

Where we used Cauchy-Schwarz in the last inequality. In virtue of Lemma 8, we conclude

$$\int_0^T \left(\int_X |\nabla_X \omega_\tau| \,\mathrm{d}x \right) \mathrm{d}t \leq T \|\nabla_X \zeta\|_\infty + \frac{\|\zeta\|_\infty}{2\kappa} \sqrt{T} (C + \mathcal{E}(\rho^0)).$$

In combination with the L^∞-bound, we thus have

$$\sup_\tau \int_0^T \mathcal{F}(\omega_\tau) \,\mathrm{d}t < \infty. \tag{74}$$

Further, the Y-averages ω_τ also inherit the uniform regularity in time from $\bar\rho_\tau$: testing with $\xi \in C^1(X)$, we obtain that

$$\begin{aligned}
\int_X \xi(x) [\omega_\tau(m\tau,x) - \omega_\tau(n\tau,x)] \,\mathrm{d}x &= \int_{X\times Y} \xi(x) [(\zeta \bar r_\tau)(m\tau,x,y) - (\zeta \bar r_\tau)(n\tau,x,y)] \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&= \int_{X\times Y} \xi(x) [\zeta(m\tau,x,y) - \zeta(n\tau,x,y)] \bar r_\tau(m\tau,x,y) \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&\quad + \int_{X\times Y} \xi(x) \zeta(n\tau,x,y) [\bar r_\tau(m\tau,x,y) - \bar r_\tau(n\tau,x,y)] \,\mathrm{d}x \,\mathrm{d}\nu(y) \\
&\leq \|\xi\|_{C^1(X)} \|\zeta\|_{C^1(\mathbb{R}_{\geq 0}\times X\times Y)} \big[\tau(m-n) + \mathbf{W}_1(\rho_\tau^m, \rho_\tau^n) \big] \\
&\leq \|\xi\|_{C^1(X)} \|\zeta\|_{C^1(\mathbb{R}_{\geq 0}\times X\times Y)} \big[\tau(m-n) + C\sqrt{\tau(m-n)} \big]
\end{aligned}$$

where we used (20) in the first inequality, and Lemmas 2 and 7 in the last. In other words, we have that

$$\int_0^{T-h} g\big(\omega_\tau(t+h), \omega_\tau(t)\big) \,\mathrm{d}t \to 0 \quad \text{as } h \searrow 0, \text{ uniformly in } \tau \in (0,1). \tag{75}$$

Since (74) and (75) are verified, it follows by means of [34, Theorem 2] that ω_τ converges to a limit $\omega_* \in L^1([0,T]\times X)$, in $L^1(X)$, in measure with respect to time. Moreover, since the limit was already identified in (72), in particular we have that $\omega_* = \omega$. Invoking the uniform L^∞-bound (73) again, we conclude that actually $\omega \in L^\infty([0,T]\times X)$, and that $\omega_\tau \to \omega$ strongly in $L^p([0,T]\times X)$, for any $p \in [1,\infty)$.

In particular, the convergence $\omega_\tau \to \omega$ is strong in $L^2([0,T]\times X)$, and so (68) follows by weak convergence of $(\nabla \bar\Pi_\tau)_\tau$.

We now turn to the weak formulation. Let $\psi \in C_c^\infty(\mathbb{R}_+ \times X \times Y)$ with $\nabla_X \psi \cdot \mathbf{n} \equiv 0$ on ∂X. Since ψ is compactly supported, for each τ there exists an N such that $\psi \equiv 0$ for $t > N\tau$. Hence,

$$\begin{aligned}\int_0^\infty \int_{X\times Y} \partial_t \psi(t,x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y) &= \sum_{n=0}^{N-1} \int_{n\tau}^{(n+1)\tau} \int_{X\times Y} \partial_t\psi(t,x,y)\,\mathrm{d}\rho_\tau^n(x,y)\,\mathrm{d}t\\ &= \sum_{n=0}^{N-1} \int_{X\times Y} \big[\psi((n+1)\tau,x,y) - \psi(n\tau,x,y)\big]\,\mathrm{d}\rho_\tau^n(x,y)\\ &= -\int_{X\times Y} \psi(0,x,y)\,\mathrm{d}\rho^0(x,y) - \sum_{n=1}^{N-1} \tau \int_{X\times Y} \psi(n\tau,x,y)\,\mathrm{d}\left[\frac{\rho_\tau^n - \rho_\tau^{n-1}}{\tau}\right]\end{aligned}$$

The first term is the last term of (69). For the second term we turn to (60):

$$\begin{aligned}&\int_0^\infty \int_{X\times Y} \partial_t\psi(t,x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y) + \int_{X\times Y} \psi(0,x,y)\,\mathrm{d}\rho^0(x,y)\\ &= -\sum_{n=1}^{N-1} \tau\left[-\int_{X\times Y} \nabla_X\psi(n\tau)\cdot\nabla_X\big[V-\Pi_\tau^n\big]\,\mathrm{d}\rho_\tau^n + \kappa\int_{X\times Y}\Delta_X\psi(n\tau)\,\mathrm{d}\rho_\tau^n + \tau R_\tau^n(\psi(n\tau))\right]. \quad (76)\end{aligned}$$

Let us first control the remainder:

$$\sum_{n=1}^{N-1} \tau^2 R_\tau^n(\psi(n\tau)) \le \tau \sum_{n=1}^{N-1} \tau R_\tau^n(\psi(n\tau)) \le \tau\|\psi\|_{C^2} \sum_{n=1}^{N-1} \frac{\mathbf{W}_F(\rho_\tau^n,\rho_\tau^{n-1})^2}{2\tau} \le \tau\|\psi\|_{C^2}\mathcal{E}(\rho^0).$$

which yields an $O(\tau)$ contribution. Lemma 10 allows to control the rest of integrands, so that replacing $\psi(n\tau)$ by $\psi(t)$ in (76) only incurs in an $o(1)$ error. Thus, back to (76):

$$\begin{aligned}\int_0^\infty \int_{X\times Y} \partial_t\psi(t,x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y) &= \int_0^\infty \int_{X\times Y} \nabla_X\psi(t,x,y)\cdot\nabla_X V(x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y)\\ &\quad - \int_0^\infty \int_X \nabla_X\bar\Pi_\tau(t,x)\cdot\left[\int_Y \nabla_X\psi(t,x,y)\bar r_\tau(t,x,y)\,\mathrm{d}\nu(y)\right]\mathrm{d}t\,\mathrm{d}x\\ &\quad - \int_0^\infty \int_{X\times Y} \Delta_X\psi(t,x,y)\,\mathrm{d}\bar\rho_\tau(t,x,y)\\ &\quad - \int_{X\times Y} \psi(0,x,y)\,\mathrm{d}\rho^0(x,y) + o(1).\end{aligned}$$

The first and third lines converge by sheer integration of the weak* convergent $(\bar\rho_\tau)_\tau$ against continuous functions, while the second line converges in virtue of (68), and the last line is constant (up to the $o(1)$ term).

Finally, (70) follows from testing (69) with $\psi(t,x,y) = \theta(t)\xi(x)$, for θ with compact support and satisfying $\theta(0) = 0$:

$$\begin{aligned}0 &= \int_0^\infty \int_X \partial_t\theta(t)\xi(x)\,\mathrm{d}\mu(x)\,\mathrm{d}t = \int_0^\infty \int_{X\times Y} \partial_t\theta(t)\xi(x)\,\mathrm{d}\rho(t,x,y)\\ &= \int_0^\infty \int_{X\times Y} \theta(t)\nabla\xi(x)\cdot\nabla_X[V-\Pi]\,\mathrm{d}\rho(t,x,y) - \kappa\int_0^\infty\int_{X\times Y}\theta(t)\Delta_X\xi(x)\,\mathrm{d}\rho(t,x,y).\\ &= \int_0^\infty \int_X \theta(t)\nabla\xi(x)\cdot\left[-\mu\nabla\Pi + \int_Y \nabla_X V(\cdot,y)r(\cdot,\cdot,y)\,\mathrm{d}\nu(y)\right] - \kappa\int_0^\infty\int_X\theta(t)\Delta_X\xi(x)\,\mathrm{d}\mu(x)\,\mathrm{d}t,\end{aligned}$$

and the claim follows from the arbitrariness of θ. □

Lemma 11. *Let $(\rho_t, \Pi_t)_t$ be a limit trajectory as in Lemma 10 and $r_t \in L^1(X \times Y, \mathcal{L} \otimes \nu)$ denote the density of ρ_t with respect to $\mathcal{L} \otimes \nu$. Define the dissipation of the energy $\mathcal{E}$ as:*

$$\mathcal{I}(\rho_t, \Pi_t) := \int_{X\times Y} |\nabla_X (V - \Pi_t + \kappa \log r_t)|^2 \,\mathrm{d}\rho_t. \tag{77}$$

Then for all $0 \le t_1 < t_2 \le \infty$ the integral of the dissipation is lower-semicontinuous along the converging subsequences in Lemma 10, i.e.:

$$\int_{t_1}^{t_2} \mathcal{I}(\rho_t, \Pi_t)\,\mathrm{d}t \le \liminf_{k\to\infty} \int_{t_1}^{t_2} \mathcal{I}((\bar\rho_n)_t, (\bar\Pi_n)_t)\,\mathrm{d}t. \tag{78}$$

In particular,

$$0 \le \mathcal{E}(\rho_{t_1}) + \int_{t_1}^{t_2} \mathcal{I}(\rho_t, \Pi_t)\,\mathrm{d}t \le \mathcal{E}(\rho^0). \tag{79}$$

