MATHEMATICS OF COMPUTATION
https://doi.org/10.1090/mcom/3492
Article electronically published on November 19, 2019

DISCRETIZATION OF FLUX-LIMITED GRADIENT FLOWS: Γ-CONVERGENCE AND NUMERICAL SCHEMES

DANIEL MATTHES AND BENJAMIN SÖLLNER

Abstract. We study a discretization in space and time for a class of nonlinear diffusion equations with flux limitation. That class contains the so-called relativistic heat equation, as well as other gradient flows of Renyi entropies with respect to transportation metrics with finite maximal velocity. Discretization in time is performed with the JKO method, thus preserving the variational structure of the gradient flow. This is combined with an entropic regularization of the transport distance, which allows for an efficient numerical calculation of the JKO minimizers. Solutions to the fully discrete equations are entropy dissipating, mass conserving, and respect the finite speed of propagation of support.

First, we give a proof of Γ-convergence of the infinite chain of JKO steps in the joint limit of infinitely refined spatial discretization and vanishing entropic regularization. The singularity of the cost function makes the construction of the recovery sequence significantly more difficult than in the L^p-Wasserstein case. Second, we define a practical numerical method by combining the JKO time discretization with a "light speed" solver for the spatially discrete minimization problem using Dykstra's algorithm, and demonstrate its efficiency in a series of experiments.

1. Introduction

1.1. **General idea.** In the field of numerical solution of transportation problems—like the estimation of Wasserstein distances, computation of barycenters, or parameter estimation—entropic regularization has been proven to be a versatile and impressively efficient tool. Based on Cuturi's adaptation of the Sinkhorn algorithm for "lightspeed computation of optimal transport" [16], a huge variety of highly efficient methods for various current applications of transport theory have been developed; see the recent book [28] for an overview. The focus has been mainly on image and data science, but these ideas have been applied for numerical approximation of gradient flows as well; see, e.g., [9, 27]. Here, we develop this approach further to define an efficient scheme for approximation of solutions to flux-limited equations of the type

$$\partial_t \rho + \nabla \cdot \left[\rho\, a\big(\nabla h'(\rho)\big)\right] = 0, \quad \rho(0,\cdot) = \rho^0. \tag{1}$$

In that problem, the sought after solution ρ is a time-dependent probability density, either on $\Omega = \mathbb{R}^d$ with a finite second moment, or on a bounded domain $\Omega \subset \mathbb{R}^d$ with no-flux boundary conditions. The given function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is convex and superlinear, and $a : \mathbb{R}^d \to \overline{\mathbb{B}}$ is a monotone map into the closed unit ball $\overline{\mathbb{B}}$ of $\mathbb{R}^d$.

Received by the editor January 31, 2019, and, in revised form, August 7, 2019.
2010 *Mathematics Subject Classification.* Primary 35K20, 35K65, 49M29, 65M12.
This research was supported by the DFG Collaborative Research Center TRR 109 "Discretization in Geometry and Dynamics".

This implies the aforementioned flux limitation, since (1) can be considered as a transport equation with velocities $a(\nabla h'(\rho))$ of modulus less than one.

Our primary example will be Rosenau's relativistic heat equation [29],

$$\partial_t \rho = \nabla \cdot \left(\rho \frac{\nabla \rho}{\sqrt{\rho^2 + |\nabla \rho|^2}} \right), \tag{2}$$

which is (1) with $h(r) = r(\log r - 1)$ and $a(p) = (1 + |p|^2)^{-1/2} p$. This equation has been analyzed in great detail, mostly by Caselles and his collaborators; see [3, 4, 14] and the references therein. Schemes for numerical solution of (2) have been developed as well (see, e.g., [10]), however, these are very different from the approach taken here.

In the definition of the entropic regularization of (1), its discretization in space and time, and the efficient numerical implementation, we closely follow the blueprint laid out in [27] for gradient flows in the L^2-Wasserstein metric. In order to make that variational approach feasible, we require a special structure of a, namely that it can be written in the form

$$a(p) = \nabla \mathbf{C}^*(-p), \tag{3}$$

where $\mathbf{C}^*$ is the Legendre transform of a convex cost function $\mathbf{C} : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$. The flux limitation is implemented by requiring further that $\mathbf{C}$ is continuous on the closed unit ball $\overline{\mathbb{B}}$, equal to one on the boundary $\partial\mathbb{B}$, and is $+\infty$ outside of $\overline{\mathbb{B}}$. As observed by Brenier [8], the relativistic heat equation (2) fits into that framework by choosing $\mathbf{C}(v) = 1 - \sqrt{1 - |v|^2}$ for $v \in \overline{\mathbb{B}}$.

1.2. **Gradient flow structure.** With the assumption (3) on a, (1) can be considered as a gradient flow on the space $\mathcal{P}(\Omega)$ of probability measures on Ω, at least formally. We briefly recall the basic idea in a language that is suitable for formulation of our approximation later. We refer, e.g., to [1, 2, 8, 24] for further details on the variational structure of (1).

The potential of that gradient flow is the entropy functional

$$\mathcal{E}(\rho) := \int_\Omega h(\rho)\, \mathrm{d}x, \tag{4}$$

and the respective dissipation $\mathfrak{D}(\rho; q)$ for a given "tangential vector" q at $\rho \in \mathcal{P}(\Omega)$ —that is, $q \in L^1(\Omega)$ is of zero average—is defined by

$$\mathfrak{D}(\rho; q) := \inf_{q = \nabla\cdot(\rho v)} \int_\Omega \mathbf{C}(v)\rho\, \mathrm{d}x. \tag{5}$$

Here the infimum runs over all vector fields $v : \Omega \to \mathbb{R}^d$ for which $q = \nabla \cdot (\rho v)$, and equals infinity if there is no such v. The integral in (5) represents the friction resulting from the infinitesimal motion of all mass elements in ρ along the vector field v; taking the infimum over v's means that the infinitesimal mass elements move in the least dissipative way to realize the macroscopic change determined by q.

A curve $\rho : \mathbb{R}_{\geq 0} \to \mathcal{P}(\Omega)$ is of *steepest descent in $\mathcal{E}$'s landscape with respect to $\mathfrak{D}$* if at each instance $t_0 > 0$, the derivative $\partial_t \rho(t_0)$ is such that the sum

$$\mathfrak{D}\big(\rho(t_0); \partial_t \rho(t_0)\big) + \frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_0} \mathcal{E}\big(\rho(t)\big) \tag{6}$$

is minimized, i.e., the decrease in energy is optimal with respect to the induced dissipation. Assuming that $\rho(t_0)$ is smooth and positive everywhere, a straight-forward calculation shows that the minimizing $\partial_t\rho(t_0) = \nabla\cdot(\rho(t_0)v(t_0))$ is determined by the vector field $v(t_0)$ that minimizes

$$v \mapsto \int_\Omega \big[\mathbf{C}(v)\rho(t_0) + h'\big(\rho(t_0)\big)\nabla\cdot\big(\rho(t_0)v\big)\big]\,\mathrm{d}x.$$

In view of (3), this produces the evolution equation (1).

1.3. **Discretization and regularization.** To connect to the variational framework of optimal transport, we perform a time-discrete approximation of (6) in the spirit of the minimizing movement scheme [2], which is often referred to as the JKO method [17] in the context of optimal transport. For a given time step $\tau > 0$, a sequence $(\rho^n)_{n=0}^\infty$ is constructed inductively: given an approximation ρ^{n-1} of $\rho((n-1)\tau)$, i.e., the solution ρ to (1) at time $t = (n-1)\tau$, choose as approximation ρ^n of $\rho(n\tau)$ the minimizer of

$$\rho \mapsto \inf_\gamma \iint_{\Omega\times\Omega} \mathbf{c}_\tau(x,y)\,\mathrm{d}\gamma(x,y) + \frac{1}{\tau}\big[\mathcal{E}(\rho) - \mathcal{E}(\rho^{n-1})\big]. \tag{7}$$

As seen above, the infimum runs over all probability measures $\gamma \in \mathcal{P}(\Omega\times\Omega)$ on the product space $\Omega\times\Omega$ whose first and second marginal, denoted by $X\#\gamma$ and $Y\#\gamma$, respectively, is equal to ρ^{n-1} and ρ. Further, $\mathbf{c}_\tau(x,y)$ is the $\mathbf{C}$-induced cost of the transport from x to y in time τ; if Ω is convex, then simply $\mathbf{c}_\tau(x,y) = \mathbf{C}(\frac{y-x}{\tau})$, i.e., $\mathbf{c}_\tau(x,y)$ is the average dissipation induced by the motion of a unit mass element with constant velocity $v = \frac{y-x}{\tau}$. The general definition of $\mathbf{c}_\tau$ is given in Section 2.4. In the language of optimal transport, γ is a transport plan from ρ^{n-1} to ρ^n: roughly speaking, $\gamma(x,y)$ determines the amount of ρ^{n-1}'s mass at position x to be moved to ρ^n's mass at position y. The double integral in (7) is visibly an approximation of the integral in (6).

The difficulty in the numerical implementation of (7) is to calculate the infimum of the integral for given ρ^{n-1} and ρ, and its variation with respect to ρ. A common approach is to go to the Lagrangian formulation, using that the optimal γ is typically concentrated on the graph of a transport map $T : \Omega \to \Omega$. This is extremely efficient in one space dimension [7,22,23], but becomes significantly more cumbersome—and difficult to analyze—in multiple dimensions [5,12,13,18]. Various alternatives to the Lagrangian approach are available, including finite volume methods [21], blob methods [11], etc.

Here, we use the "lightspeed computation" of the optimal plan γ by employing entropic regularization to the minimization problem. Recall that γ's negative entropy is

$$\mathcal{H}(\gamma) = \iint_{\Omega\times\Omega} G(x,y)\log G(x,y)\,\mathrm{d}(x,y) \tag{8}$$

if $\gamma = G\mathcal{L}^d$ is absolutely continuous, and $\mathcal{H}(\gamma) = +\infty$ otherwise. Adding this as a regularization inside the dissipation term in (7), we arrive at the new minimization problem

$$\rho \mapsto \inf_\gamma \left[\iint_{\Omega\times\Omega} \mathbf{c}_\tau(x,y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma)\right] + \frac{1}{\tau}\big[\mathcal{E}(\rho) - \mathcal{E}(\rho_\tau^{n-1})\big], \tag{9}$$

$\varepsilon \geq 0$ being the parameter of the regularization. Finally, we discretize the problem (9) in space by restricting minimization to $\mathcal{P}_\delta(\Omega)$, the set of absolutely continuous ρ's whose densities are piecewise constant on the cells Q of a given tesselation $\mathcal{Q}_\delta$

of Ω; here $\delta > 0$ parametrizes the size of the cells Q, and $\delta \to 0$ is the continuous limit. It is further admissible to approximate $\mathbf{c}_\tau$ by a more convient cost function $\mathbf{c}_{\tau,\delta}$. E.g., in the actual numerical experiments, we use a $\mathbf{c}_{\tau,\delta}$ that is piecewise constant on the products $Q \times Q'$ of cells $Q, Q' \in \mathcal{Q}_\delta$; this makes the minimization feasible in practice since it then suffices to consider only absolutely continuous γ's that are piecewise constant on $Q \times Q'$.

In summary, for given $\varepsilon \geq 0$ and $\delta \geq 0$—corresponding to a tesselation $\mathcal{Q}_\delta$ and a cost function $\mathbf{c}_{\tau,\delta}$—a time-discrete approximation $(\rho^n)_{n=0}^\infty$ of a solution to (1) is defined inductively by

$$\rho^n := Y\#\gamma^n, \quad \text{with} \quad \gamma^n := \arg\min \mathcal{E}_{\tau,\varepsilon,\delta}\big(\gamma\big|\rho^{n-1}\big), \tag{10}$$

where, using the indicator functional ι_Q that is zero if Q is true, and $+\infty$ otherwise,

$$\mathcal{E}_{\tau,\varepsilon,\delta}\big(\gamma\big|\bar\rho\big) = \iint_{\Omega\times\Omega} \mathbf{c}_{\tau,\delta}(x,y)\,\mathrm{d}\gamma(x,y) + \varepsilon\mathcal{H}(\gamma) + \frac{1}{\tau}\mathcal{E}(Y\#\gamma) + \iota_{X\#\gamma=\bar\rho} + \iota_{Y\#\gamma\in\mathcal{P}_\delta(\Omega)}. \tag{11}$$

We remark that for $\varepsilon > 0$, the minimization problem is strictly convex, so the minimizer γ^n is unique. The situation is less clear for $\varepsilon = 0$, since uniqueness results for optimal plans in relativistic costs, like in [6], only apply if Ω is bounded. Fortunately, even for $\varepsilon = 0$, it is easily seen that any minimizer γ^n leads to one and the same density ρ^n, which is uniquely determined thanks to the strict convexity of the entropy functional $\mathcal{E}$ (and the linearity of the transport term).

1.4. **Convergence result.** Our analytical result concerns the joint limit of infinitely refined spatial discretization $\delta \to 0$ and vanishing entropic regularization $\varepsilon \to 0$.

Theorem 1. *Assume $\Omega = \mathbb{R}^d$ and that ρ^0 has finite second moment. Assume further that $h(r) = r^m$ with some $m > 1$.*

Fix a time step $\tau > 0$, and nonnegative sequences (ε_k) and (δ_k) of entropic regularizations and spatial discretizations, respectively, that converge to zero. Under hypotheses on the tesselations $\mathcal{Q}_{\delta_k}$ and cost functions $\mathbf{c}_{\tau,\delta_k}$ that are detailed in Section 2.5 *below, the inductive scheme in* (10)*, with $\varepsilon = \varepsilon_k$ and $\delta = \delta_k$, is well-defined and produces time-discrete approximations $(\rho_k^n)_{n=0}^\infty$ for each k. Moreover, $\rho_k^n \to \rho^n$ narrowly and weakly in $L^m(\mathbb{R}^d)$ as $k \to \infty$, for each n, and $(\rho^n)_{n=0}^\infty$ is a sequence of minimizers in* (7)*.*

We emphasize that the special cases $\varepsilon_k \equiv 0$ (spatial discretization without entropic regularization) and $\delta_k \equiv 0$ (entropic regularization without spatial discretization) are included. Further, we remark that the choice $\Omega = \mathbb{R}^d$ is mainly made for definiteness; the proof is actually slightly more difficult than in the case of bounded Ω. Also, $h(r) = r^m$ has been chosen to simplify the presentation; the method of proof would apply to any convex h that has superlinear growth at infinity.