Proof. For simplicity in this proof we will remove the bars from $\bar\rho^n$ and $\bar\Pi^n$; simply note that they refer to the piecewise constant discrete trajectories. Then:

$$\begin{aligned}
\int_{t_1}^{t_2} \mathcal{I}((\rho_n)_t, (\Pi_n)_t)\,\mathrm{d}t &= \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X(V - \Pi_n + \kappa\log r_n)|^2 \,\mathrm{d}\rho_n \\
&= \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X(V + \kappa\log r_n)|^2 \,\mathrm{d}\rho_n && (80)\\
&\quad - \int_{t_1}^{t_2}\int_{X\times Y} 2\nabla_X(V + \kappa\log r_n)\cdot\nabla_X\Pi_n \,\mathrm{d}\rho_n && (81)\\
&\quad + \int_{t_1}^{t_2}\int_X |\nabla\Pi_n|^2 \,\mathrm{d}t\,\mathrm{d}\mu. && (82)
\end{aligned}$$

For (80), defining $\sigma_n := r_n e^{V/\kappa}$ we have

$$\begin{aligned}
\int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X(V + \kappa\log r_n)|^2 \,\mathrm{d}\rho_n &= \kappa^2 \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X\sigma_n|^2 \sigma_n e^{-V/k}\,\mathrm{d}x\,\mathrm{d}\nu(y) \\
&= 4\kappa^2 \int_{t_1}^{t_2}\int_{X\times Y} e^{-V/k} |\nabla_X\sqrt{\sigma_n}|^2 \,\mathrm{d}x\,\mathrm{d}\nu(y).
\end{aligned}$$

which is lower-semicontinuous, since σ_n inherits ρ_n's regularity properties outlined in Lemma 12. The next term (81) can be rewritten as follows:

$$\begin{aligned}
&\int_{t_1}^{t_2}\int_{X\times Y} \nabla_X(V + \kappa\log r_n)\cdot\nabla_X\Pi_n \,\mathrm{d}\rho_n = \\
&\qquad = \int_{t_1}^{t_2}\int_X \nabla\Pi_n \cdot \left[\int_Y \rho_n \nabla_X V \,\mathrm{d}\nu(y)\right] \mathrm{d}t\,\mathrm{d}x + \kappa\int_{t_1}^{t_2}\int_X \nabla\Pi_n\cdot\nabla\log\mu\,\mathrm{d}\mu,
\end{aligned}$$

where the first contribution converges thanks to Lemma 13 and the second by sheer weak convergence of $\nabla(\Pi_n)_t$ (plus the fact that $\log\mu \in H^1(X,\mu)$, thanks to the finite Fisher information of μ). The last term (82) is lower-semicontinuous under weak convergence of $(\Pi_n)_n$ in $L^2([t_1,t_2]; H^1(X,\mu))$.

Finally, (79) is a direct consequence of (48) and the lower-semicontinuity of both energy and the integrated dissipation. □

Lemma 12. *The weak solutions* (ρ, Π) *of* (69)-(70) *extracted as the limit of the minimizing movement scheme in Lemma 10 satisfy, for all* $0 < t_1 < t_2 < \infty$*:*

$$\mathbf{W}^2(\rho_{t_1}, \rho_{t_2}) \leq \mathbf{W}_F^2(\rho_{t_1}, \rho_{t_2}) \leq \mathcal{E}(\rho_0)(t_2 - t_1). \tag{83}$$

$$\int_{t_1}^{t_2} \int_X \left|\nabla \Pi_t\right|^2 \mathrm{d}t \, \mathrm{d}\mu(x) \leq C(t_2 - t_1). \tag{84}$$

$$2\kappa^2 \int_{t_1}^{t_2} \int_{X\times Y} \left|\nabla_X \sqrt{r_t}\right|^2 \mathrm{d}t \, \mathrm{d}x \, \mathrm{d}\nu(y) \leq \mathcal{E}(\rho^0) + C(t_2 - t_1). \tag{85}$$

with C *a finite constant depending only on* $\|\nabla_X V\|_\infty$, $\|\nabla \sqrt{\mu}\|_2$ *and* κ*.*

Remark 4. Even though the weak solutions constructed in Lemma 10 can (a priori) only be shown to be continuous in the $\mathbf{W}$ metric, a posteriori Lemma 12 yields regularity in time with respect to the $\mathbf{W}_F$ metric.

Proof of Lemma 12. For (83) we will use the Benamou-Brenier formula fiberwise. For given $t_1 < t_2$, and for each $s \in [0,1]$ consider $\varrho_s := \rho((1-s)t_1 + st_2)$ with density given by $\tilde{r}_s \in L^1(X \times Y, \mathcal{L} \otimes \nu)$ and $\pi_s := \Pi((1-s)t_1 + st_2)$. For each $y \in Y$, define the velocity field:

$$v_s : (x, y) \mapsto (t_2 - t_1)\nabla_X [V(x, y) - \pi_s(x) + \kappa \log \tilde{r}_s(x, y)], \tag{86}$$

The pair $(\tilde{r}_s, v_s)$ thus constructed is a feasible weak solution for the continuity equation:

$$\begin{aligned} \partial_s \tilde{r}_s(x, y) + \nabla_X \cdot (\tilde{r}_s(x, y) v_s(x, y)) = 0 \\ \nabla_X (\tilde{r}_s v_s) = 0 \quad \text{on } \partial X \end{aligned} \tag{87}$$

Now, by the Benamou-Brenier formula,

$$\begin{aligned} \mathbf{W}_F(\varrho_0, \varrho_1)^2 &= \int_Y \mathbf{W}(\tilde{r}_0(\cdot, y), \tilde{r}_0(\cdot, y)) \, \mathrm{d}\nu(y) \leq \int_Y \left[\int_0^1 \int_X |v_s(x, y)|^2 \tilde{r}_s(x, y) \, \mathrm{d}x \, \mathrm{d}s \right] \mathrm{d}\nu(y) \\ &= \int_0^1 \int_{X\times Y} |v_s(x, y)|^2 \, \mathrm{d}\varrho_s(x, y) \, \mathrm{d}s \\ &= (t_2 - t_1)^2 \int_0^1 \int_{X\times Y} |\nabla_X [V(x, y) - \pi_s(x) + \kappa \log \tilde{r}_s(x, y)]|^2 \, \mathrm{d}\varrho_s(x, y) \, \mathrm{d}s, \end{aligned}$$

and after a change of variables we eventually obtain:

$$\begin{aligned} \mathbf{W}_F(\rho(t_1), \rho(t_2))^2 &= (t_2 - t_1) \int_{t_1}^{t_2} \int_{X\times Y} |\nabla_X [V(x, y) - \Pi_t(x) + \kappa \log r_t(x, y)]|^2 \, \mathrm{d}\rho_t(x, y) \, \mathrm{d}t \\ &\leq (t_2 - t_1)\mathcal{E}(\rho^0), \end{aligned}$$

where in the last inequality we used Lemma 11.

For (84) we test (70) with (an approximating sequence of) Π_t and integrate by parts to obtain:

$$\begin{aligned} \int_X |\nabla \Pi_t|^2 \, \mathrm{d}\mu(x) &= \int_{X\times Y} \nabla \Pi_t \cdot \nabla_X V \, \mathrm{d}\rho_t + \kappa \int_X \nabla \Pi_t \cdot \nabla \log \mu \, \mathrm{d}\mu(x) \\ &\leq \|\nabla_X V\|_\infty \left(\int_{X\times Y} |\nabla \Pi_t|^2 \, \mathrm{d}\rho_t \right)^{1/2} + \kappa \left(\int_{X\times Y} |\nabla \Pi_t|^2 \, \mathrm{d}\mu \right)^{1/2} \left(\int_{X\times Y} |\nabla \log \mu|^2 \, \mathrm{d}\mu \right)^{1/2}, \end{aligned}$$

where we used Cauchy-Schwartz. Then, using that $\|\nabla \log \mu\|_{L^2(X,\mu)} = 2\|\nabla\sqrt{\mu}\|_{L^2(X)}$ and simplifying one $\|\nabla \Pi_t\|_{L^2(X,\mu)}$ term yields:

$$\left(\int_{X\times Y} |\nabla \Pi_t|^2 \,\mathrm{d}\mu\right)^{1/2} \leq \|\nabla_X V\|_\infty + 2\kappa \|\nabla\sqrt{\mu}\|_2, \tag{88}$$

and (84) follows by integration in time.

To obtain (85) we expand (79):

$$\begin{aligned}
\mathcal{E}(\rho_0) &\geq \int_{t_1}^{t_2} \mathcal{I}(\rho_t, \Pi_t)\,\mathrm{d}t = \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X(V - \Pi + \kappa \log r)|^2 \,\mathrm{d}\rho && (89)\\
&= \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X V|^2 \,\mathrm{d}\rho + \int_{t_1}^{t_2}\int_X |\nabla \Pi|^2 \,\mathrm{d}\mu\,\mathrm{d}t + \kappa^2 \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X \log r|^2 \,\mathrm{d}\rho && (90)\\
&\quad + 2\int_{t_1}^{t_2}\int_{X\times Y} [\kappa \nabla_X V\cdot\nabla_X \log r - \kappa \nabla_X \log r\cdot\nabla_X \Pi - \nabla_X V\cdot\nabla_X \Pi]\,\mathrm{d}\rho && (91)
\end{aligned}$$

To bound the Fisher information of ρ we solve for the last term in (90):

$$\begin{aligned}
4\kappa^2 \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X \sqrt{r}|^2 \,\mathrm{d}x\,\mathrm{d}\nu(y) \leq \mathcal{E}(\rho^0) &- 2\int_{t_1}^{t_2}\int_{X\times Y} \kappa \nabla_X V\cdot\nabla_X \log r\,\mathrm{d}\rho\\
&+ 2\int_{t_1}^{t_2}\int_{X\times Y} [\kappa\nabla_X \log r\cdot\nabla_X\Pi - \nabla_X V\cdot\nabla_X\Pi]\,\mathrm{d}\rho
\end{aligned}$$

Now all the inner products are straightforward to bound:

$$\begin{aligned}
-2\kappa &\int_{t_1}^{t_2}\int_{X\times Y} \nabla_X V\cdot\nabla_X \log r\,\mathrm{d}\rho\\
&\leq 2\kappa \int_{t_1}^{t_2} \left(\int_{X\times Y} |\nabla_X V|^2\,\mathrm{d}\rho_t\right)^{1/2} \left(\int_{X\times Y} |\nabla_X \log r|^2\,\mathrm{d}\rho_t\right)^{1/2}\mathrm{d}t\\
&\leq 4\kappa \|\nabla_X V\|_\infty \sqrt{t_2 - t_1}\left(\int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X\sqrt{r}|^2\,\mathrm{d}x\,\mathrm{d}\nu(y)\,\mathrm{d}t\right)^{1/2}\\
&\leq 2\|\nabla_X V\|_\infty^2 (t_2 - t_1) + 2\kappa^2 \int_{t_1}^{t_2}\int_{X\times Y} |\nabla_X\sqrt{\rho}|^2\,\mathrm{d}x\,\mathrm{d}\nu(y)\,\mathrm{d}t,
\end{aligned}$$

where we used $ab \leq a^2/2 + b^2/2$ for $a = 2\|\nabla_X V\|_\infty\sqrt{t_2 - t_1}$ and b encompassing the rest. For the next term:

$$\begin{aligned}
2\kappa \int_{t_1}^{t_2}\int_{X\times Y} \nabla_X \log r\cdot\nabla_X \Pi\,\mathrm{d}\rho \leq 2\kappa \int_{t_1}^{t_2}\int_X \nabla\mu\cdot\nabla\Pi\,\mathrm{d}x &\leq 2\kappa\int_{t_1}^{t_2}\int_X \nabla\log\mu\cdot\nabla\Pi\,\mathrm{d}\mu\\
&\leq 4\kappa\left(\int_{t_1}^{t_2}\int_X |\nabla\Pi|^2\,\mathrm{d}\mu\right)^{1/2}\sqrt{t_2 - t_1}\left(\int_X |\nabla\sqrt{\mu}|^2\,\mathrm{d}x\right)^{1/2}\\
&\leq 2\kappa\int_{t_1}^{t_2}\int_X |\nabla\Pi|^2\,\mathrm{d}\mu + 2\kappa(t_2 - t_1)\int_X |\nabla\sqrt{\mu}|^2\,\mathrm{d}x,
\end{aligned}$$

where we used a similar bound. For the remaining term:

$$2\int_{t_1}^{t_2}\int_{X\times Y}\nabla_X V\cdot\nabla_X\Pi\,\mathrm{d}\rho \le 2\|\nabla_X V\|_\infty\sqrt{t_2-t_1}\left(\int_{t_1}^{t_2}\int_X|\nabla\Pi|^2\,\mathrm{d}\mu\right)^{1/2}$$
$$\le \|\nabla_X V\|_\infty^2(t_2-t_1)+\int_{t_1}^{t_2}\int_X|\nabla\Pi|^2\,\mathrm{d}\mu$$

Finally, collecting all the contributions and using (84) yields the desired bound. □

5. Stability of weak solutions

In the following sections, for a weak solution (ρ,Π) of (69)-(70) we will denote by ρ_t the measure $\rho(t)$ (which is for a.e. t a feasible coupling in $\Gamma(\mu,\nu)$ with density with respect to $\mathcal{L}\otimes\nu$ given by $r_t\in L^1(X\times Y,\mathcal{L}\otimes\nu)$), and analogously we will denote by Π_t the function $\Pi(t)$.

We start with a preliminary result that will be useful below:

Lemma 13 (Stability of solutions)**.** *Let $(\mu_n)_n\subset\mathcal{P}(X)$ be a sequence with uniformly bounded Fisher information $\|\nabla\sqrt{\mu_n}\|_2\le C<\infty$ and uniformly bounded from above and away from zero, converging strongly to μ in $L^2(X)$, and let $(\nu_n)_n\subset\mathcal{P}(Y)$ converge to ν weak*. For each $n\in\mathbb{N}$ let (ρ_n,Π_n) be a weak solution to the continuity equation (69)-(70) with initial condition $\rho_n^0\in\Gamma(\mu_n,\nu_n)$ and satisfying the estimates in Lemma 12. Further assume that $(\rho_n^0)_n$ converges weak* to $\rho^0\in\Gamma(\mu,\nu)$, and that $\mathcal{E}(\rho_n^0\mid\mu_n,\nu_n)$ is uniformly bounded in n.*

Then, up to the extraction of a subsequence, $(\rho_n)_n$ converges on compact sets to a weak solution of (69)-(70) with initial condition ρ^0. More precisely, for each $T>0$,

$$\rho_n\rightharpoonup\rho\quad\text{locally uniformly w.r.t. } t\in\mathbb{R}_{\ge0}\text{ in }\mathbf{W},\tag{92}$$
$$\Pi_n\rightharpoonup\Pi\quad\text{weakly in } L^2([0,T];H^1(X)).\tag{93}$$

Moreover, define for any given $\zeta\in C_c^1(\mathbb{R}_{\ge0}\times X\times Y)$:

$$\omega_n(t,x):=\int_Y\zeta(t,x,y)r_n(t,x,y)\,\mathrm{d}\nu(y),\qquad\omega(t,x):=\int_Y\zeta(t,x,y)r(t,x,y)\,\mathrm{d}\nu(y).\tag{94}$$

where $r_n\in L^1(X\times Y,\mathcal{L}\otimes\nu_n)$ and $r\in L^1(X\times Y,\mathcal{L}\otimes\nu)$ denote respectively the densities of ρ_n and ρ. Then, along the same sequence as above,

$$\omega_n\to\omega\quad\text{strongly in } L^2([0,T]\times X),\tag{95}$$
$$\omega_n\nabla\Pi_n\rightharpoonup\omega\nabla\Pi\quad\text{weakly in } L^1([0,T]\times X).\tag{96}$$

Proof. As in Lemma 10, for the convergence of the trajectories (92) we employ the refined version of Ascoli-Arzela in [1, Proposition 3.3.1]: we use Lemma 12 for the equicontinuity of ρ w.r.t. $\mathbf{W}$, while the pointwise compactness follows from the compactness of $(\mathcal{P}(X\times Y),\mathbf{W})$ for X and Y compact. This yields existence of a $\mathbf{W}$-continuous limit trajectory ρ, to which $(\rho_n)_n$ converges locally uniformly. Besides, ρ_t has marginals μ and ν for all $t\ge0$ by weak* convergence of (μ_n,ν_n) to (μ,ν) and continuity of the projection operators P_X and P_Y.

To show density of ρ_t with respect to $\mathcal{L}\otimes\nu$ note first that, by Lemma 11, $\mathcal{E}(\rho_n(t)\mid\mu_n,\nu_n)\le\mathcal{E}(\rho_n^0\mid\mu_n,\nu_n)\le C$, since by assumption the energy of the initial datum is uniformly bounded, and by Lemma 79 the energy is non-increasing. On the other hand, the joint lower-semicontinuity of the energy with respect to weak* convergence (Lemma 5) implies:

$$\mathcal{E}(\rho(t)\mid\mu,\nu)\le\liminf_{n\to\infty}\mathcal{E}(\rho_n(t)\mid\mu_n,\nu_n)\le\liminf_{n\to\infty}\mathcal{E}(\rho_n^0\mid\mu_n,\nu_n)\le C,\tag{97}$$

which shows that $\rho(t)$ has a density with respect to $\mathcal{L} \otimes \nu$ for all $t \geq 0$.

Now, the bounds (84) and (85) are uniform in n, since they only depend on the energy of the initial datum, the Fisher information of μ_n, κ and $\|\nabla_X V\|_\infty$, all of which are uniformly bounded by assumption. In view of this, (84) together with the the weighted Poincaré inequality for functions with zero mean [22] imply the weak convergence of the potentials Π_n to a limit trajectory $\Pi \in L^2([0,T]; H^1(X,\mu))$ for all $T > 0$. Further, the uniform control on the horizontal Fisher information of $(\rho^n)_n$ given by (85) implies convergence of the vertical averages (95) and (96) in a similar fashion to the proof in Lemma 10. The main difference is that now the marginals μ_n and ν_n vary along the converging sequence, but this can be factored into the proof without major issues.