The proof is based on the Γ-convergence of the functional in (11) to the one in (7) without $\mathcal{E}(\rho^{n-1})$, which is made precise in Proposition 1 below. That Γ-limit would be fairly easy to obtain in the situation of regular cost functions, i.e., when $\mathbf{C}$ is a continuous and strictly convex function on all of $\mathbb{R}^d$. In the flux limited situation that we consider here, the construction of the recovery sequence is surprisingly delicate.

We emphasize that we do not consider the passage $\tau \to 0$ from the JKO method (7) to a solution of the PDE (1). That kind of limit has been studied extensively, albeit rarely in the flux-limited case. Particularly for L^2-Wasserstein gradient flows, corresponding to $\mathbf{C}(v) = \frac{1}{2}|v|^2$ and to $a(p) = p$, the existing literature is huge and also covers much more general nonlinearities in (1) than just $h'(\rho)$. The JKO method has been used to construct solutions to linear and nonlinear Fokker–Planck equations [26], to degenerate fourth-order parabolic equations [25], to PDEs with nonlocal terms [7], to coupled systems [20], and many more. There are fewer results on a JKO-like variational approximation of (1) with a nonlinear power function $a(\xi) = |\xi|^{p-2}\xi$, with $p \neq 2$; this includes in particular the p-Laplace equations. The corresponding theory of gradient flows in the L^q-Wasserstein metric with $\mathbf{C}(v) = \frac{1}{q}|v|^q$ (with $q = p' \neq 2$) has been developed in [1,2]. Finally, concerning the situation of interest here, which is (1) with flux-limitation, the analysis is significantly more challenging in that situation, but still, the limit $\tau \to 0$ has been carried out successfully on the JKO-like variational approximation of the relativistic heat equation in a work of McCann and Puel [24]. The techniques developed therein should apply to the more general class (1) considered here. We remark that the concept of solution used in [24] is much weaker than in the situation of convex gradient flows in the L^2-Wasserstein distance [2]. For instance, uniqueness of the limit curve for $\tau \to 0$ is unknown, despite unique solvability of the minimization problem (7).

To the best of our knowledge, our result is the first one that rigorously shows the stability of the minimizers in the JKO scheme under entropic regularization. In a related problem, namely for (1) with $a(\xi) = \xi$, i.e., in the L^2-Wasserstein case, the combined limit of $\tau \to 0$ and $\varepsilon \to 0$ (without spatial discretization, $\delta = 0$) has been carried out by Carlier et al. [9]. Also there, the Γ-limit of an entropically regularized transport is studied, however in a different sense, namely for fixed marginals and for quadratic costs, both of which make the analysis much easier. We remark further that a joint limit of spatio-temporal refinement has been performed recently [19] for a structurally different fully discrete approximation of the relativistic heat equation in one space dimension, using Lagrangian maps.

2. Notation and general hypotheses

Below, we summarize several basic notation and hypotheses, most of which have been mentioned in the introduction in an informal way.

2.1. **Domains and measures.** In the proof of Theorem 1, $\Omega = \mathbb{R}^d$. In the numerical experiments, $\Omega \subset \mathbb{R}^d$ is an open, bounded and connected set with Lipschitz boundary. $\mathcal{L}^d$ is the d-dimensional Lebesgue measure on Ω.

For a measurable subset M of a euclidean space $\mathbb{R}^m$, we denote by $\mathcal{P}(M)$ the affine space of probability measures on M that have finite second moment (which is irrelevant if M is bounded). By abuse of notation, we shall frequently identify absolutely continuous $\mu = \rho\mathcal{L}^d \in \mathcal{P}(M)$ and their Lebesgue densities $\rho \in L^1(M)$.

For a measurable map $T : M \to M'$, the push-forward $T\#\mu \in \mathcal{P}(M')$ of $\mu \in \mathcal{P}(M)$ is defined via $T\#\mu[A] = \mu[T^{-1}(A)]$ for all measurable sets $A \subset M$. Canonical projections $X, Y : M \times M \to M$ are given by $X(x,y) = x$ and $Y(x,y) = y$. With these notation, the two marginals of $\gamma \in \mathcal{P}(\Omega \times \Omega)$ are given by $X\#\gamma, Y\#\gamma \in \mathcal{P}(\Omega)$, respectively.

The natural notion of convergence in $\mathcal{P}(M)$ is narrow convergence, that is weak convergence as measures in duality to bounded continuous functions $\varphi \in C_b(M)$. For $M = \mathbb{R}^m$, we shall occasionally use a slightly stronger kind of convergence, namely convergence in $\mathbf{W}_2$ (the Wasserstein distance is recalled below), which means narrow convergence plus convergence of the second moment.

2.2. **Wasserstein distance.** The L^2-Wasserstein distance between $\rho_0, \rho_1 \in \mathcal{P}(M)$ is given by

$$\mathbf{W}_2(\rho_0,\rho_1) = \left(\inf_{\gamma\in\mathcal{P}(M\times M)} \left[\iint_{M\times M} |x-y|^2 \,\mathrm{d}\gamma(x,y) + \iota_{X\#\gamma=\rho_0} + \iota_{Y\#\gamma=\rho_1} \right] \right)^{1/2}.$$

The infimum above is actually a minimum, and minimizers γ are called *optimal plans* for the transport from ρ_0 to ρ_1. We use the following fact: if ρ_0 is absolutely continuous, then there exists a measurable $T : M \to M$, called an *optimal map*, such that $T\#\rho_0 = \rho_1$, and

$$\mathbf{W}_2(\rho_0,\rho_1) = \left(\int_M |T(x)-x|^2 \rho_0(x)\,\mathrm{d}x \right)^{1/2}.$$

$\mathbf{W}_2$ is a genuine metric on $\mathcal{P}(M)$. Convergence in $\mathbf{W}_2$ is equivalent to narrow convergence and convergence of the second moment.

2.3. **Energy functional.** By abuse of notation, the definition of $\mathcal{E} : \mathcal{P}(\Omega) \to \mathbb{R}\cup\{+\infty\}$ in (4) has to be understood in the sense that if $\mu = \rho\mathcal{L}^d$ is absolutely continuous, then $\mathcal{E}(\mu) = \mathcal{E}(\rho)$ is given by the integral, and $\mathcal{E}(\mu) = +\infty$ otherwise. Since h is convex, l.s.c., and superlinear at infinity, $\mathcal{E}$ is lower semicontinuous with respect to narrow convergence.

The methods we present are suited to study general energy functionals of the form (4) with a smooth and convex function h of superlinerar growth at infinity. In the proof of Theorem 1, we restrict ourselves to $h(r) = r^m$ with $m > 1$ to facilitate readability. In the numerical experiments, we additionally use $h(r) = r(\log r - 1)$.

2.4. **Derived cost function.** We assume that $\mathbf{C} : \mathbb{R}^d \to [0,\infty]$ is strictly convex, continuous and bounded on $\overline{\mathbb{B}}$, and $+\infty$ outside of $\overline{\mathbb{B}}$, with unique minimum $\mathbf{C}(0) = 0$. For technical reasons, we further assume that $\mathbf{C} \equiv 1$ on $\partial\overline{\mathbb{B}}$. Then the gradient of the Legendre dual $\mathbf{C}^\star$ lies in $\overline{\mathbb{B}}$.

The cost function $\mathbf{c} : \Omega\times\Omega \to [0,\infty]$ is derived from $\mathbf{C}$ via

(12)
$$\mathbf{c}_\tau(x,y) = \inf\left\{ \frac{1}{\tau}\int_0^\tau \mathbf{C}(\dot z(t))\,\mathrm{d}t \,\middle|\, z : [0,\tau] \to \Omega \text{ differentiable, } z(0) = x,\ z(\tau) = y \right\}.$$

If Ω is convex (e.g., $\Omega = \mathbb{R}^d$), then thanks to the convexity of $\mathbf{C}$,

$$\mathbf{c}_\tau(x,y) = \mathbf{C}\left(\frac{y-x}{\tau}\right). \tag{13}$$

2.5. **Spatial discretization.** We assume that for each $\delta > 0$, a tesselation $\mathcal{Q}_\delta$ of Ω is given. That is, $\mathcal{Q}_\delta$ consists of finitely (if Ω bounded) or infinite-countably (if $\Omega = \mathbb{R}^d$) many open nonoverlapping cells Q such that the union of their closures $\overline{Q}$ cover Ω. We further require that there is a constant $\underline{r} > 0$ such that

$$\operatorname{diam}(Q) \le \sqrt{d}\delta \quad \text{and} \quad |Q| := \mathcal{L}^d(Q) \ge (\underline{r}\delta)^d \quad \text{for all } Q \in \mathcal{Q}_\delta. \tag{14}$$

A canonical example for $\Omega = \mathbb{R}^d$ is, setting $\underline{r} := 1$,

$$\mathcal{Q}_\delta = \big\{\delta(\{\mathbf{j}\} + K) \mid \mathbf{j} \in \mathbb{Z}^d\big\} \qquad \text{where} \qquad K := (-\tfrac{1}{2}, \tfrac{1}{2})^d\,.$$

Accordingly, we define $\mathcal{P}_\delta(\Omega)$ as the space of those $\rho\mathcal{L}^d \in \mathcal{P}(\Omega)$ for which ρ is constant on each $Q_i \in \mathcal{Q}_\delta$. Further, $\mathcal{P}_\delta(\Omega\times\Omega)$ consists of those $\gamma \in \mathcal{P}(\Omega\times\Omega)$ for which $Y\#\mathcal{P} \in \mathcal{P}_\delta(\Omega)$. We emphasize that the condition is only on the y-marginal $Y\#\gamma$, not on the x-marginal $X\#\gamma$, which does not even need to be absolutely continuous. For convenience, we set $\mathcal{P}_0(\Omega) := \mathcal{P}(\Omega)$.

For a probability density $\bar\rho \in L^1(\Omega)$, let

$$\Gamma_\delta(\bar\rho) = \big\{\gamma \in \mathcal{P}_\delta(\Omega\times\Omega)\,;\, X\#\gamma = \bar\rho\mathcal{L}^d\big\}$$

be the subset of measures with $\bar\rho\mathcal{L}^d$ as first marginal.

Moreover, we assume that for each $\delta > 0$, a function $\mathbf{c}_{\tau,\delta} : \Omega\times\Omega \to [0,\infty]$ is given that approximates the cost function $\mathbf{c}_\tau$ as follows: there are $\alpha_{\tau,\delta} \in (0,1)$ with $\alpha_{\tau,\delta} \to 0$ as $\delta \to 0$ for fixed $\tau > 0$ such that

$$|\mathbf{c}_{\tau,\delta}(x,y) - \mathbf{c}_\tau(x,y)| \le \alpha_{\tau,\delta} \quad \text{for } |x-y| \le \tau \quad \text{and} \tag{15}$$

$$\mathbf{c}_{\tau,\delta}(x,y) \ge 1 - \alpha_{\tau,\delta} + \frac{1}{\alpha_{\tau,\delta}}\big(|y-x| - \tau\big)^2 \quad \text{for } |x-y| > \tau. \tag{16}$$

Naturally, one can always take $\mathbf{c}_{\tau,\delta} \equiv \mathbf{c}_\tau$. Note that any $\mathbf{c}_{\tau,\delta}$ with $\mathbf{c}_{\tau,\delta} = +\infty$ on $|x-y| > \tau$ automatically satisfies (16).

For brevity, we write $\mathbf{c}_k$ for $\mathbf{c}_{\tau,\delta_k}$, and accordingly α_k for the constants α_{τ,δ_k} appearing in (15) and (16).

3. Proof of Theorem 1

The proof of Theorem 1 immediately follows from a Γ-convergence result that we formulate below.

Proposition 1. *In addition to the hypotheses of Theorem 1, let a sequence $(\rho_k)_{k=1}^\infty$ of densities $\rho_k \in \mathcal{P}_2(\mathbb{R}^d)$ be given such that ρ_k converges in $\mathbf{W}_2$ to some $\rho_* \in \mathcal{P}_2(\mathbb{R}^d)$, and $\sup_k \mathcal{E}(\rho_k) < \infty$. Let, furthermore, $\delta_k > 0$ be a sequence tending to zero slowly enough such that*

$$\varepsilon_k \log(\delta_k^{-1}) \to 0 \tag{17}$$

holds. Then the sequence of functionals $\mathcal{E}_k^\tau : \mathcal{P}(\Omega\times\Omega) \to [0,+\infty]$ with (cf. (11))

$$\mathcal{E}_k^\tau(\gamma) := \mathcal{E}^\tau_{\varepsilon_k,\delta_k,\mathbf{c}_k}(\gamma|\rho_k)$$

Γ-converges in the narrow topology to $\mathcal{E}_*^\tau : \mathcal{P}(\Omega\times\Omega) \to [0,+\infty]$ with

$$\mathcal{E}_*^\tau(\gamma) = \iint_{\Omega\times\Omega} \mathbf{c}_{\tau,\delta}(x,y)\,\mathrm{d}\gamma(x,y) + \frac{1}{\tau}\mathcal{E}(Y\#\gamma) + \iota_{X\#\gamma=\rho_*}.$$

Moreover, each $\mathcal{E}_k^\tau$ possesses a (unique, if $\varepsilon_k > 0$) minimizer $\hat\gamma_k \in \Gamma_{\delta_k}(\rho_k)$, and a subsequence of these minimizers converges in $\mathbf{W}_2$ to a minimizer $\hat\gamma \in \Gamma(\rho)$ of $\mathcal{E}^\tau(\cdot;\rho_*)$.