The remaining step is to show that ρ and Π verify their respective continuity equations. Let $\psi \in C_c^\infty(\mathbb{R}_+ \times X \times Y)$ with $\nabla_X \psi \cdot \mathbf{n} \equiv 0$ on ∂X, and test (62) with ψ for the weak solution (ρ_n, Π_n):

$$\int_{X\times Y} \psi(0)\,\mathrm{d}\rho_n^0 = \int_{\mathbb{R}_{\geq 0}\times X\times Y} \left[\nabla_X\psi\cdot\nabla_X V - \nabla_X\psi\cdot\nabla\Pi^k - \kappa\Delta_X\psi - \partial_t\psi\right]\,\mathrm{d}\rho^n. \tag{98}$$

The left-hand side converges by weak* convergence of the initial conditions. Likewise, the first, third and last term of the right-hand side converge by sheer weak* convergence of ρ^n to ρ. The remaining term can be shown to converge in virtue of (96): defining $(\mathbf{u}^k)_n$ and $\mathbf{u}$ as

$$\mathbf{u}_n(t,x) := \int_Y \nabla_X\psi(t,x,y) r_n(t,x,y)\,\mathrm{d}\nu(y), \qquad \mathbf{u}(t,x) := \int_Y \nabla_X\psi(t,x,y) r(t,x,y)\,\mathrm{d}\nu(y), \tag{99}$$

and using (96) we obtain convergence also of the remaining term from (98):

$$\begin{aligned}\int_0^T\int_{X\times Y} \nabla_X\psi\cdot\nabla\Pi_n\,\mathrm{d}\rho_n &= \int_0^T\int_X \nabla\Pi_n(x)\cdot\mathbf{u}_n(t,x)\,\mathrm{d}t\,\mathrm{d}x\\ &\to \int_0^T\int_X \nabla\Pi(x)\cdot\mathbf{u}(t,x)\,\mathrm{d}t\,\mathrm{d}x = \int_0^T\int_{X\times Y}\nabla_X\psi\cdot\nabla\Pi\,\mathrm{d}\rho.\end{aligned}$$

where T the threshold after which $\psi(t)$ is identically zero.

Finally, the fact that Π obeys its respective weak equation (70) follows by testing (98) with a test function with no dependence in y, plus noting that ρ_n converges weak* to ρ and that μ_n converges strongly to μ in $L^2(X)$:

$$\begin{aligned}\int_0^T\int_X \nabla_X\xi\cdot\nabla\Pi\;\mathrm{d}t\,\mathrm{d}\mu &= \lim_n \int_0^T\int_X \nabla_X\xi\cdot\nabla\Pi_n\;\mathrm{d}t\,\mathrm{d}\mu_n\\ &= \lim_n -\kappa\int_0^T\int_X \Delta_X\xi\,\mathrm{d}t\,\mathrm{d}\mu_n + \int_0^T\int_{X\times Y}\nabla_X\xi\cdot\nabla_X V\,\mathrm{d}\rho_n\\ &= -\kappa\int_0^T\int_X\Delta_X\xi\,\mathrm{d}t\,\mathrm{d}\mu + \int_0^T\int_{X\times Y}\nabla_X\xi\cdot\nabla_X V\,\mathrm{d}\rho\end{aligned}$$

□

Remark 5. Lemma 13 allows to extract the continuum limit of finite, multiphasic flows such as those studied in [11] by crafting a pertinent sequence of discrete second marginals ν_n converging to a weak* limit ν. We illustrate this convergence with numerical examples in Section 8.

6. Asymptotic limit of weak solutions

In this section we show that any weak solution (ρ_t, Π_t) converges as $t \to \infty$ to the unique minimizer of the energy $\mathcal{E}$. This contrasts with the behavior of the solutions of the AHT scheme (corresponding to $\kappa = 0$ and ρ_t being of Monge type, cf. Section 1.2 and also Section 7 further below), which are not guaranteed to converge to the optimizer and can stay stationary in suboptimal configurations.

Lemma 14. *For a sequence of weak solutions (ρ_n, Π_n) with initial conditions $(\rho_n^0)_n$ and fixed marginals μ and ν converging to (ρ, Π) on a time interval $(t_1, t_2) \subset [0, \infty)$ under the same conditions of Lemma 13, the integral of the dissipation is lower-semicontinuous, i.e.:*

$$\int_{t_1}^{t_2} \mathcal{I}(\rho_t, \Pi_t)\,\mathrm{d}t \leq \liminf_{k\to\infty} \int_{t_1}^{t_2} \mathcal{I}((\rho_n)_t, (\Pi_n)_t)\,\mathrm{d}t. \tag{100}$$

Proof. The proof is analogue to that of Lemma 11, replacing the regularity estimates coming from the minimizing movement scheme by the weak solution estimates in Lemma 12. □

Lemma 15. *Let $\rho_\infty \in \Gamma(\mu, \nu)$, with a non-negative density $r_\infty \geq 0$ in $L^1(X \times Y, \mathcal{L} \otimes \nu)$, and let Π_∞ be the corresponding pressure given by (70). If $\mathcal{I}(\rho_\infty, \Pi_\infty) = 0$, then ρ_∞ is the unique minimizer of the energy $\mathcal{E}$.*

Proof. By assumption:

$$0 = \int_{X\times Y} \left|\nabla_X \big(\kappa \log r_\infty(x,y) + V(x,y) - \Pi_\infty(x)\big)\right|^2 \mathrm{d}\rho_\infty(x,y). \tag{101}$$

First we will show that Π_∞ is bounded. Since r_∞ is non-negative, the gradient in (101) must vanish wherever r_∞ is strictly positive, i.e.

$$\nabla\Pi_\infty(x) = \nabla_X V(x,y) + \kappa \nabla_X \log r_\infty(x,y) \quad \text{for } \rho_\infty\text{-a.e. } (x,y). \tag{102}$$

Multiplying by ρ_∞ and integrating in Y one finds:

$$\nabla\Pi_\infty(x)\mu(x) = \int_Y \nabla_X V(x,y) r_\infty(x,y)\,\mathrm{d}\nu(y) + \kappa\nabla\mu(x) \quad \text{for a.e. } x, \tag{103}$$

or

$$\nabla\Pi_\infty(x) = \int_Y \nabla_X V(x,y) \frac{r_\infty(x,y)}{\mu(x)}\,\mathrm{d}\nu(y) + \kappa\nabla\log\mu(x) \quad \text{for a.e. } x. \tag{104}$$

Since $\int_Y r_\infty(x,y)\,\mathrm{d}\nu(y) = \mu(x)$ and $\nabla_X V$ is bounded, we deduce $\Pi_\infty - \kappa\log\mu$ is bounded on X, and in view of the uniform bounds on μ so is Π_∞. Now, defining $\sigma := e^{(V-\Pi_\infty)/\kappa} r_\infty$ we can rewrite (101) as:

$$0 = \int_{X\times Y} |\kappa\nabla_X \log\sigma|^2\, \sigma e^{-(V-\Pi_\infty)/\kappa}\,\mathrm{d}x\,\mathrm{d}\nu = 4\kappa^2 \int_{X\times Y} \left|\nabla_X\sqrt{\sigma}\right|^2 e^{-(V-\Pi_\infty)/\kappa}\,\mathrm{d}x\,\mathrm{d}\nu. \tag{105}$$

And since both V and Π_∞ are bounded, κ is positive and X has connected interior, $\sigma(x,y)$ must be for each y equal to a positive constant $e^{\Psi(y)/\kappa}$. Rearranging terms yields:

$$r_\infty(x,y) = \exp\left(\frac{\Pi_\infty(x) + \Psi(y) - V(x,y)}{\kappa}\right). \tag{106}$$

Finally, since the measure $\rho_\infty = r_\infty\,\mathrm{d}x\,\mathrm{d}\nu(y)$ has marginals μ and ν and its density takes the form of the diagonal scaling (106), we conclude by Lemma 4 that it is the unique minimizer of $\mathcal{E}$,

or analogously, the optimal entropic plan between μ and ν with cost function V and regularization strength κ. The optimal entropic dual potentials are thus Π_∞ and Ψ. □

Lemma 16. *The weak solutions $(\rho_t, \Pi_t)_t$ obtained in Lemma 10 tend as $t \to \infty$ to the unique minimizer of $\mathcal{E}$.*

Proof. Our convergence argument follows a similar rationale to that in [28]. Let (ρ, Π) be a weak solution of (69)-(70). There exist cluster points of the trajectory $(\rho_t)_t$ by sheer weak* compactness in the space of probability measures on $X \times Y$; let us denote such a cluster point by ρ_* and the converging subsequence of times by $(t_n)_n$. By weak* continuity, ρ_* preserves the marginals and positivity of the trajectory ρ.

Let us define functions $\rho_n(s) := \rho(t_n + s)$, $\Pi_n := \Pi(t_n + s)$ for $s \in [0, 1]$. Since each (ρ_n, Π_n) is a weak solution of (69)-(70) with initial condition $\rho_n(0) = \rho(t_n) \to \rho_*$, the stability of weak solutions (Lemma 13) grants convergence of (a subsequence of) $(\rho_n, \Pi_n)_n$ to a weak solution (ρ', Π') with initial condition ρ_*. Now, since the dissipation $\mathcal{I}$ of the original weak solution (ρ, Π) is positive and summable, the integrals of the form $\int_0^1 \mathcal{I}(\rho_n(s), \Pi_n(s))\,\mathrm{d}s$ must converge to zero, which combined with Lemma 14 implies that the dissipation vanishes in the asymptotic limit:

$$0 \leq \int_0^1 \mathcal{I}(\rho'(s), \Pi'(s))\,\mathrm{d}s \leq \liminf \int_0^1 \mathcal{I}(\rho_n(s), \Pi_n(s))\,\mathrm{d}s = 0 \tag{107}$$

Since $\mathcal{I}$ is a non-negative quantity (for $\rho \geq 0$), this means that $\mathcal{I}(\rho'(s), \Pi'(s))$ is identically zero, which by Lemma 15 implies that $\rho'(s)$ and ρ_* are equal to the unique minimizer of $\mathcal{E}$, given by (106). Since this convergence holds along any subsequence, the whole trajectory converges to ρ_*. □

7. Stationarity in the unregularized case

Section 6 showed that solutions to (1)-(3) converge as $t \to \infty$ to the minimizer of $\mathcal{E}$. In contrast, the AHT scheme does not enjoy such a global convergence. As explained in Section 1.2, equations (1)-(3) introduce two kinds of modifications with respect to the AHT scheme: a relaxation (enlarging the space of transport maps to that of transport plans) and a smoothing (adding an entropic regularization term to the energy). In this section we show that the relaxation step is not sufficient to endow the AHT scheme with convergence properties analogous to those of (1)-(3), and that some kind of regularization appears to be necessary.

As reviewed in Section 1.2, the problem of existence of solutions becomes in general much harder for $\kappa = 0$, since the a priori estimate on the horizontal gradients of ρ (or r) is lost. In fact, for an absolutely continuous ν, weak solutions to (1)-(3) may be extremely degenerate, possibly giving mass to a single point in each y-fiber. Due to these difficulties we will not attempt to provide a comprehensive picture of the $\kappa = 0$ case. Instead, we will exemplify that, when weak solutions exist, they face the same type of obstacle for asymptotic convergence to the energy minimizer that is observed in the AHT scheme. The reason is that in the $\kappa = 0$ regime there exist suboptimal stationary points of the evolution equation. We will demonstrate this by showing that, whenever $V \in C^2(X \times Y)$ satisfies the *twist condition* (i.e., $\nabla_X V(x, \cdot)$ is injective for all x), any initial datum ρ_0 of the form

$$\rho_0 = (\mathrm{id}, T_*)_\sharp \mu, \qquad \text{with } T_*(x) = \nabla_X V(x, \cdot)^{-1}(\nabla \Pi_*(x)), \qquad T_{*\sharp}\mu = \nu \tag{108}$$

with Π_* in $C^2(X)$ stays stationary by the discrete minimizing movement when τ is small enough. As a consequence, in this case a limit trajectory exists and is constantly equal to the initialization, which in general will be suboptimal for the energy.

Let us show this in detail. For $\kappa = 0$ the minimizing movement (38) becomes:

$$\rho_\tau^n \in \operatorname*{argmin}_{\rho\in\Gamma(\mu,\nu)} \frac{1}{2\tau}\mathbf{W}_F(\rho,\rho_\tau^{n-1})^2 + \mathcal{E}_0(\rho), \qquad \mathcal{E}_0(\rho) := \int_{X\times Y} V \,\mathrm{d}\rho. \tag{109}$$

Our argument hinges on a primal-dual argument for the minimizing movement. In analogy to Lemma 6, (109) can be rewritten as

$$\rho^n = \mathrm{P}_{13}\gamma^n, \quad \text{with } \gamma^n \in \operatorname*{argmin}_{\gamma\in\Gamma(\mu,\rho^{n-1})} \int_{X\times X\times Y} \left(\frac{|x-x'|^2}{2\tau} + V(x,y)\right) \mathrm{d}\gamma(x,x',y), \tag{110}$$

with respective (pre)-dual problem

$$\begin{aligned} &\sup_{\substack{\Pi\,\in\, C(X)\\ \Psi\,\in\, C(X\times Y)}} \int_X \Pi(x)\,\mathrm{d}\mu(x) + \int_{X\times Y} \Psi(x',y)\,\mathrm{d}\rho^{n-1}(x',y) \\ &\text{s.t. } \Pi(x) + \Psi(x',y) \le \frac{|x-x'|^2}{2\tau} + V(x,y) \text{ for all } (x,x',y)\in X\times X\times Y. \end{aligned} \tag{111}$$

We can show that ρ_0 is stationary by the minimizing movement —or, equivalently, optimal for the primal problem (110) when choosing $\rho^{n-1} = \rho_0$ — by building a dual certificate. Choosing $\Pi = \Pi_*$, we can obtain a feasible Ψ by computing the c-transform of Π_* with respect to the cost function $c(x,x',y) := \frac{|x-x'|^2}{2\tau} + V(x,y)$:

$$\Psi_*(x',y) := \inf_{x\in X} \frac{|x-x'|^2}{2\tau} + V(x,y) - \Pi_*(x).$$

For τ sufficiently small, and for each $(x',y)\in X\times Y$, the expression in the infimum is a uniformly convex function of x (thanks to the assumed smoothness of Π_*). So a sufficient condition for $\hat{x}$ to attain the infimum is that it satisfies:

$$(\hat{x}-x')/\tau + \nabla_X V(\hat{x},y) - \nabla_* \Pi(\hat{x}) = 0. \tag{112}$$

Now, for (x',y) in the support of $\rho_0 = (\mathrm{id}, T_*)_\sharp\mu$, (108), one has $y = \nabla_X V(x',\cdot)^{-1}(\nabla\Pi_*(x'))$, so choosing $\hat{x} = x'$ verifies (112). Thus, for ρ^{n-1}-a.e. (x',y), Ψ_* is given by:

$$\Psi_*(x',y) = V(x',y) - \Pi_*(x'). \tag{113}$$

The dual score then reads:

$$\int_X \Pi_*\,\mathrm{d}\mu + \int_{X\times Y} \Psi_*\,\mathrm{d}\rho^{n-1} = \int_X \Pi_*\,\mathrm{d}\mu + \int_{X\times Y} (V-\Pi_*)\,\mathrm{d}\rho^{n-1} = \int_{X\times Y} V\,\mathrm{d}\rho^{n-1}, \tag{114}$$

which matches the primal score (110) for $\gamma_* := (\mathrm{id},\mathrm{id},T)_\sharp\mu$. Hence we obtained a primal-dual certificate, showing that for τ sufficiently small ρ_0 is a minimizer of (110) and therefore a stationary point of the minimizing movement scheme (109).

8. Numerical simulations

In this Section we showcase with numerical simulations the main theoretical results discussed above. Our numerical experiments are based on solving the minimizing movement scheme (38), or more precisely its formulation as a convex optimization problem (39), using the Chambolle-Pock primal-dual method [16]. Our approach differs from traditional numerical approaches to simulate multiphase porous media flows (see for instance [19, 3, 23]) as it builds on the variational interpretation of the problem rather than on a direct discretization of the partial differential equations (1)–(3). Let us mention the contribution [9] where an augmented Lagrangian approach extending the ideas

of [5] to the context of multiphase porous media flows was compared to the more classical (and non-variational) phase-by-phase upstream mobility finite volume scheme [23].

To keep the figures intuitive and simple we fix $X = Y = [0, 1]$, $V(x, y) = |x - y|^2/2$, and $\mu = \mathrm{d}x$ the Lebesgue measure on X. We set $\kappa = 0.01$. Unless otherwise stated we discretize μ to a set of $M = 256$ equispaced points with equal weights in $[0, 1]$. We set the second marginal ν likewise to a sum of $N \in \{4, 16, 64, 256\}$ equispaced dirac deltas with equal weights. For the initialization ρ^0 we choose either the *product* initialization $\rho^0 = \mu \otimes \nu$ or the *flipped* initialization $\rho^0 = (\mathrm{id}, 1 - \mathrm{id})_\sharp \mu$, each discretized appropriately.

We fix the timestep to $\tau = 0.25$. For smaller values of τ we encounter discretization artefacts that hinder the convergence of the JKO step: When X is discretized, the minimal cost for moving a particle horizontally becomes quantized, since it must move by at least one pixel, and it may therefore be more expensive than the potential gain by following a small X-gradient of V, especially when τ is very small. To alleviate this issue one could use entropic regularization in the transport term $\mathbf{W}_F$ (see e.g. [15]); however in this work we prefer to illustrate faithfully the behavior of the minimizing movement without introducing additional terms.

Figure 1 shows an example of the first iterations in the minimizing scheme for a product initialization and $N = 4$. Note how the accumulation of mass of species 1 at the left boundary and of species 4 at the right boundary makes up for the gap left by the other species.

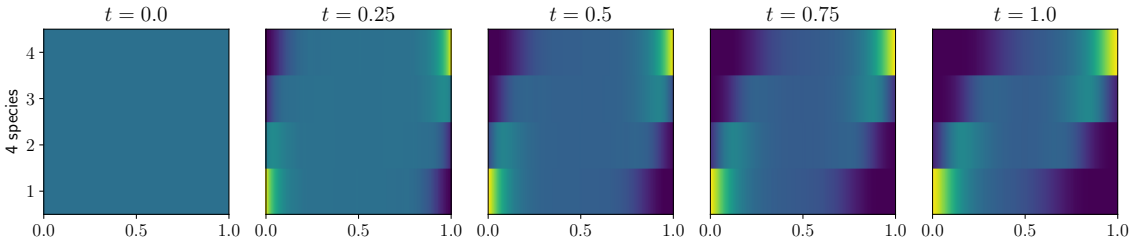

FIGURE 1. Discrete trajectory $\bar{\rho}_\tau$ for $N = 4$ and product initialization. Each band represents a phase, with hue indicating density.

Figure 2 shows a comparison of the same evolution for several choices of the number of species N. It is interesting to note how the evolution is consistent for the several values of N (i.e. convergence is achieved in approximately the same time), and the intermediate steps in the evolution are strikingly similar, in agreement with the stability of weak solutions shown in Lemma 13.

An analogous observation can be made for Figures 3 and 4. Here the initialization is a flipped plan, which constitutes a worse initial score for $\mathcal{E}$ than the product coupling from Figures 1 and 2. This explains the longer time to convergence; however the final coupling remains essentially identical to the final coupling for the product initialization. Note that as $N \to \infty$ this example tends to a trajectory which does not fulfill our assumptions since the initial entropy $\mathcal{H}(\rho^0)$ tends to ∞ as $N \to \infty$. Nevertheless, the convergence to the global optimizer still appears to hold.