Remark 1. Note that (17), which is needed for the construction of the recovery sequence in Section 4.3, imposes no additional restriction if the tesselation $\mathcal{Q}_\delta$ is made of identical cubes, since then $\Gamma_{\delta'}(\tilde\rho) \subseteq \Gamma_\delta(\tilde\rho)$ if $\delta' > 0$ is an integer multiple of $\delta > 0$, or is arbitrary if $\delta = 0$, (recall that the additional condition induced by δ is only on the Y-marginal, not on the X-marginal), and we can replace (δ_k)

by a sequence (δ_k') with $\delta_k' \geq \delta_k$ that still goes to zero and satisfies (17), and the recovery sequence $\gamma_k \in \Gamma_{\delta_k'}(\rho_k)$ that we obtain is clearly also a recovery sequence with $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$.

It is now easy to conclude Theorem 1 by induction on n. Trivially, $\rho_k^0 = \rho^0$ converges to $\rho_*^0 = \rho^0$. Assume that for some $n = 1, 2, \ldots$ there is a (nonrelabeled) subsequence $(\rho_k^{n-1})_{k=1}^\infty$ that converges in $\mathbf{W}_2$ and weakly in $L^m(\mathbb{R}^d)$ to a limit ρ_*^{n-1}. That sequence $(\rho_k^{n-1})_{k=1}^\infty$ satisfies the hypotheses of Proposition 1, since weak convergence in $L^m(\mathbb{R}^d)$ implies that $\mathcal{E}(\rho_k^{n-1}) = \|\rho_k^{n-1}\|_{L^m}^m$ remains bounded. Hence the respective functionals $\mathcal{E}_k^\tau$ with $\rho_k := \rho_k^{n-1}$ Γ-converge narrowly to $\mathcal{E}_*^\tau$, with ρ_*^{n-1} in place of ρ_*, and a (nonrelabeled) subsubsequence $(\gamma_k^n)_{k=1}^\infty$ of the minimizers converges to a limit γ_*^n in $\mathbf{W}_2$. It is obvious that $\rho_*^n := Y\#\gamma_*^n$ is a minimizer in (7). It is further obvious that for the subsubsequence under consideration, the convergence of γ_k^n in $\mathbf{W}_2$ is inherited by the marginal ρ_{k-1}^n. Finally, to conclude the weak convergence in $L^m(\mathbb{R}^d)$, possibly after passing to yet another subsequence, observe that the γ_k^{n-1} are minimizers of the respective $\mathcal{E}_k^\tau$, that $\|\rho_k\|_{L^m}^m = \mathcal{E}(\rho_k) \leq \mathcal{E}_k^\tau(\gamma_k^{n-1})$ by the definition of $\mathcal{E}_{\varepsilon,\delta,\mathbf{c}}^\tau$, and that $\mathcal{E}_k^\tau$ Γ-converges to $\mathcal{E}_*^\tau$. Alaoglou's theorem allows us to select a subsequence that converges weakly in $L^m(\mathbb{R}^d)$.

Note that above we have used that some subsequence of the γ_k^n converges to a minimizer γ_*^n of $\mathcal{E}_*^\tau$. However, since the respective marginal $\rho_*^n = Y\#\gamma_*^n$ is a global minimizer in (7), it is uniquely determined by ρ_*, thanks to the strict convexity of $\mathcal{E}$. Therefore, no matter which convergent subsequence of $(\gamma_k^n)_{k=1}^\infty$ is chosen, the respective $\rho_k^n = Y\#\gamma_k^n$ all converge to the same limit, implying convergence of the entire sequence $(\rho_k^n)_{k=1}^\infty$.

The rest of the analytical part of this paper is devoted to proving Proposition 1.

4. Proof of Proposition 1

Throughout the proof, let a sequence $(\rho_k)_{k=1}^\infty$ be fixed that satisfies the hypotheses of Proposition 1, i.e., $\rho_k \in \mathcal{P}_2(\mathbb{R}^d)$, $\sup_k \mathcal{E}(\rho_k) < \infty$, and $\rho_k \to \rho_*$ in $\mathbf{W}_2$.

The proof is divided into three steps. First, we prove the *liminf condition* for Γ-convergence: if $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$ converges to $\gamma_* \in \Gamma(\rho_*)$ narrowly, then

$$\mathcal{E}_*^\tau(\gamma_*) \leq \liminf_{k\to\infty} \mathcal{E}_k^\tau(\gamma_k). \tag{18}$$

Second, and by far more difficult, is the construction of a *recovery sequence*: if $\gamma_* \in \Gamma(\rho_*)$ is such that $\mathcal{E}_*^\tau(\gamma_*) < \infty$, then there are $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$ such that $\gamma_k \to \gamma_*$ narrowly, and

$$\mathcal{E}_*^\tau(\gamma_*) \geq \limsup_{k\to\infty} \mathcal{E}_k^\tau(\gamma_k). \tag{19}$$

These two steps together verify the Γ-convergence of the $\mathcal{E}_k^\tau$. In particular, it follows that if $\hat\gamma_k$ are minimizers of the $\mathcal{E}_k^\tau$ which converge to $\hat\gamma \in \Gamma(\rho_*)$, then $\hat\gamma$ is a minimizer of $\mathcal{E}_*^\tau$. Now, in the final step, we verify that each $\mathcal{E}_k^\tau$ actually possesses a minimizer $\hat\gamma_k \in \Gamma_{\delta_k}(\rho_k)$, and that a subsequence of those converges narrowly to a limit $\hat\gamma \in \Gamma(\rho_*)$, which then is necessarily a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.

4.1. **Preliminary results.** Before starting with the core of the proof, we draw two immediate conclusions from the hypotheses stated above.

Lemma 1. *The ρ_k have k-uniformly bounded second moments, and*

$$\int_{\mathbb{R}^d} \rho_k(x) \log \rho_k(x)\,\mathrm{d}x$$

is k-uniformly bounded from above and below.

Proof. By hypothesis, ρ_k converges to ρ_* in $\mathbf{W}_2$, which implies in particular the convergence of ρ_k's second moment to the one of ρ_*. Boundedness of the integral is obtained by means of a classical estimate: first, observe that $r \log r \geq -\frac{d+1}{e} r^{\frac{d}{d+1}}$ for all $r > 0$. By Hölder's inequality, it follows that

$$\begin{aligned}\int_{\mathbb{R}^d} \rho(x) \log \rho(x)\,\mathrm{d}x &\geq -\frac{d+1}{e} \int_{\mathbb{R}^d} \rho(x)^{\frac{d}{d+1}}\,\mathrm{d}x \\ &\geq -\frac{d+1}{e} \left(\int_{\mathbb{R}^d} \frac{\mathrm{d}x}{\big(1+|x|^2\big)^d} \right)^{\frac{1}{d+1}} \left(\int_{\mathbb{R}^d} \rho(x)\big(1+|x|^2\big)\,\mathrm{d}x \right)^{\frac{d}{d+1}},\end{aligned}$$

which yields a finite lower bound that only depends on the second moment of ρ_k. An upper bound easily follows from the k-uniform boundedness of $\mathcal{E}(\rho_k)$ and the fact that $r \log r \leq \frac{1}{(m-1)e} r^m$ for all $r > 0$. □

For the next result, recall that $\alpha_k = \alpha_{\tau,\delta_k}$ are the quantities that appear in conditions (15) and (16).

Lemma 2. *There is a constant C such that—uniformly for all k large enough—the second moment of each $\gamma \in \Gamma(\rho_k)$ is controlled via*

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} \big(|x|^2 + |y|^2\big)\,\mathrm{d}\gamma(x,y) \leq C\big(1 + \alpha_k \mathcal{E}_k^\tau(\gamma)\big), \tag{20}$$

and $\mathcal{E}_k^\tau$ is bounded from below as follows:

$$\mathcal{E}_k^\tau(\gamma) \geq (\tau - C\alpha_k\varepsilon_k) \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma + \mathcal{E}(Y\#\gamma). \tag{21}$$

In particular, $\mathcal{E}_k^\tau$ is nonnegative for all sufficiently large k such that $C\alpha_k\varepsilon_k \leq \tau$.

Proof. On the one hand, with the same idea as in the proof of Lemma 1 above, we find for every $\gamma = G\mathcal{L}^d \otimes \mathcal{L}^d$ that

$$\mathcal{H}(\gamma) \geq -C_d \left(1 + \iint_{\mathbb{R}^d \times \mathbb{R}^d} \big(|x|^2 + |y|^2\big)\,\mathrm{d}\gamma(x,y) \right),$$

where

$$C_d := \frac{2d+1}{e} \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\mathrm{d}(x,y)}{\big(1+|x|^2+|y|^2\big)^{2d}} \right)^{\frac{1}{2d+1}}$$

is a finite constant that only depends on d. On the other hand, using hypothesis (16) on $\mathbf{c}_k$, it follows that

$$\begin{aligned}\iint_{\mathbb{R}^d \times \mathbb{R}^d} |y|^2\,\mathrm{d}\gamma(x,y) &\leq \iint_{\mathbb{R}^d \times \mathbb{R}^d} \big[|x| + \big(|y-x| - \tau\big)\mathbf{1}_{|y-x|\geq\tau} + \tau\big]^2\,\mathrm{d}\gamma(x,y) \\ &\leq 2\iint_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2\,\mathrm{d}\gamma(x,y) + 4\tau^2 \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{d}\gamma(x,y) + 4\iint_{|y-x|\geq\tau} \big(|x-y| - \tau\big)^2\,\mathrm{d}\gamma(x,y) \\ &\leq 2\int_{\mathbb{R}^d} |x|^2 \rho_k(x)\,\mathrm{d}x + 4\tau^2 + 4\alpha_k \iint_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma,\end{aligned}$$

which yields

(22)
$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x|^2+|y|^2\right)\mathrm{d}\gamma(x,y) \le 4\left[\tau^2+\int_{\mathbb{R}^d}|x|^2\rho_k(x)\,\mathrm{d}x+\alpha_k\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{c}_k\,\mathrm{d}\gamma\right].$$

In view of Lemma 1 above, the second moment of ρ_k is uniformly controlled, and therefore

$$\mathcal{H}(\gamma)\ge -C\left(1+\alpha_k\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{c}_k\,\mathrm{d}\gamma\right), \tag{23}$$

with a k-independent C. This induces the bound (21). The other bound (20) follows for all k such that, say, $C\alpha_k\varepsilon_k\le\tau/2$, by reinserting (21) into (22) and once again using the uniform bound on ρ_k's second moment. □

4.2. **Liminf condition.**

Proposition 2. *Assume that a sequence of measures $\gamma_k\in\Gamma(\rho_k)$ converges narrowly to $\gamma_*\in\Gamma(\rho)$. Then* (18) *holds.*

Proof. Recall from Lemma 2 that $\mathcal{E}_k^\tau$ is nonnegative for k large enough, and if $\mathcal{E}_k^\tau(\gamma_k)\to+\infty$, there is nothing to prove. Hence, it suffices to consider a sequence (γ_k) such that $\mathcal{E}_k^\tau(\gamma_k)$ converges to a finite value. From (21), one directly concludes k-uniform boundedness of $\iint\mathbf{c}_k\,\mathrm{d}\gamma_k$. Thanks to the bound (16) on $\mathbf{c}_k$, it follows for every $t>0$ that γ_k's mass in $|x-y|\ge\tau+t$ goes to zero as $k\to\infty$. Thus, γ_* is supported in $|x-y|\le\tau$.

Define the continuous function $\tilde{\mathbf{c}}:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ by $\hat{\mathbf{c}}(x,y)=\mathbf{c}(x,y)$ for $|x-y|\le\tau$, and $\hat{\mathbf{c}}\equiv 1$ otherwise. From (15) and (16) it is clear that $\mathbf{c}_k\ge\hat{\mathbf{c}}-\alpha_k$, and so

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{c}_k\,\mathrm{d}\gamma_k &\ge \iint_{\mathbb{R}^d\times\mathbb{R}^d}\left(\hat{\mathbf{c}}-\alpha_k\right)\mathrm{d}\gamma_k=\iint_{\mathbb{R}^d\times\mathbb{R}^d}\hat{\mathbf{c}}\,\mathrm{d}\gamma_k-\alpha_k \overset{k\to\infty}{\longrightarrow}\iint_{\mathbb{R}^d\times\mathbb{R}^d}\hat{\mathbf{c}}\,\mathrm{d}\gamma_*\\ &=\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{c}\,\mathrm{d}\gamma_*.\end{aligned}$$

So, by (21),

$$\liminf_{k\to\infty}\mathcal{E}_k^\tau(\gamma_k)\ge\tau\iint_{\mathbb{R}^d\times\mathbb{R}^d}\mathbf{c}\,\mathrm{d}\gamma_*+\liminf_{k\to\infty}\mathcal{E}(Y\#\gamma_k). \tag{24}$$

Finally, since the projection Y is a continuous map, the pushed-forward measure $Y\#\gamma_k$ converges narrowly to $Y\#\gamma_*$, and since $r\mapsto r^m$ is a convex function, it follows that

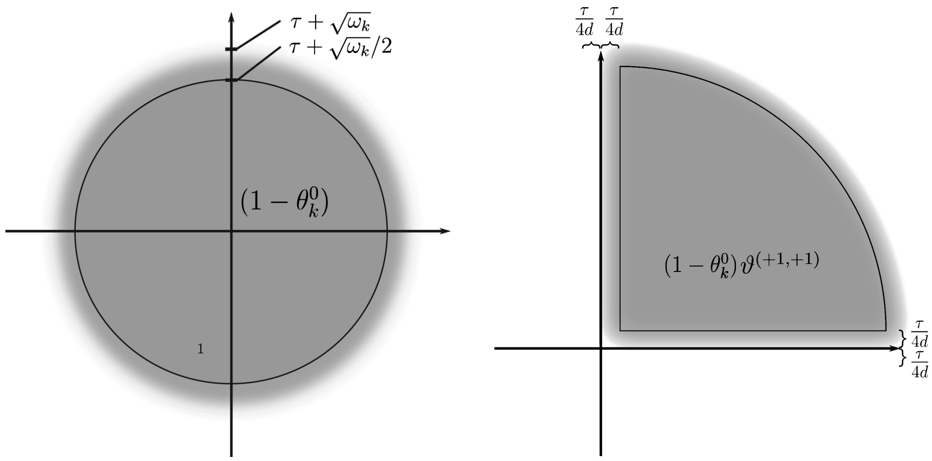
$$\liminf_{k\to\infty}\mathcal{E}(Y\#\gamma_k)\ge\mathcal{E}(Y\#\gamma_*),$$

so the sum on the right-hand side of (24) is greater than or equal to $\mathcal{E}^\tau(\gamma_*|\rho_*)$. □

4.3. **Limsup condition.**

Proposition 3. *For every $\gamma_*\in\Gamma(\rho_*)$ with $\mathcal{E}^\tau(\gamma_*|\rho_*)<\infty$, there exists a sequence of $\gamma_k\in\Gamma_{\delta_k}(\rho_k)$ such that $\gamma_k\to\gamma_*$ narrowly, and* (19) *holds.*

For future reference, define $\eta_*:=Y\#\gamma_*$. From our hypothesis $\mathcal{E}(Y\#\gamma_*)<\infty$, it follows that $\eta_*\in L^m(\mathbb{R}^d)$.