Convergence to the minimizer of $\mathcal{E}$ is examined in greater detail in Figures 5 and 6. Figure 5 shows the final discrete iterate (corresponding to $t = 10$) under the two studied initializations and several choices of N with the corresponding optimizer of $\mathcal{E}$, namely ρ_*, obtained with the Sinkhorn algorithm. There appar to be no discernable differences in the images. For a quantitative analysis,

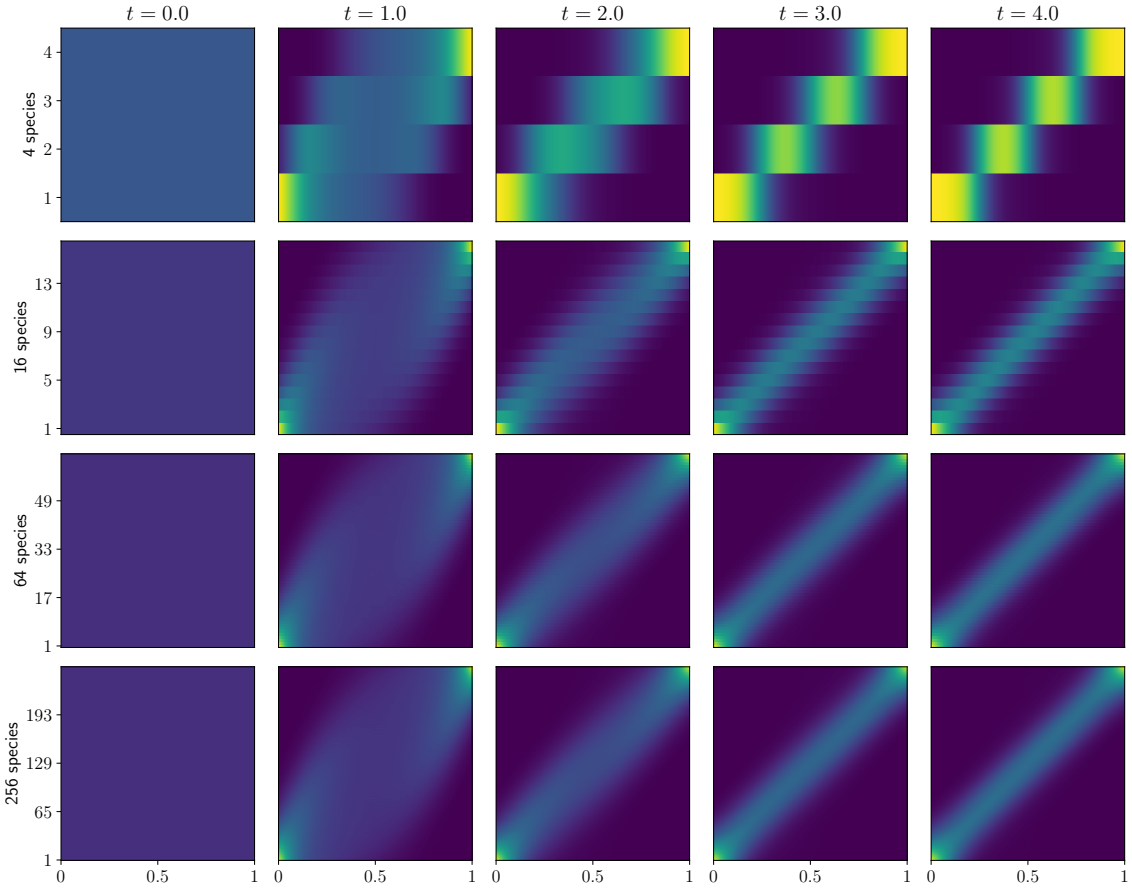

FIGURE 2. Comparison of discrete trajectories for different number of species N and product initialization. Note how the trajectories tend to a consistent limit as N increases.

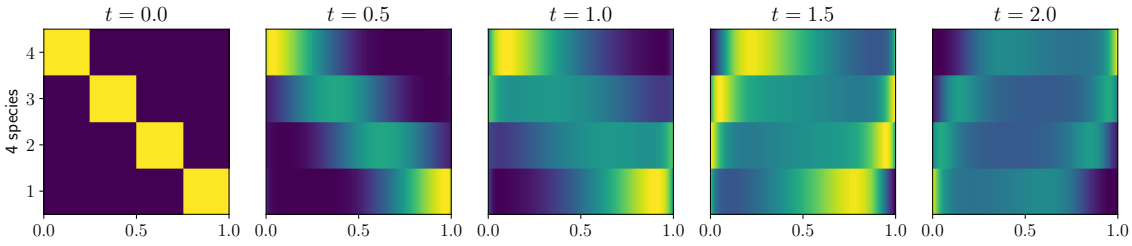

FIGURE 3. Discrete trajectory $\bar{\rho}_\tau$ for $N = 4$ and flipped initialization.

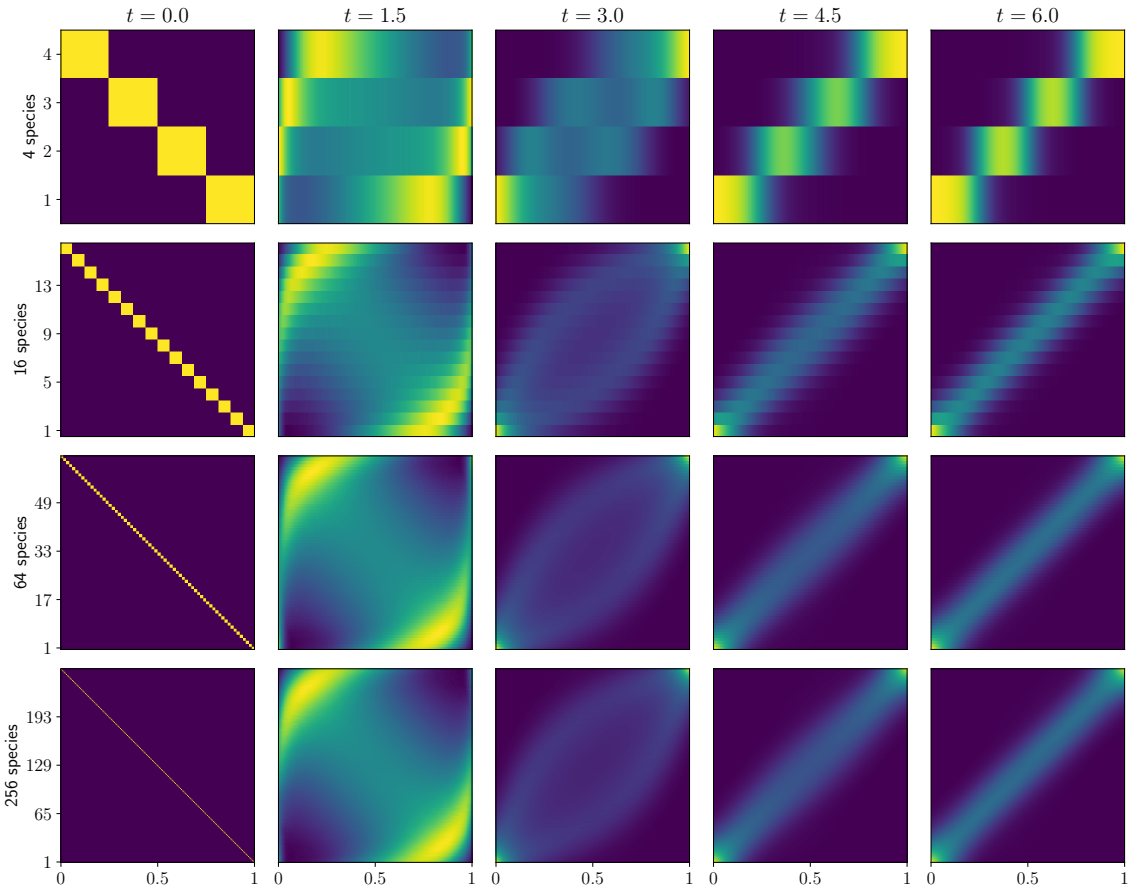

FIGURE 4. Comparison of discrete trajectories for different number of species N and flipped initialization. Note how the trajectories again tend to a consistent limit, even when the initial conditions have unbounded entropy as $N \to \infty$

in Figure 6 we plot the relative score difference

$$\Delta\mathcal{E}(\rho_t) := \frac{\mathcal{E}(\rho_t) - \mathcal{E}(\rho_*)}{\mathcal{E}(\rho_*)} \geq 0 \tag{115}$$

with respect to the elapsed time t, for the two initializations and either $M = N = 64$ or $M = N = 256$. There are two features of interest in the curves. The first is that $\Delta\mathcal{E}$ decreases until it reaches a plateau, which is encountered first for lower resolutions. This is due to discretization artefacts, as mentioned at the beginning of this section. The second interesting feature is that, even though the flipped initialization takes a longer time to converge (as it has a higher initial score), the final score is approximately independent on the initialization. Furthermore, the convergence rate seems to be independent of the initialization and approximately linear, indicated by the straight, parallel lines Figure 6.

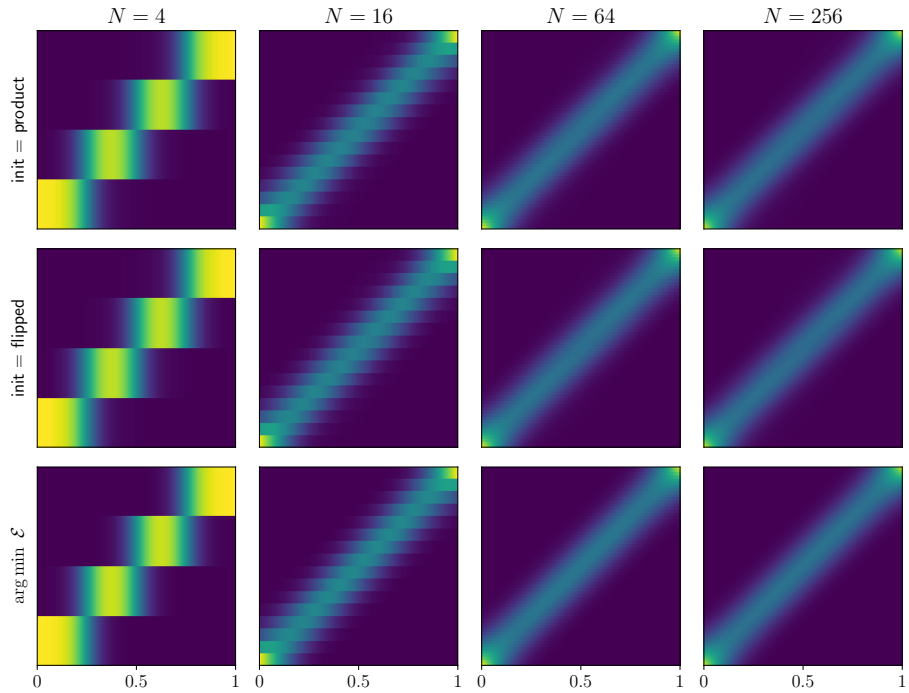

FIGURE 5. Comparison of the final discrete iterate (corresponding to $t = 10$) and the minimizer of $\mathcal{E}$.

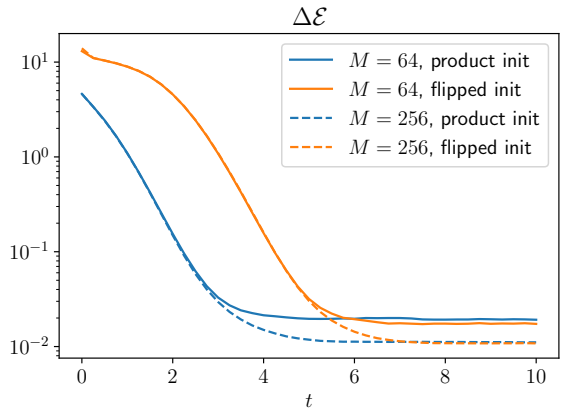

FIGURE 6. Evolution of the relative difference between $\mathcal{E}(\rho_t)$ and $\mathcal{E}(\rho_*)$.

Finally, Figures 7 and 8 show a trajectory for a μ with discontinuous density, hence not satisfying our Fisher information bound assumption. Concretely, we use the density

$$\mu(x) := \begin{cases} a & \text{for } x \in [0,1] \setminus [0.5-\delta, 0.5+\delta], \\ b & \text{for } x \in [0.5-\delta, 0.5+\delta]. \end{cases} \tag{116}$$

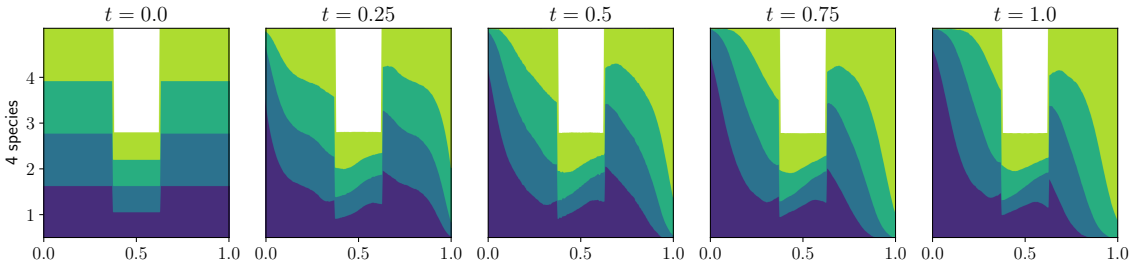

FIGURE 7. Discrete trajectory $\bar{\rho}_\tau$ for μ featuring a bottleneck, $N = 4$ and product initialization. We change the visualization with respect to previous figures to showcase better the effect of the bottleneck on the density of the different phases. Here the width the different bands represents the density of the respective species. Their combined densities add up to μ, which is given by (116).

with $\delta = 1/8$, $b = a/2$ and a chosen so that $\|\mu\| = 1$ This creates a bottleneck with reduced capacity for mass transport in the middle region. The discontinuity in μ is reproduced by each slice $r(\cdot, y)$ and we observe congestion at the boundaries of the bottleneck region; however the relative density $r(\cdot, y)/\mu(\cdot)$ remains continuous for all times. Again, the trajectories show a consistent behavior for all values of N considered, which may indicate that the Fisher information bound hypothesis could be further relaxed.

In conclusion, the numerical simulations of this section exemplify the stability of the weak solutions of (69)-(70) as the second marginal ν tends to a continuum, as well as the asymptotic convergence to the minimizer of the energy $\mathcal{E}$ as time goes to infinitiy. Moreover, there is some numerical evidence that the convergence rate may be linear. The simulations provide interesting insights on how a continuum of particles (the species $y \in Y$) find an optimal arrangement of positions (the points $x \in X$) by local exchanges, subject to a volume constraint, that may include obstructions and bottlenecks.

APPENDIX A. A TECHNICAL LEMMA

Lemma 17. *Let $\gamma \in \mathcal{P}(X \times Y)$ with respective marginals $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be given, and assume that γ possesses a density $u \in L^1(X \times Y, \mu \otimes \nu)$ with respect to $\mu \otimes \nu$. Further, let $f \in L^1(X \times Y, \gamma)$ have the following property: for every $\theta \in L^\infty(X \times Y, \mu \otimes \nu)$ satisfying*

$$\begin{aligned} \int_X f(x,y)\,\theta(x,y)u(x,y)\,\mathrm{d}\mu(x) = 0 \quad &\text{for } \nu\text{-a.e. } y \in Y, \\ \int_Y f(x,y)\,\theta(x,y)u(x,y)\,\mathrm{d}\nu(y) = 0 \quad &\text{for } \mu\text{-a.e. } x \in X, \end{aligned} \tag{117}$$

it follows that

$$\int_{X\times Y} f(x,y)\,\theta(x,y)\,\mathrm{d}\gamma(x,y) = 0. \tag{118}$$

Then there are $P \in L^1(X, \mu)$ and $Q \in L^1(Y, \nu)$ such that $f(x, y) = P(x) + Q(y)$ for γ-a.e. $(x, y) \in X \times Y$.

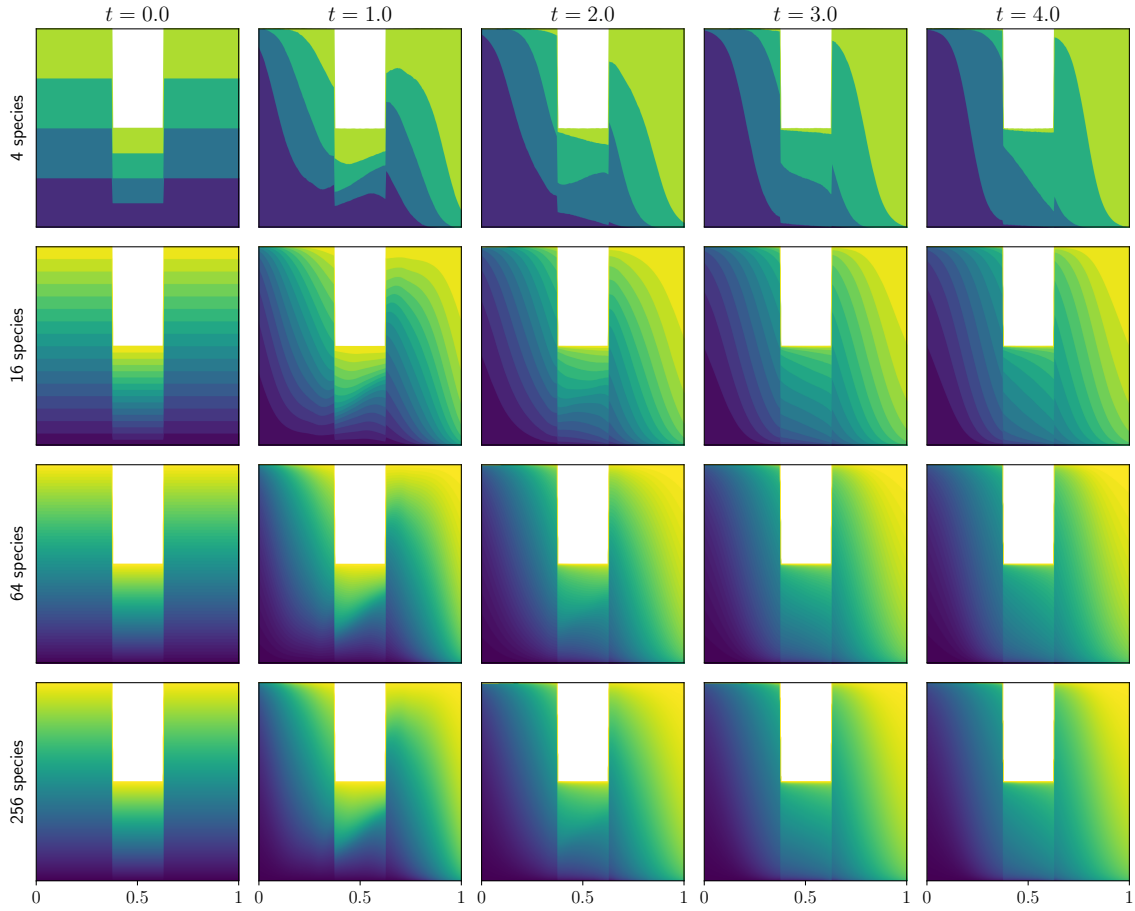

FIGURE 8. Comparison of discrete trajectories for μ featuring a bottleneck, product initialization and different number of species N. Note how the trajectories again tend to a consistent limit, even when μ has (in the continuum limit) unbounded Fisher information.