(a) The density of $1-\theta_k^0$ in $\mathbb{R}^2$.

(b) The density of $\vartheta^\beta(1-\theta_k^0)$ in $\mathbb{R}^2$ for $\beta = (+1,+1)$.

FIGURE 4.3.1. The two smoothed indicator functions used in Step 2 to cup up the transport map $\gamma_k^{(1)}$ displayed for $d=2$. Note that the set, on which both have density 1, has been emphasized by an additional black border and a small step in the grayscale.

4.3.1. *Construction of the recovery sequence.* In the following, let $k = 1, 2, \ldots$ be fixed. We are going to construct $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$ in several steps.

Step 1. *Modify γ_* into $\gamma^{(1)}$ such that $X\#\gamma_k^{(1)} = \rho_k\mathcal{L}^d$ and $Y\#\gamma_k^{(1)} = \eta_*\mathcal{L}^d$.* To that end, let $T_k : \mathbb{R}^d \to \mathbb{R}^d$ be an optimal map for the transport from ρ_* to ρ_k in $\mathbf{W}_2$; such a map exists since ρ_* is a probability density, and both ρ_k and ρ_* have a finite second moment. Then $\gamma_k^{(1)} := (T_k \circ X, Y)\#\gamma_*$ has the desired marginals. For later use, define

$$\omega_k := \left(\int_{\mathbb{R}^d} |T_k(x) - x|^2 \rho_*(x)\,\mathrm{d}x\right)^{\frac{1}{2}} = \mathbf{W}_2(\rho_*, \rho_k), \tag{25}$$

which goes to zero by our hypothesis that ρ_k converges to ρ_* in $\mathbf{W}_2$.

Step 2. *Decompose $\gamma_k^{(1)}$ into the sum of 2^d nonnegative measures $\gamma_k^{(2,\beta)}$—each of which fits into the cylinder $|x-y| \le \tau$ after proper translation—and a remainder $\gamma_k^{(2,0)}$ of small mass.*

This is done with the help of several cut-off functions that we define now: for each $\beta \in \{+1,-1\}^d$, choose $\vartheta^\beta \in C^\infty(\mathbb{R}^d)$ such that

- $0 \le \vartheta^\beta \le 1$ and
$$\sum_\beta \vartheta^\beta = 1 \quad \text{on } \mathbb{R}^d,$$
- ϑ^β is supported on the set where $\beta_j x_j \ge -\frac{\tau}{4d}$ for all $j = 1, \ldots, d$.

Thus, each ϑ^β is essentially a smoothed indicator function for one of the 2^d orthants in $\mathbb{R}^d$. The label β corresponds the signs of the d coordinates in the respective orthant. Next, let $\theta_k^0 \in C^\infty(\mathbb{R}^d)$ be a smoothed indicator function of the complement of the closed ball $\overline{\mathbb{B}}_\tau$ of radius τ with the following properties:

- $0 \leq \theta_k^0 \leq 1$ and $|\nabla \theta_k^0| \leq 3\omega_k^{-1/2}$,
- θ_k^0 vanishes on $\overline{\mathbb{B}}_{\tau+\sqrt{\omega_k}/2}$ and is identical to one on the complement of $\overline{\mathbb{B}}_{\tau+\sqrt{\omega_k}}$.

Now define $\theta_k^\beta := \vartheta^\beta(1-\theta_k^0)$ for all $\beta \in \{-1,+1\}^d$, which are smoothed indicator functions of the sectors of the ball $\overline{\mathbb{B}}_\tau$ corresponding to the respective β-orthant (cf. *Figure* 4.3.1 *(b)*). Note that

$$\theta_k^0 + \sum_\beta \theta_k^\beta = 1 \quad \text{on } \mathbb{R}^d. \tag{26}$$

For brevity, further introduce $\Theta_k^\beta(x,y) = \theta_k^\beta(x-y)$ as well as $\Theta_k^0(x,y) = \theta_k^0(x-y)$, and define

$$\gamma_k^{(2,\beta)} := \Theta_k^\beta \gamma_k^{(1)}, \quad \gamma^{(2,0)} := \Theta_k^0 \gamma_k^{(1)}.$$

From (26), it follows that

$$\gamma_k^{(2,0)} + \sum_\beta \gamma_k^{(2,\beta)} = \gamma_k^{(1)}. \tag{27}$$

Roughly speaking, $\gamma^{(2,0)}$ contains the part of $\gamma^{(1)}$ corresponding to transport with speed that exceeds—by $\sqrt{\omega_k}/\tau$ or more—the limit set by the flux limitation. The part $\gamma^{(2,\beta)}$ corresponds to transport that either respects the flux limitation or violates it—by no more than $\sqrt{\omega_k}/\tau$—in the β-directions.

Step 3a. *Translate each of the $\gamma_k^{(2,\beta)}$ in the y-direction to obtain a $\gamma^{(3,\beta)}$ that fits in the cylinder $|x-y| \leq \tau - \delta_k$.*

With

$$\sigma_k := 12\big(\delta_k + \sqrt{\omega_k}\big),$$

we define $\gamma_k^{(3,\beta)} := (X, Y - \sigma_k\beta)\#\gamma_k^{(2,\beta)}$. The fact that $\gamma_k^{(3,\beta)}$ is supported in the aforementioned cylinder is not completely obvious and is verified in Lemma 5 below.

Step 3b. *From the remainder $\gamma_k^{(2,0)}$, define a measure $\gamma_k^{(3,0)}$, which has the same first marginal as $\gamma_k^{(2,0)}$ and a smooth second marginal, and is supported in the cylinder $|x-y| \leq \tau/2$.*

Let λ be a some smooth probability density on $\mathbb{R}^d$ with support in $\overline{\mathbb{B}}_{\tau/2}$. Consider the product measure $\gamma_k^{(2,0)} \otimes \lambda$ on $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$. With $(X, X+Z)$ being the map $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \ni (x,y,z) \mapsto (x, x+z) \in \mathbb{R}^d \times \mathbb{R}^d$, one easily sees that $\gamma_k^{(3,0)} := (X, X+Z)\#(\gamma_k^{(2,0)} \otimes \lambda)$ has the desired properties. Intuitively, on each vertical fiber $\{x\} \times \mathbb{R}^d$, one redistributes the disintegrated mass of $\gamma^{(2,0)}$ in a smooth way around the point $y = x$.

In summary of Steps 1–3, define

$$\gamma_k^{(3)} := \gamma_k^{(3,0)} + \sum_\beta \gamma_k^{(3,\beta)}.$$

Step 4. *Project* $\gamma_k^{(3)} \in \Gamma(\rho_k)$ *onto a* $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$.

For each $Q \in \mathcal{Q}_{\delta_k}$, consider the Borel measure γ_k^Q on $\mathbb{R}^d$ defined by $\gamma_k^Q(A) := \gamma_k^{(3)}(A \times Q)$ for each measurable $A \subset \mathbb{R}^d$. Since

$$\sum_{Q \in \mathcal{Q}_{\delta_k}} \gamma_k^Q = X\#\gamma_k^{(3)} = \rho_k \mathcal{L}^d, \tag{28}$$

it follows that γ_k^Q possesses a nonnegative Lebesgue density $g_k^Q \in L^1(\mathbb{R}^d)$. From the g_k^Q, we define a probability density function $G_k \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ via

$$G_k(x,y) := \frac{g_k^Q(x,y)}{|Q|}, \quad \text{where } Q \in \mathcal{Q}_{\delta_k} \text{ is chosen such that } y \in Q. \tag{29}$$

Our final definition of the recovery sequence is $\gamma_k := G_k \mathcal{L}^d \otimes \mathcal{L}^d$.

4.3.2. *Properties of the recovery sequence.* We prove various properties of the sequence (γ_k) that eventually allow us to conclude (19).

Lemma 3. $\gamma_k \in \Gamma_{\delta_k}(\rho_k)$. *Moreover, its second moment is k-uniformly bounded.*

Proof. This is essentially clear from the construction.

First, γ_k is a probability measure since the construction is a combination of push-forwards (Steps 1 and 3), decomposition into a finite sum of nonnegative measures (Step 2), rearrangement of these components (Step 3), and finally a projection (Step 4), each of which is easily checked to preserve nonnegativity and total mass of the measure.

Second, the X-marginal of γ_k is $\rho_k \mathcal{L}^d$, since Step 1 is made such that $X\#\gamma^{(1)} = T_k\#(X\#\gamma_*) = T_k\#(\rho_*\mathcal{L}^d) = \rho_k\mathcal{L}^d$, and all further steps keep the X-marginal fixed.

Third, γ_k has a finite and, in fact, even k-uniformly bounded second moment. Indeed, since γ_k is supported in $|x-y| \le \tau$ (which follows from the purely geometric considerations in Lemma 5 below), one has γ_k-a.e. that

$$|y|^2 = |x + (y-x)|^2 \le 2|x|^2 + 2\tau^2,$$

and therefore, recalling that γ_k has X-marginal $\rho_k\mathcal{L}^d$,

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} \big(|x|^2 + |y|^2\big)\,\mathrm{d}\gamma_k \le \iint_{\mathbb{R}^d \times \mathbb{R}^d} \big(3|x|^2 + 2\tau^2\big)\,\mathrm{d}\gamma_k = 2\tau^2 + 3\int_{\mathbb{R}^d} |x|^2 \rho_k(x)\,\mathrm{d}x.$$

The last expression is finite and is even k-uniformly bounded since the same is true for ρ_k's second moment; see Lemma 1. □

Lemma 4. *There is a constant C such that*

$$\iint_{\mathbb{R}^d \times \mathbb{R}^d} G_k(x,y) \log G_k(x,y)\,\mathrm{d}(x,y) \le C + d\log(\delta_k^{-1}).$$

Consequently, $\varepsilon_k \mathcal{H}(\gamma_k) \to 0$.

Proof. By definition of G_k,

$$\begin{aligned}
&\iint_{\mathbb{R}^d\times\mathbb{R}^d} G_k(x,y)\log G_k(x,y)\,\mathrm{d}(x,y)\\
&\quad= \sum_{Q\in\mathcal{Q}_{\delta_k}} \iint_{\mathbb{R}^d\times Q} \left(\frac{g_k^Q(x)}{|Q|}\right)\log\left(\frac{g_k^Q(x)}{|Q|}\right)\mathrm{d}(x,y)\\
&\quad= \sum_{Q\in\mathcal{Q}_{\delta_k}} \left[\int_{\mathbb{R}^d} g_k^Q(x)\log g_k^Q(x)\,\mathrm{d}x - \log(|Q|)\int_{\mathbb{R}^d} g_k^Q(x)\,\mathrm{d}x\right]\\
&\quad\le \int_{\mathbb{R}^d}\rho_k(x)\log\rho_k(x)\,\mathrm{d}x - d\log(\delta_k)\int_{\mathbb{R}^d}\rho_k(x)\,\mathrm{d}x,
\end{aligned}$$

where we have estimated $|Q|\ge\delta_k^d$ on grounds of (14), and have used (28) in combination with the superadditivity of the function $s\mapsto s\log s$, that is,

$$a\log a + b\log b \le (a+b)\log(a+b)\quad\text{for arbitrary } a,b\ge 0.$$

The latter is an immediate consequence of the monotonicity of the logarithm. Recalling Lemma 1 and our assumption (17), the convergence follows. □

Lemma 5. *For all k large enough, the γ_k are supported in $|x-y|\le\tau$.*

Proof. The main step is to show that the measures $\gamma_k^{(3)}$ are supported in $|x-y|\le\tau-\delta_k$. The function θ_k^β is supported on the set

$$S^\beta := \left\{y\in\mathbb{R}^d;\ \beta_j y_j\ge -\frac{\tau}{4d}\text{ for all } j,\ |y|\le\tau+\sqrt{\omega_k}\right\}.$$

We show that the translate $S^\beta-\sigma_k\beta$ is a subset of $\overline{\mathbb{B}}_{\tau-\delta_k}$. Observe that S^β is the convex hull of the point $o^\beta := -\frac{\tau}{4d}\beta$ and the spherical cap

$$\mathfrak{S}^\beta = \left\{y\in\mathbb{R}^d;\ \beta_j y_j\ge -\frac{\tau}{4d}\text{ for all } j,\ |y|=\tau+\sqrt{\omega_k}\right\}.$$

Since $\overline{\mathbb{B}}_{\tau-\sqrt{\omega_k}}$ is convex, it thus suffices to verify that the translate of o^β, i.e., the point $-\left(\frac{\tau}{4d}+\sigma_k\right)\beta$, and the translate of the cap, i.e., $\mathfrak{S}^\beta-\sigma_k\beta$, belong to $\overline{\mathbb{B}}_{\tau-\delta_k}$. For all k large enough so that $\sigma_k\le\frac{\tau}{4d}$, the claim $-\left(\frac{\tau}{4d}+\sigma_k\right)\beta\in\overline{\mathbb{B}}_{\tau-\delta_k}$ is obvious. To prove that also $\mathfrak{S}^\beta\subset\overline{\mathbb{B}}_{\tau-\delta_k}$, consider an arbitrary point $x\in\mathfrak{S}^\beta$. Observing that

$$\beta\cdot x = \sum_j\beta_j x_j \ge \sum_j\left(|x_j|-\frac{\tau}{2d}\right) = \sum_j|x_j|-\frac{\tau}{2}\ge\tau+\sqrt{\omega_k}-\frac{\tau}{2}\ge\frac{\tau}{2},$$

it follows that

$$|x-\sigma_k\beta|^2 = |x|^2+\sigma_k^2|\beta|^2-2\sigma_k\beta\cdot x \le \left(\tau+\sqrt{\omega_k}\right)^2+d\sigma_k^2-\tau\sigma_k.$$

Recall that k is large enough such that $\sigma_k\le\frac{\tau}{4d}$; on the one hand, this yields that

$$d\sigma_k^2-\tau\sigma_k\le-\frac{3}{4}\tau\sigma_k,$$

and on the other hand, we obtain

$$\left(\tau+\sqrt{\omega_k}\right)^2-\left(\tau-\delta_k\right)^2 = \left(2\tau+\sqrt{\omega_k}-\delta_k\right)\left(\sqrt{\omega_k}+\delta_k\right)\le 3\tau\left(\sqrt{\omega_k}+\delta_k\right)\le\frac{\tau}{4}\sigma_k.$$

In summary, we conclude that

$$|x - \sigma_k\beta|^2 \le \big(\tau - \delta_k\big)^2,$$

which verifies that $\mathfrak{S}^\beta - \sigma_k\beta \subset \overline{\mathbb{B}}_{\tau-\delta_k}$.

By definition, $\gamma_k^{(2,\beta)}$ is supported in the region where $y - x \in S^\beta$. Its translate $\gamma_k^{(3,\beta)} = (X, Y - \sigma_k\beta)\#\gamma_k^{(2,\beta)}$ is therefore supported where $y - x \in S^\beta - \sigma_k\beta \subset \overline{\mathbb{B}}_{\tau-\delta_k}$, where the inclusion is a consequence of the considerations above.

This proves that each $\gamma_k^{(3)}$ is supported in $|x-y| \le \tau - \delta_k$. From the construction of γ_k it is clear that $\operatorname{supp}\gamma_k$ intersects $\{x\}\times Q$ for some $x \in \mathbb{R}^d$ and $Q \in \mathcal{Q}_{\delta_k}$ only if $\operatorname{supp}\gamma_k^{(3)}$ intersects $\{x\}\times Q$. Since the distance of two points in Q is less than δ_k by (14), it follows that γ_k is supported in $|x-y| \le \tau$. □

Lemma 6. $\gamma_k^{(2,0)}$*'s total mass does not exceed* $4\omega_k$.

Proof. Recall that $|x - y| \le \tau$ for γ_*-a.e. (x, y). Hence $|T_k(x) - y| \ge \tau + \sqrt{\omega_k}/2$ implies that $|T_k(x) - x| \ge \sqrt{\omega_k}/2$ for γ_*-a.e. (x, y). Consequently, recalling that $\gamma_k^{(2,0)} = \Theta_k^0\,(T_k \circ X, Y)\#\gamma_*$,

$$\begin{aligned}
\gamma_k^{(2,0)}[\mathbb{R}^d\times\mathbb{R}^d] &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \theta_k^0(T_k(x) - y)\,\mathrm{d}\gamma_*(x,y)\\
&\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_{|T_k(x)-y|\ge\tau+\sqrt{\omega_k}/2}\,\mathrm{d}\gamma_*(x,y)\\
&\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{1}_{|T_k(x)-x|\ge\sqrt{\omega_k}/2}\,\mathrm{d}\gamma_*(x,y)\\
&= \int_{\mathbb{R}^d} \mathbf{1}_{|T_k(x)-x|^2\ge\omega_k/4}\,\rho_k(x)\,\mathrm{d}x\\
&\le \frac{4}{\omega_k}\int_{\mathbb{R}^d} |T_k(x) - x|^2\rho_k(x)\,\mathrm{d}x = 4\omega_k,
\end{aligned}$$

where we have used the definition (25) of ω_k in the last step. □

Lemma 7. γ_k *converges narrowly to* γ_**, and moreover,*

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\gamma_k \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}\,\mathrm{d}\gamma_*. \tag{30}$$

Proof. To start with, we show that $\gamma_k^{(1)}$ converges narrowly to γ_*. Since both each $\gamma_k^{(1)}$ and the proposed limit γ_* are probability measures, it suffices to show convergence in distribution, i.e., for all test functions $\psi \in C_c^\infty(\mathbb{R}^d\times\mathbb{R}^d)$. Since $\omega_k \to 0$ in (25), it follows that T_k converges to the identity map in measure with respect to ρ_*, and hence also $(T_k\circ X, Y)$ converges to (X, Y) in measure with respect to γ_*. As well—ψ being smooth and compactly supported—$\psi(T_k \circ X, Y)$ converges to ψ in measure with respect to γ_*. By the dominated convergence theorem,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi\,\mathrm{d}\gamma_k^{(1)} = \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(T_k\circ X, Y)\,\mathrm{d}\gamma_* \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi\,\mathrm{d}\gamma_*.$$

Next, we show that also $\gamma_k^{(3)}$ converges to γ_*:

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_k^{(3,0)}(x,y) \\
&\quad + \sum_\beta \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_k^{(3,\beta)}(x,y) \\
= \iint_{\mathbb{R}^d\times\mathbb{R}^d} &\left[\int_{\mathbb{R}^d} \psi(x,x+z)\lambda(z)\,\mathrm{d}z\right] \mathrm{d}\gamma_k^{(2,0)}(x,y) \\
&\quad + \sum_\beta \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y-\sigma_k\beta)\,\mathrm{d}\gamma_k^{(2,\beta)}(x,y)
\end{aligned}$$

$$= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \psi(x,x+z)\lambda(z)\,\mathrm{d}z - \sum_\beta \psi(x,y-\sigma_k\beta)\vartheta^\beta(x-y)\right] \mathrm{d}\gamma_k^{(2,0)}(x,y) \tag{31}$$

$$+ \sum_\beta \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y-\sigma_k\beta)\vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(1)}(x,y). \tag{32}$$

Here we have used that, by definition of $\gamma^{(2,\beta)}$ from $\gamma^{(1)}$ in Step 2,

$$\begin{aligned}
\mathrm{d}\gamma_k^{(2,\beta)} &= \vartheta^\beta(x-y)\big(1-\theta_k^0(x-y)\big)\,\mathrm{d}\gamma_k^{(1)}(x,y) \\
&= \vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(1)}(x,y) - \vartheta^\beta(x-y)\,\mathrm{d}\gamma_k^{(2,0)}(x,y).
\end{aligned}$$

The integral in (31) converges to zero thanks to Lemma 6; observe that the expression inside the square brackets is a continuous function that is bounded independently of k. Concerning the sum in (32), observe that $\psi(x,y-\sigma_k\beta)\to\psi(x,y)$ uniformly in (x,y) since ψ is compactly supported, and recall from above that $\gamma_k^{(1)}$ converges narrowly to γ_*. This suffices to conclude that

$$\begin{aligned}
\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y) &\to \sum_\beta \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\vartheta^\beta(x-y)\,\mathrm{d}\gamma_*(x,y) \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_*(x,y),
\end{aligned}$$

where we have used that the smooth expressions $\vartheta^\beta(x-y)$ sum up to unity on the support of γ_*.

As the last step, we show that γ_k converges to γ_* as well. For each $Q\in\mathcal{Q}_{\delta_k}$, define $\Psi_k^Q\in C_c(\mathbb{R}^d)$ by

$$\Psi_k^Q(x) = \frac{1}{|Q|}\int_Q \psi(x,y)\,\mathrm{d}y.$$

Note that there is one common compact set on which all the Ψ_k^Q are supported. From the definition of γ_k, it follows that

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} &\psi(x,y)\,\mathrm{d}\gamma_k(x,y)\\ &= \sum_{Q\in\mathcal{Q}_{\delta_k}} \int \Psi^Q(x) g_k^Q(x)\,\mathrm{d}x = \sum_{Q\in\mathcal{Q}_{\delta_k}} \iint_{\mathbb{R}^d\times Q} \Psi^Q(x)\,\mathrm{d}\gamma_k^{(3)}(x,y)\\ &= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(x,y)\,\mathrm{d}\gamma_k^{(3)}(x,y) + \sum_{Q\in\mathcal{Q}_{\delta_k}} \iint_{\mathbb{R}^d\times Q} \big[\Psi^Q(x)-\psi(x,y)\big]\,\mathrm{d}\gamma_k^{(3)}(x,y).\end{aligned}$$

Now since the term in square brackets converges uniformly to zero as the mesh is refined, and since $\gamma_k^{(3)}$ converges to γ_* narrowly, distributional—and subsequently narrow—convergence of γ_k to γ_* follows.

Finally, in combination with the fact that—thanks to Lemma 5—all the γ_k are supported inside $|x-y|\le\tau$, where $\mathbf{c}_k$ converges to $\mathbf{c}$ uniformly by hypothesis (15), the claimed convergence (30) is proven. ☐

Lemma 8. *$Y\#\gamma_k^{(3)}$ has a Lebesgue density $\eta_k^{(3)} \in L^m(\mathbb{R}^d)$, and $\eta_k^{(3)} \to \eta_*$ in $L^m(\mathbb{R}^d)$.*

Proof. By Step 3b, $Y\#\gamma_k^{(3,0)} = \eta_k^{(3,0)}\mathcal{L}^d$ for a smooth density $\eta_k^{(3,0)} \in L^1\cap L^\infty(\mathbb{R}^d)$. Moreover, for each $\beta\in\{-1,+1\}^d$, the marginal $Y\#\gamma_k^{(3,\beta)}$ is a translate of $Y\#\gamma^{(2,\beta)}$, and from (27) it follows that

$$Y\#\gamma_k^{(2,0)} + \sum_\beta Y\#\gamma_k^{(2,\beta)} = Y\#\gamma_k^{(1)} = \eta_*\mathcal{L}^d,$$

hence $Y\#\gamma^{(3,\beta)}$ has a Lebesgue density

$$\eta_k^{(3,\beta)} \le \eta_*(\cdot+\sigma_k\beta) \in L^1\cap L^m(\mathbb{R}^d). \tag{33}$$

Further define η_*^β as the density of $Y\#(\Theta_k^\beta\gamma_*)$; this definition is independent of the index k, since γ_* is supported in the region $|x-y|\le\tau$ where $\theta_k^0(x-y)$ vanishes. Obviously

$$\eta_* = \sum_\beta \eta_*^\beta, \quad \eta_k^{(3)} = \eta_k^{(3,0)} + \sum_\beta \eta_k^{(3,\beta)}. \tag{34}$$

In the convergence proof that follows, we use the dual representation of the norm on $L^q(\mathbb{R}^d)$:

$$\|f\|_{L^q} = \sup\left\{\int \psi(x)f(x)\,\mathrm{d}x\,;\ \psi\in C_c(\mathbb{R}^d),\ \|\psi\|_{L^{q'}}\le 1\right\},$$

where $q'=\frac{q}{q-1}$ is the Hölder conjugate exponent of $q>1$.