Proof. This is an application of the Hahn-Banach separation theorem. Define

$$U := \{P(x) + Q(y) \,|\, P \in L^1(X, \mu),\, Q \in L^1(Y, \nu)\}.$$

This is a subspace of $L^1(X \times Y, \gamma)$, since

$$\begin{aligned}
&\int_{X\times Y} |P(x) + Q(y)| \,\mathrm{d}\gamma(x,y) \\
&\leq \int_X |P(x)| \left(\int_Y u(x,y) \,\mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) + \int_Y |Q(y)| \left(\int_X u(x,y) \,\mathrm{d}\mu(x) \right) \mathrm{d}\nu(y) \\
&\quad = \|P\|_{L^1(X,\mu)} + \|Q\|_{L^1(Y,\nu)}.
\end{aligned}$$

Further, the subspace U is closed since convergence of f_n with $f_n(x,y) = P_n(x) + Q_n(y)$ in $L^1(X \times Y, \gamma)$ is equivalent to convergence of P_n and Q_n in the respective Banach spaces $L^1(X,\mu)$ and $L^1(Y,\nu)$.

Fix some $f_* \in L^1(X \times Y, \gamma)$ that is *not* in U; the claim is proven if we can verify the existence of some $\theta_* \in L^\infty(X \times Y, \mu \otimes \nu)$ that satisfies (117), but *not* (118) with $f = f_*$. Since U is closed, f_* has a positive distance to U, and the Hahn-Banach theorem thus guarantees the existence of a continuous linear functional $\varphi_* : L^1(X \times Y, \gamma) \to \mathbb{R}$ with $\varphi_*(U) = \{0\}$ and $\varphi_*(f_*) = 1$. By L^1-L^∞ duality, there exists some $\theta_* \in L^\infty(X \times Y, \mu \otimes \nu)$ such that

$$\varphi_*(f) = \int_{X\times Y} f\,\theta_*\,\mathrm{d}\gamma \quad \text{for all } f \in L^1(X \times Y, \gamma).$$

Further, $\varphi_*(U) = \{0\}$ implies for $f(x,y) := a(x) + b(y)$ with arbitrary $a \in L^\infty(X,\mu)$ and $b \in L^\infty(Y,\nu)$ that

$$0 = \varphi_*(f) = \int_X a(x) \left(\int_Y \theta_*(x,y) u(x,y)\,\mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) + \int_Y b(y) \left(\int_X \theta_*(x,y) u(x,y)\,\mathrm{d}\mu(x) \right) \mathrm{d}\nu(y).$$

Hence θ_* satisfies the conditions (117). On the other hand, one has

$$1 = \varphi_*(f_*) = \int_{X\times Y} f_*\,\theta_*\,\mathrm{d}\gamma,$$

i.e., f_* does *not* satisfy (118) for the choice $\theta = \theta_*$. □

Remark 6. In case that $\gamma = \mu \otimes \nu$ is a product measure, i.e., $u \equiv 1$, one does not need the abstract machinery of the Hahn-Banach theorem. Then, for any given $f_* \in L^1(X \times Y, \gamma)$ that is not of the form $f(x,y) = P(x) + Q(y)$, a suitable test function θ_* satisfying (117) but not (118) for $f = f_*$ can be obtained very explicitly. The generalization of that construction to γ that are not of product form is unclear.

Acknowledgements. CC acknowledges partial support from the French National Research Agency through LabEx CEMPI (ANR-11-LABX-0007) and project MATHSOUT of the PEPR Mathematics in Interaction (ANR-23-EXMA-0010). DM's research was supported by the DFG Collaborative Research Center TRR 109, "Discretization in Geometry and Dynamics". IM and BS were supported by the Emmy Noether programme of the DFG.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] Sigurd Angenent, Steven Haker, and Allen Tannenbaum. Minimizing flows for the Monge-Kantorovich problem. *SIAM J. Math. Anal.*, 35(1):61–97, 2003.

[3] K. Aziz and A. Settari. *Petroleum Reservoir Simulation*. Elsevier Applied Science Publishers, Londres, 1979.

[4] J. Bear and Y. Bachmat. *Introduction to modeling of transport phenomena in porous media*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.

[5] J.-D. Benamou, G. Carlier, and M. Laborde. An augmented Lagrangian approach to Wasserstein gradient flows and applications. In *Gradient flows: from theory to application*, volume 54 of *ESAIM Proc. Surveys*, pages 1–17. EDP Sci., Les Ulis, 2016.

[6] Mauro Bonafini and Bernhard Schmitzer. Domain decomposition for entropy regularized optimal transport. *Numerische Mathematik*, 149:819–870, 2021.

[7] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[8] Yann Brenier. Optimal transport, convection, magnetic relaxation and generalized Boussinesq equations. *Journal of Nonlinear Science*, 19(5):547–570, 10 2009.

[9] C. Cancès, T. O. Gallouët, M. Laborde, and L. Monsaingeon. Simulation of multiphase porous media flows with minimizing movement and finite volume schemes. *European J. Appl. Math.*, 30(6):1123–1152, 2019.
[10] C. Cancès, T. O. Gallouët, and L. Monsaingeon. The gradient flow structure of immiscible incompressible two-phase flows in porous media. *C. R. Acad. Sci. Paris Sér. I Math.*, 353:985–989, 2015.
[11] C. Cancès, T. O. Gallouët, and L. Monsaingeon. Incompressible immiscible multiphase flows in porous media: a variational approach. *Anal. PDE*, 10(8):1845–1876, 2017.
[12] C. Cancès, D. Matthes, and F. Nabet. A two-phase two-fluxes degenerate Cahn-Hilliard model as constrained Wasserstein gradient flow. *Arch. Ration. Mech. Anal.*, 233(2):837–866, 2019.
[13] C. Cancès and A. Zurek. A convergent finite volume scheme for dissipation driven models with volume filling constraint. *Numer. Math.*, 151:279–328, 2022.
[14] C. Cancès and D. Matthes. Construction of a two-phase flow with singular energy by gradient flow methods. *J. Diff. Equ.*, 344:439–470, 2023.
[15] Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.
[16] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.
[17] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169:671–691, 2016.
[18] Z. Chen. Degenerate two-phase incompressible flow. I. Existence, uniqueness and regularity of a weak solution. *J. Diff. Equ.*, 171(2):203–232, 2001.
[19] Z. Chen and R. E. Ewing. Fully discrete finite element analysis of multiphase flow in groundwater hydrology. *SIAM J. Numer. Anal.*, 34(6):2228–2253, 1997.
[20] D. Cordoba, D. Faraco, and F. Gancedo. Lack of uniqueness for weak solutions of the incompressible porous media equation. *Arch. Ration. Mech. An.*, 200:725–746, 2011.
[21] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2292–2300, 2013.
[22] D. E. Edmunds and B. Opic. Weighted Poincaré and Friedrichs inequalities. *Journal of the London Mathematical Society*, s2-47(1):79–96, 02 1993.
[23] R. Eymard, R. Herbin, and A. Michel. Mathematical study of a petroleum-engineering scheme. *M2AN Math. Model. Numer. Anal.*, 37(6):937–972, 2003.
[24] C. Förster and L. Székelyhidi. Piecewise constant subsolutions for the Muskat problem. *Commun. Math. Phys.*, 363:1051–1080, 2018.
[25] Wilfrid Gangbo and Robert J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.
[26] N. Gigli and F. Otto. Entropic Burgers' equation via a minimizing movement scheme based on the Wasserstein metric. *Calc. Var. Partial Differential Equations*, 47(1-2):181–206, 2013.
[27] Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem. *SIAM Journal on Control and Optimization*, 59(4):2428–2453, 2021.
[28] Philippe Laurençot and Bogdan-Vasile Matioc. Self-similarity in a thin film Muskat problem. *SIAM Journal on Mathematical Analysis*, 49(4):2790–2842, 2017.
[29] Hugo Lavenant, Stephen Zhang, Young-Heon Kim, and Geoffrey Schiebinger. Towards a mathematical theory of trajectory inference. *Ann. Appl. Probab.*, 34(1A), 2024.
[30] F. Otto and W. E. Thermodynamically driven incompressible fluid mixtures. *J. Chem. Phys.*, 107(23):10177–10184, 1997.
[31] Felix Otto. Evolution of microstructure in unstable porous media flow: a relaxational approach. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 52(7):873–915, 1999.
[32] Jan Peszek and David Poyato. Heterogeneous gradient flows in the topology of fibered optimal transport. *Calculus of Variations and Partial Differential Equations*, 62(9):258, Oct 2023.
[33] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.
[34] Riccarda Rossi and Giuseppe Savaré. Tightness, integral equicontinuity and compactness for evolution problems in Banach spaces. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 2(2):395–431, 2003.

[35] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser Boston, 2015.
[36] L. Székelyhidi Jr. Relaxation of the incompressible porous media equation. *Ann. Sci. École Norm. Sup.*, 45(3):491–509, 2012.
[37] M. H. Vignal. Convergence of a finite volume scheme for an elliptic-hyperbolic system. *RAIRO Modél. Math. Anal. Numér.*, 30(7):841–872, 1996.
[38] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.

Univ. Lille, CNRS, Inria, UMR8524 – Laboratoire Paul Painlevé, F-59000 Lille
Email address: clement.cances@inria.fr

Department of Mathematics, TUM School of Computation, Information and Technology, Technical University of Munich, Germany
Email address: matthes@ma.tum.de

Institute of Computer Science, University of Göttingen, Göttingen, Germany
Email address: ismael.medina@cs.uni-goettingen.de

Institute of Computer Science, University of Göttingen, Göttingen, Germany
Email address: schmitzer@cs.uni-goettingen.de