To begin with, observe that $\eta_k^{(3,0)}$ converges to zero in $L^m(\mathbb{R}^d)$. For that, let $\psi\in C(\mathbb{R}^d)$ with $\|\psi\|_{L^{m'}}\le 1$. Then, with the help of Hölder's inequality and Lemma 6 above,

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)\,\mathrm{d}\gamma_k^{(3,0)}(x,y) &= \iint_{\mathbb{R}^d\times\mathbb{R}^d}\left[\int_{\mathbb{R}^d}\psi(x+z)\lambda(z)\,\mathrm{d}z\right]\mathrm{d}\gamma_k^{(2,0)}(x,y)\\ &\le \iint_{\mathbb{R}^d\times\mathbb{R}^d} \|\psi\|_{L^{m'}}\|\lambda\|_{L^m}\,\mathrm{d}\gamma_k^{(2.0)} \le 4\|\lambda\|_{L^m}\omega_k.\end{aligned}$$

Next, we show that $\eta_k^{(3,\beta)} \to \eta_*^\beta$ in $L^q(\mathbb{R}^d)$, for each β, where $q := \frac{2m}{m+1} < m$; note that $q' = 2m'$. For $\psi \in C(\mathbb{R}^d)$ with $\|\psi\|_{L^{q'}} \leq 1$, we have

$$
\begin{aligned}
&\int_{\mathbb{R}^d} \psi(y)\big[\eta_k^{(3,\beta)}(y) - \eta_*^{(3,\beta)}(y)\big]\,\mathrm{d}y \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} \Big[\psi(y-\sigma_k\beta)\Theta_k^\beta(T_k(x),y) - \psi(y)\Theta_k^\beta(x,y)\Big]\,\mathrm{d}\gamma_*(x,y) \\
&= \iint_{\mathbb{R}^d\times\mathbb{R}^d} [\psi(y-\sigma_k\beta) - \psi(y)]\Theta_k^\beta(x,y)\,\mathrm{d}\gamma_*(x,y) \\
&\qquad + \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y-\sigma_k\beta)\big[\theta_k^\beta(T_k(x)-y) - \theta_k^\beta(x-y)\big]\,\mathrm{d}\gamma_*(x,y) \\
&\leq \int_{\mathbb{R}^d} [\psi(y-\sigma_k\beta) - \psi(y)]\eta_*^\beta(y)\,\mathrm{d}y \\
&\qquad + \left(\int_{\mathbb{R}^d} |\psi(y-\sigma_k\beta)|^2\eta_*^\beta(y)\,\mathrm{d}y\right)^{\frac12} \left(\|\nabla\theta_k^\beta\|_{L^\infty}^2 \int_{\mathbb{R}^d} |T_k(x)-x|^2\rho_*(x)\,\mathrm{d}x\right)^{\frac12} \\
&\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} \psi(y)\big[\eta_*^\beta(y+\sigma_k\beta) - \eta_*^\beta(y)\big]\,\mathrm{d}y + \|\psi\|_{L^{2m'}}^{\frac12}\|\eta_*\|_{L^m}^{\frac12}\|\nabla\theta_k^\beta\|_{L^\infty}\omega_k \\
&\leq \big\|(\mathrm{id}-\sigma_k\beta)\#\eta_*^\beta - \eta_*^\beta\big\|_{L^q} + 3\|\eta_*\|_{L^m}^{\frac12}\omega_k^{\frac12}.
\end{aligned}
$$

In the last step, we have used that $\nabla\theta_k^\beta = (1-\theta_k^0)\nabla\vartheta^\beta - \vartheta^\beta\nabla\theta_k^0$, and hence $\|\nabla\theta_k^\beta\|_{L^\infty} \leq 4\omega_k^{-1/2}$ by our hypotheses on θ_k^0 and ϑ^β, at least for all sufficiently large k. The first term of the final sum above goes to zero, since $\sigma_k \to 0$, and the translation semigroup is continuous in $L^q(\mathbb{R}^d)$; the second term goes to zero since $\omega_k \to 0$.

From this, we conclude convergence of $\eta_k^{(3,\beta)}$ to η_*^β in $L^q(\mathbb{R}^d)$, and in particular also in measure. Further, from the bound (33), it follows that $\eta_k^{(3,\beta)}$ is equi-integrable in $L^m(\mathbb{R}^d)$. Hence $\eta_k^{(3,\beta)} \to \eta_*^\beta$ also in $L^m(\mathbb{R}^d)$. In view of (34), this verifies the claim. □

Lemma 9. *Define η_k by $\eta_k(y) = \int_{\mathbb{R}^d} G_k(x,y)\,\mathrm{d}x$, with G_k from (29). Then $\eta_k \in L^m(\mathbb{R}^d)$, and $\eta_k \to \eta$ in $L^m(\mathbb{R}^d)$. Consequently, $\mathcal{E}(Y\#\gamma_k) \to \mathcal{E}(Y\#\gamma_*)$.*

Proof. First, we recall two properties of the linear projection operator $\Pi_\delta : L^m(\mathbb{R}^d) \to L^m(\mathbb{R}^d)$ given by

$$\Pi_\delta[f](y) = \fint_Q f(y')\,\mathrm{d}y', \qquad \text{where } Q \in \mathcal{Q}_\delta \text{ is such that } y \in Q.$$

Namely,

(a) $\|\Pi_\delta[f] - \Pi_\delta[g]\|_{L^m(\mathbb{R}^d)} \leq \|f-g\|_{L^m(\mathbb{R}^d)}$ for all $f,g \in L^m(\mathbb{R}^d)$;
(b) $\Pi_\delta[f] \to f$ in $L^m(\mathbb{R}^d)$ for each $f \in L^m(\mathbb{R}^d)$ as $\delta \searrow 0$.

Indeed, claim (a) is an easy consequence of Jensen's inequality:

$$\begin{aligned}
&\big\|\Pi_\delta[f]-\Pi_\delta[g]\big\|_{L^m(\mathbb{R}^d)}^m \\
&\quad= \sum_{Q\in\mathcal{Q}_{\delta_k}} \big\|\Pi_\delta[f]-\Pi_\delta[g]\big\|_{L^m(Q)}^m = \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left| \fint_Q \big[f(y')-g(y')\big]\mathrm{d}y'\right|^m \mathrm{d}y \\
&\quad\le \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left[\fint_Q \big|f(y')-g(y')\big|^m \mathrm{d}y'\right]\mathrm{d}y = \sum_{Q\in\mathcal{Q}_{\delta_k}} \|f-g\|_{L^m(Q)}^m = \|f-g\|_{L^m(\mathbb{R}^d)}^m.
\end{aligned}$$

Concerning claim (b), we use that thanks to hypothesis (14), arbitrary $y'\in Q$ lie in a ball of radius δ_k around any given $y\in Q$:

$$\begin{aligned}
&\big\|\Pi_\delta[f]-f\big\|_{L^m(\mathbb{R}^d)}^m \\
&\quad= \sum_{Q\in\mathcal{Q}_{\delta_k}} \big\|\Pi_\delta[f]-f\big\|_{L^m(Q)}^m = \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left| \fint_Q \big[f(y')-f(y)\big]\,\mathrm{d}y'\right|^m \mathrm{d}y \\
&\quad\le \sum_{Q\in\mathcal{Q}_{\delta_k}} \int_Q \left[\fint_Q \big|f(y')-f(y)\big|^m \mathrm{d}y'\right]\mathrm{d}y \le \int_{\overline{\mathbb{B}}} \big\|f-f(\cdot+\delta_k z)\big\|_{L^m(\mathbb{R}^d)}^m\,\mathrm{d}z.
\end{aligned}$$

The norm inside the final integral goes to zero as $\delta_k\to 0$, since $f(\cdot+\delta_k z)\to f$ in $L^m(\mathbb{R}^d)$, uniformly with respect to $z\in\overline{\mathbb{B}}$.

To connect this auxiliary result to the claim of the lemma, recall that $\eta_k^{(3)}\to\eta_*$ in $L^m(\mathbb{R}^d)$ by Lemma 8 above, and observe that $\eta_k=\Pi_{\delta_k}[\eta_k^{(3)}]$. Therefore,

$$\begin{aligned}
\|\eta_k-\eta_*\|_{L^m} &\le \|\Pi_{\delta_k}[\eta_k^{(3)}]-\Pi_{\delta_k}[\eta_*]\|_{L^m} + \|\Pi_{\delta_k}[\eta_*]-\eta_*\|_{L^m} \\
&\le \|\eta_k^{(3)}-\eta_*\|_{L^m} + \|\Pi_{\delta_k}[\eta_*]-\eta_*\|_{L^m}
\end{aligned}$$

tends to zero. □

4.4. **Existence and convergence of minimizers.**

Lemma 10. *For each k large enough, $\mathcal{E}_k^\tau$ has a (unique if $\varepsilon_k>0$) minimizer $\hat\gamma_k\in\Gamma_{\delta_k}(\rho_k)$.*

Proof. We use the estimates from Lemma 2: thanks to (21), the $\mathcal{E}_k^\tau$ are bounded below for all sufficiently small k. And thanks to (20), the γ's in the sublevels of $\mathcal{E}_k^\tau$ have uniformly bounded second moment, hence are relatively compact with respect to narrow convergence. Moreover, it is easily seen that $\mathcal{E}_k^\tau$ is the sum of three convex (in the sense of convex combinations of measures) functionals, and thus is lower semicontinuous with respect to narrow convergence. Moreover, $\mathcal{H}$ is a strictly convex functional on $\Gamma(\rho_k)$, so $\mathcal{E}_k^\tau$ is strictly convex if $\varepsilon_k>0$. This together allows us to invoke the direct methods from the calculus of variations and conclude the existence of a minimizer, which is unique if $\varepsilon_k>0$. □

Lemma 11. *Let $\hat\gamma_k\in\Gamma_{\delta_k}(\rho_k)$ be minimizers of the respective $\mathcal{E}_k^\tau$. Then a subsequence of $(\hat\gamma_k)$ converges in $\mathbf{W}_2$ to a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.*

Proof. We begin by showing that the second momenta of the $\hat\gamma_k$ are k-uniformly bounded. In view of estimate (20), it suffices to show that $\mathcal{E}_k^\tau(\hat\gamma_k)$ is k-uniformly bounded. But this is a consequence of Γ-convergence: since $\mathcal{E}^\tau(\cdot|\rho_*)$ is not identically $+\infty$—for instance, $\mathcal{E}^\tau((X,X)\#\rho_*\mathcal{L}^d|\rho_*)=\mathcal{E}(\rho_*)<\infty$—there is a recovery sequence γ_k such that $\mathcal{E}_k^\tau(\gamma_k)$ is bounded, and hence also $\mathcal{E}_k^\tau(\hat\gamma_k)$ is bounded.

Consequently, there is a subsequence that converges narrowly to a limit $\hat{\gamma}_*$. Since $X\#\gamma_k = \rho_k\mathcal{L}^d \to \rho_*\mathcal{L}^d$ narrowly by hypothesis, and since the projection X is continuous, it follows that $\gamma^* \in \Gamma(\rho_*)$. Thus, by the fundamental properties of Γ-convergence, γ_* is a minimizer of $\mathcal{E}^\tau(\cdot|\rho_*)$.

It remains to be shown that actually $\hat{\gamma}_k \to \hat{\gamma}_*$ in $\mathbf{W}_2$. It suffices to verify that $\hat{\gamma}_k$'s second moment converges to that of $\hat{\gamma}_*$. The second moment of $\hat{\gamma}_k$ amounts to

$$\begin{aligned}\iint_{\mathbb{R}^d\times\mathbb{R}^d} \big(|x|^2+|y|^2\big)\,\mathrm{d}\hat{\gamma}_k = 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\hat{\gamma}_k + \iint_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,\mathrm{d}\hat{\gamma}_k \\ + 2\iint_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_k.\end{aligned} \tag{35}$$

Thanks to Lemma 1,

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\gamma_k = \int_{\mathbb{R}^d} |x|^2\rho_k(x)\,\mathrm{d}x \to \int_{\mathbb{R}^d} |x|^2\rho_*(x)\,\mathrm{d}x = \iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\gamma_*.$$

Further, recalling the lower bound (16) on $\mathbf{c}_k$ and estimate (21), we obtain for all sufficiently large k that

$$\begin{aligned}\iint_{|y-x|\ge 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k &\le 4\iint_{|y-x|\ge 2\tau} \big(|y-x|-\tau\big)^2\,\mathrm{d}\hat{\gamma}_k \\ &\le 4\alpha_k\iint_{\mathbb{R}^d\times\mathbb{R}^d} \mathbf{c}_k\,\mathrm{d}\hat{\gamma}_k \le \frac{8\alpha_k}{\tau}\mathcal{E}_k^\tau(\hat{\gamma}_k),\end{aligned}$$

which converges to zero as $k\to\infty$ since $\mathcal{E}_k^\tau(\hat{\gamma}_k)$ is bounded. In the same spirit, also

$$\left|\iint_{|y-x|\ge 2\tau} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_k\right| \le \frac{\sqrt{\alpha_k}}{2}\iint_{\mathbb{R}^d\times\mathbb{R}^d} |x|^2\,\mathrm{d}\hat{\gamma}_k + \frac{1}{2\sqrt{\alpha_k}}\iint_{|y-x|\ge 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k$$

converges to zero. The continuous function $|y-x|^2$ is bounded on the set where $|y-x|\le 2\tau$, so narrow convergence $\hat{\gamma}_k \to \hat{\gamma}_*$ implies

$$\iint_{|y-x|\le 2\tau} |y-x|^2\,\mathrm{d}\hat{\gamma}_k \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2\,\mathrm{d}\hat{\gamma}_*.$$

Finally, for $|y-x|\le 2\tau$, the function $x\cdot(y-x)$ is bounded in modulus by $2\tau|x|$. Since the $\hat{\gamma}_k$ have k-uniformly bounded second momenta, Prokhorov's theorem yields

$$\iint_{|y-x|\le 2\tau} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_k \to \iint_{\mathbb{R}^d\times\mathbb{R}^d} x\cdot(y-x)\,\mathrm{d}\hat{\gamma}_*.$$

In summary, we can pass to the limit $k\to\infty$ in each term on the right-hand side of (35), obtaining the second moment of $\hat{\gamma}_*$. □

5. Numerical scheme

5.1. **Formulation of the minimization problem.** Throughout this section, we assume that the following are fixed: a bounded domain $\Omega\subset\mathbb{R}^d$, a tesselation $\mathcal{Q}_\delta$ of Ω with cells of diameter at most $\delta>0$ (see (14)), an entropic regularization parameter $\varepsilon>0$, a time step $\tau>0$, and an approximation $\tilde{\mathbf{c}} := \mathbf{c}_{\tau,\delta}$ of the distance cost function $\mathbf{c}$, which is such that $\tilde{\mathbf{c}}$ is constant (possibly $+\infty$) on each $Q\times Q'$ where $Q, Q'\in\mathcal{Q}_\delta$, and such that $\tilde{\mathbf{c}}(x,y)<\infty$ at each (x,y) with $|x-y|\le\tau$. We assume that the elements Q_i of $\mathcal{Q}_\delta$ are enumerated with an index $i\in I$, where I is a finite index set, and for each $i\in I$, a point $x_i\in Q_i$ is given.

We need to fix some further notation: indexed quantities $u = (u_i)_{i\in I}$ are considered as (column) vectors, quantities $g = (g_{i,j})_{i,j\in I}$ with double index as matrices. Below, we use $\odot$ to denote the entrywise products of vectors and matrices, $[u\odot v]_j = u_j v_j$ and $[g\odot h]_{i,j} = g_{i,j}h_{i,j}$, respectively. In the same spirit, $\frac{u}{v}$ and $\frac{g}{h}$ denote entrywise division. Further, for a vector u, we denote by $\operatorname{diag}(u)$ the diagonal matrix with the vector u on the diagonal $[\operatorname{diag}(u)]_{i,j} = u_i\delta_{i,j}$ where $\delta_{i,j}$ denotes the Kronecker delta. For the sake of disambiguation, the usual matrix-vector product is written as $g\cdot u$, i.e., $[g\cdot u]_i = \sum_j g_{i,j}u_j$, and $u\otimes v$ denotes outer product of the vectors u and v, that is $[u\otimes v]_{i,j} = u_i v_j$.

Remark 2. With the x_i at hand, a practical choice for $\tilde{\mathbf{c}}$ that conforms with (15) and (16) is the following:

$$\tilde{\mathbf{c}}(x,y) = \tilde{\mathbf{c}}_{i,j} := \mathbf{C}\left(\frac{|x_i - y_j|}{\tau+\delta}\right) \text{ for all } x\in Q_i,\ y\in Q_j, \tag{36}$$

and extend $\tilde{\mathbf{c}}$ by lower semicontinuity to all of $\mathbb{R}^d\times\mathbb{R}^d$. The modified denominator $\tau+\delta$ has been chosen such that $\tilde{\mathbf{c}}$ is finite on each $2d$-cube $Q_i\times Q_j$ that intersects the region $|x-y|\le\tau$.

A density $\rho\in\mathcal{P}_\delta(\Omega)$ is the conveniently identified with the vector $r = (r_i)$, where r_i is the constant density on Q_i. Now, if $\rho\in\mathcal{P}_\delta(\Omega)$, and if $\gamma = G\mathcal{L}^d\otimes\mathcal{L}^d$ is a minimizer of $\mathcal{E}^\tau_{\varepsilon,\delta,\tilde{\mathbf{c}}}(\cdot|\rho)$ on $\Gamma_\delta(\rho)$, then G is constant on each $2d$-cube $Q_i\times Q_j$; this follows by Jensen's inequality and strict convexity of $\mathcal{H}$. Accordingly, the set of all possible minimizers γ can be parametrized by matrices g, where $g_{i,j}$ is the constant value of γ's density on $Q_i\times Q_j$.

For notational simplicity, introduce the vector $\mathbb{I}_\delta$ with $[\mathbb{I}_\delta]_j = |Q_j|$ for all j, so that

$$[\mathbb{I}_\delta^T\cdot g]_j = \sum_i |Q_i|g_{i,j}, \quad [g\cdot\mathbb{I}_\delta]_i = \sum_j |Q_j|g_{i,j}.$$

In this notation, the constraint $X\#\gamma = \rho\mathcal{L}^d$ then becomes $g\cdot\mathbb{I}_\delta = r$, and we have

$$\mathcal{H}(\gamma) = \sum_{i,j}|Q_i||Q_j|\big[g_{i,j}\log g_{i,j} - g_{i,j}\big], \quad \mathcal{E}(Y\#\gamma) = \sum_j\left[|Q_j|h\left(\sum_i|Q_i|g_{i,j}\right)\right].$$

In terms of the notation introduced above, the variational problem (10) turns into

$$\begin{aligned} g^n = \operatorname*{arg\,min}_{g=(g_{i,j})}\Bigg(&\varepsilon\sum_{i,j}\left[|Q_i||Q_j|\left(\frac{\tau}{\varepsilon}\tilde{\mathbf{c}}_{i,j} + \log g_{i,j}\right)g_{i,j} - g_{i,j}\right] \\ &+ \sum_j\left[|Q_j|h\left(\sum_i|Q_i|g_{i,j}\right)\right] + \iota_{(\bar{r}^{(n-1)} - g\mathbb{I}_\delta)}\Bigg), \end{aligned}\tag{37}$$

where $\bar{r}^{n-1} = g^{n-1}\cdot\mathbb{I}_\delta$ encodes the datum from the previous step.

5.2. **Excursion: Dykstra's algorithm.** In this section, we briefly summarize the concept of the generalized Dykstra algorithm that is the basis for the efficient numerical approximation of Wasserstein gradient flows in the spirit of [27].

Let $F : X\to\mathbb{R}$ be a convex differentiable function defined on a Hilbert space X, and let F^* be its Legendre dual. Below, we identify at each $x\in X$ the differentials

$F'(x), (F^*)'(x) \in X'$ by their respective Riesz duals in X. The *Bregman divergence* $D_F(x, y)$ of $x \in X$ relative to $y \in X$ is defined by

$$D_F(x|y) = F(x) - F(y) - \langle F'(y), x - y\rangle. \tag{38}$$

By convexity, $D_F(x|y) \geq 0$. Further, let $\phi_1, \phi_2 : X \to \mathbb{R} \cup \{+\infty\}$ be two proper, convex and lower semicontinuous functionals on X, and consider, for a given $y \in X$, the variational problem

$$D_F(x|y) + \phi_1(x) + \phi_2(x) \longrightarrow \min. \tag{39}$$

In this setting, the generalized Dykstra algorithm for approximation of a minimizer $x^* \in X$ is the following. Let $x^{(0)} := y$, let $q^{(0)} := q^{(-1)} := 0$, and define for $k = 0, 1, 2, \ldots$ inductively:

$$\begin{aligned} x^{(k+1)} &:= \operatorname*{arg\,min}_{x \in X} \big[D_F\big(x\big|(F^*)'(F'(x^{(k)}) + q^{(k-1)})\big) + \phi_{[k]}(x)\big], \\ q^{(k+1)} &:= F'(x^{(k)}) + q^{(k-1)} - F'(x^{(k+1)}), \end{aligned} \tag{40}$$

where $[k] = 1$ if k is even, and $[k] = 2$ if k is odd. In the special case that $F(x) = \frac{1}{2}\langle x, x\rangle$ and ϕ_1, ϕ_2 are the indicator functions of two convex sets with nonempty intersection, then (40) reduces to the original Dykstra projection algorithm.

Under certain hypotheses (for instance, if X is finite-dimensional), it can be proven that $x^{(k)}$ converges to a minimizer x^* of (39) in X. The core idea of the convergence proof is to study the dual problem for (39), for which the iteration (40) attains a considerably easier form. We refer to [9,15,27] for further discussion of the algorithm, including questions of well-posedness and convergence, in the context of fully discrete approximation of gradient flows.

5.3. **From the minimization problem to the iteration.** In this section, we once again follow closely [27] with the goal is to rewrite (37) in the form (39), and then to apply the algorithm (40) to its solution. The Hilbert space is that of matrices $g = (g_{i,j})_{i,j \in I}$ endowed with the scalar product

$$\langle g, g'\rangle = \sum_{i,j} |Q_i||Q_j| g_{i,j} g'_{i,j},$$

and we shall choose F in (38) as

$$F(g) = \sum_{i,j} |Q_i||Q_j| g_{i,j} \log g_{i,j},$$

with the convention that $0 \log 0 = 0$ and $r \log r = +\infty$ for any $r < 0$, which has Legendre dual

$$F^*(\omega) = \sum_{i,j} |Q_i||Q_j| e^{\omega_{i,j}},$$

and respective derivatives (recall that we identify the functional $F'(g)$ with its Riesz dual)

$$\big[F'(g)\big]_{i,j} = \log g_{i,j}, \quad \big[(F^*)'(\omega)\big]_{i,j} = \exp \omega_{i,j}.$$

The corresponding Bregman distance is the Kullback–Leibler divergence,

$$\mathrm{KL}(g|\omega) := D_F(g|\omega) = \sum_{i,j} |Q_i||Q_j| \big[g_{i,j}(\log g_{i,j} - \log \omega_{i,j}) - g_{i,j} + \omega_{i,j}\big],$$

which is defined for matrices g and ω with nonnegative entries. The correct interpretation of the logarithmic terms is the following: if $\omega_{i,j}=0$, then the entire term in square brackets is $+\infty$ unless $g_{i,j}=0$ as well, in which case this term is zero.

Next, we rewrite our minimization problem (37) in the form (39). As the reference density $\xi=(\xi_{i,j})$ for the divergence, we choose

$$\xi_{i,j}=\begin{cases}\exp\left(-\frac{\tau}{\varepsilon}\tilde{\mathbf{c}}_{i,j}\right) & \text{if } \tilde{\mathbf{c}}_{i,j} \text{ is finite,}\\ 0 & \text{if } \tilde{\mathbf{c}}_{i,j}=+\infty.\end{cases}$$

Thus $\tau\tilde{\mathbf{c}}_{i,j}g_{i,j}=-\varepsilon g_{i,j}\log\xi_{i,j}$, with the convention that $0\log 0=0$, but $(-a)\log(-a)=+\infty$ and $-a\log 0=+\infty$ for any $a>0$. The sum of the first two terms in the variational functional (37) takes the convenient form

$$\begin{aligned}\sum_{i,j}\Big[|Q_i||Q_j|\Big(\frac{\tau}{\varepsilon}\tilde{\mathbf{c}}_{i,j}+\log g_{i,j}-1\Big)g_{i,j}\Big] &= \sum_{i,j}|Q_i||Q_j|g_{i,j}\big(\log g_{i,j}-\log\xi_{i,j}-1\big)\\ &= \mathrm{KL}(g|\xi)-\sum_{i,j}|Q_i||Q_j|\xi_{i,j}.\end{aligned}$$

Recall that $\mathrm{KL}(g|\xi)\geq 0$ by construction, and that $\mathrm{KL}(g|\xi)=+\infty$ unless $g_{i,j}=0$ for all (i,j) with $\tilde{\mathbf{c}}_{i,j}=+\infty$. Neglecting irrelevant factors and constants, the minimization problem (37) attains the form

$$g^n=\arg\min_g\big[\varepsilon\mathrm{KL}(g|\xi)+\phi_1(\mathbb{I}_\delta^T\cdot g)+\phi_2^n(g\cdot\mathbb{I}_\delta)\big], \tag{41}$$

where

$$\phi_1(s)=\mathcal{E}_\delta(s)=\sum_j|Q_j|h(s_j),\quad \phi_2^n(r)=\iota_{(\bar{r}^{n-1}-r)}=\begin{cases}0 & \text{if } r=\bar{r}^{(n-1)},\\ +\infty & \text{otherwise.}\end{cases}$$

Using that for our choice of F,

$$\big[(F^*)'\big(F'(x)+q\big)\big]_{i,j}=\exp\big(\log x_{i,j}+q_{i,j}\big)=(x\odot s)_{i,j},\quad \text{with } s_{i,j}:=e^{q_{i,j}},$$

Dykstra's algorithm (40) translates into the following: from $g^{(0)}=\xi$ and $s^{(0)}=s^{(-1)}\equiv 1$, inductively define

$$g^{(k+1)}=\Phi_{[k]}(g^{(k)}\odot s^{(k-1)}),\quad s^{(k+1)}=\frac{g^{(k)}\odot s^{(k-1)}}{g^{(k+1)}}, \tag{42}$$

again with $[k]=1$ for even k, and $[k]=2$ for odd k, where $\Phi_1(\omega)$ and $\Phi_2(\omega)$ are, respectively, the solutions to the minimization problems

$$\varepsilon\mathrm{KL}(g|\omega)+\phi_1(\mathbb{I}_\delta^T\cdot g)\to\min\quad\text{and}\quad\varepsilon\mathrm{KL}(g|\omega)+\phi_2^n(g\cdot\mathbb{I}_\delta)\to\min. \tag{43}$$

These minimization problems can be solved almost explicitly. Their respective Euler–Lagrange equations are, at each (i,j) with $\omega_{i,j}>0$,

$$0=\varepsilon\log\frac{g_{i,j}}{\omega_{i,j}}+h'\left(\sum_i|Q_i|g_{i,j}\right)\quad\text{and}\quad 0=\varepsilon\log\frac{g_{i,j}}{\omega_{i,j}}+\lambda_i,$$

where the λ_i are Lagrange multipliers to realize the constraint $g \cdot \mathbb{I}_\delta = \bar{r}^{n-1}$. After dividing these equations by ε, taking the exponential, and evaluating the marginals, one obtains in a straightforward way the following representation of the minimizers in (43):

$$\Phi_1(\omega) = \omega \cdot \operatorname{diag}\left(\frac{H_\varepsilon^{-1}(\mathbb{I}_\delta^T \cdot \omega)}{\mathbb{I}_\delta^T \cdot \omega}\right) \quad \text{and} \quad \Phi_2(\omega) = \operatorname{diag}\left(\frac{\bar{r}^{n-1}}{\omega \cdot \mathbb{I}_\delta}\right) \cdot \omega,$$

where $[H_\varepsilon^{-1}(\eta)]_j$ for given $\eta_j \geq 0$ is the solution z to the nonlinear relation

$$H_\varepsilon(z) = z \exp\left(\frac{h'(z)}{\varepsilon}\right) = \eta_j;$$

note that the equations for the components of $H_\varepsilon^{-1}(\eta)$ are decoupled.

Finally, a significant reduction in the computational complexity of the algorithm is achieved by taking advantage of the Dyadic structure of g and s that is inherited from each iteration to the next: at each stage k, there are vectors $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$ such that

$$g^{(k)} = \left(\alpha^{(k)} \otimes \beta^{(k)}\right) \odot \xi, \quad s^{(k)} = u^{(k)} \otimes v^{(k)}. \tag{44}$$

Inserting this special form into (42), one obtains iteration rules for $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$ that are summarized below.

Proposition 4. *Initialize $\alpha_i^{(0)} = \beta_j^{(0)} = 1$ and $u_i^{(0)} = u_i^{(-1)} = v_j^{(0)} = v_j^{(-1)} = 1$ for all i, j, and inductively calculate $\alpha^{(k)}$, $\beta^{(k)}$ and $u^{(k)}$, $v^{(k)}$ for $k = 1, 2, \ldots$ from*

$$\alpha^{(k+1)} = \begin{cases} \alpha^{(k)} \odot u^{(k-1)} & \text{if } k \text{ odd}, \\ \frac{\bar{r}^{n-1}}{\xi \cdot (\beta^{(k+1)} \odot \mathbb{I}_\delta)} & \text{if } k \text{ even}, \end{cases}$$

$$\beta^{(k+1)} = \begin{cases} \frac{H_\varepsilon^{-1}\left((\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}\right)}{\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)} & \text{if } k \text{ odd}, \\ \beta^{(k)} \odot v^{(k-1)} & \text{if } k \text{ even}, \end{cases}$$

$$u^{(k+1)} = \frac{\alpha^{(k)} \odot u^{(k-1)}}{\alpha^{(k+1)}}, \quad v^{(k+1)} = \frac{\beta^{(k)} \odot v^{(k-1)}}{\beta^{(k+1)}},$$

with the understanding that for odd k, one calculates $\alpha^{(k+1)}$ first and $\beta^{(k+1)}$ next, and the other way around for even k. Further, the quotient $\frac{0}{0}$ is interpreted as 0.

Then (44) produces the iterates $g^{(k)}$ and $s^{(k)}$ of (42).

Proof. We assume that $g^{(\ell)} = (\alpha^{(\ell)} \otimes \beta^{(\ell)}) \odot \xi$ and $s^{(\ell)} = u^{(\ell)} \otimes v^{(\ell)}$ are in the form of (44) for all $\ell = 0, 1, 2, \ldots, k$; we show that with $\alpha^{(k+1)}$, $\beta^{(k+1)}$ and $u^{(k+1)}$, $v^{(k+1)}$ defined as above, $g^{(k+1)} = (\alpha^{(k+1)} \otimes \beta^{(k+1)}) \odot \xi$ and $s^{(k+1)} = u^{(k+1)} \otimes v^{(k+1)}$ satisfy the original induction formula (42).

First, note that

$$\begin{aligned} g^{(k)} \odot s^{(k-1)} &= (\alpha^{(k)} \otimes \beta^{(k)}) \odot \xi \odot (u^{(k-1)} \otimes v^{(k-1)}) \\ &= \left((\alpha^{(k)} \odot u^{(k-1)}) \otimes (\beta^{(k)} \odot v^{(k-1)})\right) \odot \xi. \end{aligned}$$

Further, we shall use the rule that for arbitrary vectors p, q, and x, and matrices h,

$$[(p \otimes q) \odot h] \cdot x = p \odot [h \cdot (q \odot x)].$$

Now, if k is odd, then

$$\begin{aligned}
&\alpha^{(k+1)} \otimes \beta^{(k+1)} = \frac{\bar{r}^{n-1}}{\xi \cdot (\beta^{(k+1)} \odot \mathbb{I}_\delta)} \otimes \beta^{(k+1)} \\
&= \left(\frac{r^{n-1}}{\alpha^{(k)} \odot u^{(k-1)} \odot (\xi \cdot (\beta^{(k)} \odot v^{(k-1)} \odot \mathbb{I}_\delta))} \odot \alpha^{(k)} \odot v^{(k-1)} \right) \otimes (\beta^{(k)} \odot v^{(k-1)}) \\
&= \operatorname{diag}\left(\frac{r^{n-1}}{(g^{(k)} \odot s^{(k-1)}) \cdot \mathbb{I}_\delta} \right) \cdot \frac{g^{(k)} \odot s^{(k-1)}}{\xi} = \frac{\Phi_1(g^{(k)} \odot s^{(k-1)})}{\xi} = \frac{g^{(k+1)}}{\xi}.
\end{aligned}$$

In the same spirit, for k even, one shows that

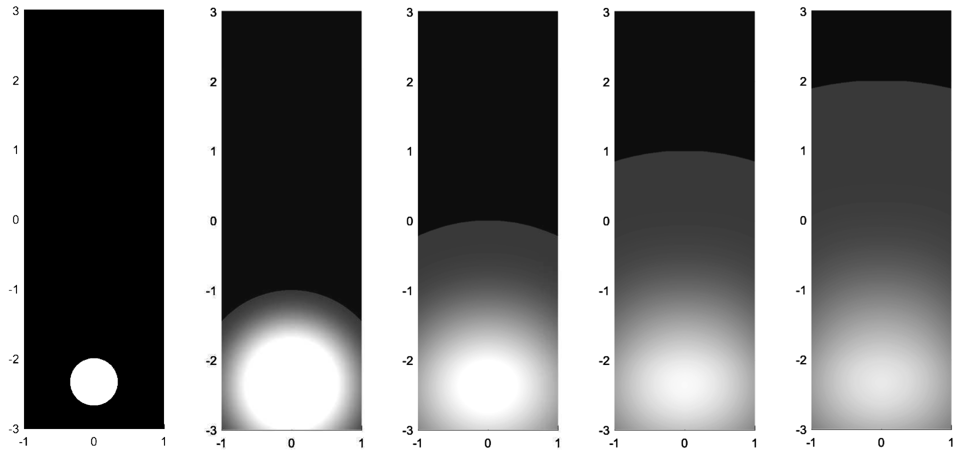
$$\begin{aligned}
\alpha^{(k+1)} \otimes \beta^{(k+1)} &= (\alpha^{(k)} \odot u^{(k-1)}) \\
&\quad \otimes \left(\beta^{(k)} \odot v^{(k-1)} \odot \frac{H_\varepsilon^{-1}\big((\xi^T \cdot (\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}\big)}{(\xi^{\cdot}(\alpha^{(k+1)} \odot \mathbb{I}_\delta)) \odot \beta^{(k)} \odot v^{(k-1)}} \right) \\
&= \frac{\Phi_2(g^{(k)} \odot s^{(k-1)})}{\xi} = \frac{g^{(k+1)}}{\xi}.
\end{aligned}$$

Finally,

$$\begin{aligned}
u^{(k+1)} \otimes v^{(k+1)} &= \frac{\alpha^{(k)} \odot u^{(k-1)}}{\alpha^{(k+1)}} \otimes \frac{\beta^{(k)} \odot v^{(k-1)}}{\beta^{(k+1)}} \\
&= \frac{(\alpha^{(k)} \otimes \beta^{(k)}) \odot \xi \odot (u^{(k-1)} \otimes v^{(k-1)})}{(\alpha^{(k+1)} \otimes \beta^{(k+1)}) \odot \xi} \\
&= \frac{g^{(k)} \odot s^{(k-1)}}{g^{(k+1)}} = s^{(k+1)}. \qquad \square
\end{aligned}$$

5.4. **Implementation.** Based on the discussion above, we introduce a numerical scheme for an approximate solution of the initial value problem for (1) as follows. Choose a spatial mesh width $\delta > 0$ and an entropic regularization parameter $\varepsilon > 0$. Further, define a suitable approximation $\tilde{\mathbf{c}}$ of the cost function $\mathbf{c}$ that is constant on cubes $Q_i \times Q_j$, for instance as in (36), and an approximation r^0 of the initial condition, for instance $r_i^0 = ⨍_{Q_i} \rho^0(x)\,\mathrm{d}x$.

From a given r^{n-1}, the next iterate r^n is obtained as a second marginal, $r_j^n = \sum_i |Q_i| g_{i,j}^n$, of the minimizer g^n to the variational problem (37) or, equivalently, (41). To calculate g^n from r^{n-1}, we use Dykstra's algorithm (42) as shown in Proposition 4 above. That is, we alternately calculate the scaling factors $\alpha^{(k)}$, $\beta^{(k)}$, and the auxiliary vectors $u^{(k)}$ and $v^{(k)}$, using the iteration from Proposition 4 with $\bar{r} := r^{n-1}$. The updates of $\alpha^{(k+1)}$, $u^{(k+1)}$, and $v^{(k+1)}$ are obviously very efficient. To calculate the term involving H_ε^{-1} in the update for $\beta^{(k+1)}$, we use a Newton iteration, which converges in a few steps. The iteration in k is repeated until the changes in α and β from one iteration to the next meets a smallness condition. Then $g_{i,j}^n := \alpha_i^{(k)} \xi_{i,j} \beta_j^{(k)}$.



(a) The initial probability density $\rho^{(0)}$.

(b) The density after the first iteration $\rho^{(1)}$.

(c) The density after the second iteration $\rho^{(2)}$.

(d) The density after the third iteration $\rho^{(3)}$.

(e) The density after the fourth iteration $\rho^{(4)}$.

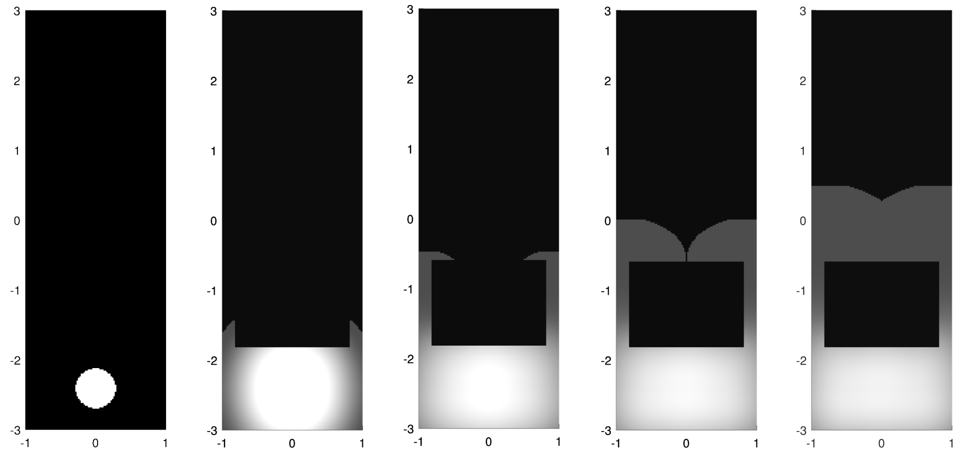
FIGURE 5.5.1. A support of the density propagates at most with "light speed". The greyscale possesses a step from black (representing density 0) to the darkest displayed gray (representing the smallest double-precision floating-point number greater than 0) in order to illustrate the support of ρ moving with finite speed. The iteration was performed on a grid of 400×1200 uniformly distributed gridpoints on $[-1, 1] \times [-3, 3] \subset \mathbb{R}^2$ with parameters $\tau = 1$, $\varepsilon = 0.5$, $m = 2$ and lightspeed 1. As an initial distribution we used $\rho^{(0)}$ with its mass equally distributed on its support, a ball with radius 0.8 centered at $(0, -2.8)$. This way, the uppermost points in the support of $\rho^{(0)}$ have ordinate $y = -2$ and the propagation with lightspeed can be observed over the displayed plots.

5.5. **Numerical experiments.** In our expriments, we study the application of our discretization method to the equation

$$\partial_t \rho = \nabla \cdot \left[\rho \, \frac{\nabla \rho}{\sqrt{1 + |\nabla \rho|^2}} \right],$$

which is (1) with the relativistic cost $\mathbf{C}(v) = 1 - \sqrt{1 - |v|^2}$ and the energy from (4) with $h(r) = r^2/2$. Naturally, all experiments are carried out on finite domains Ω, which are either of dimension $d = 1$ or $d = 2$.

5.5.1. *Finite speed of propagation.* In the first experiment, we study how the flux limitation numerically becomes manifest. We consider the rectangular box $\Omega = (-1, 1) \times (-3, 3)$ in $\mathbb{R}^2$, and a discretization by squares of edge length 0.005. Our time step is $\tau = 1$. The chosen discrete approximation $\tilde{\mathbf{c}}$ to the cost function $\mathbf{c}$ is of the type (36), so in particular we set $\xi_{i,j} = 0$ if $|x_i - x_j| > 1$. We chose a (discontinuous) initial condition $\rho^{(0)}$ that is uniformly distributed on a ball.



(a) The initial probability density ρ^0.

(b) The density after two iterations ρ^2.

(c) The density after four iterations ρ^4.

(d) The density after five iterations ρ^5.

(e) The density after six iterations ρ^6.

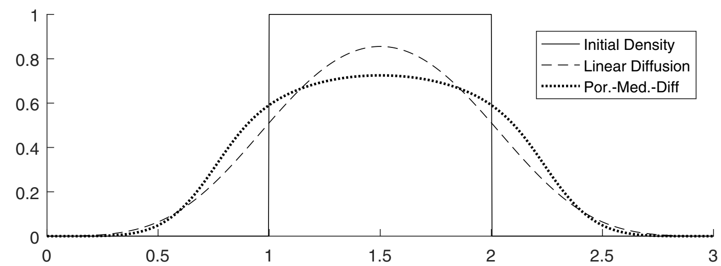
FIGURE 5.5.2. Evolution of a density around a obstacle. Grayscale as in Figure 5.5.1 . The iteration was performed on the remaining part of a 100×300, equidistant, quadrilateral grid on $[-1, 1] \times [-3, 3]$, after the obstacle points were removed. The parameters were $\tau = 0.5$, $\varepsilon = 0.1$, $m = 2$ and again, lightspeed was set to 1. As an initial distribution we used ρ^0 with its mass equally distributed over a small ball with center $(x, y) = (2, 1.2)$.

Figure 5.5.1 shows (from left to right) the initial density, and then the solution at $t = \tau, 2\tau, 3\tau$ and $t = 4\tau$. In order to make the finite speed of propagation of the support visible, we set the grayscale to black for $\rho(x) = 0$, and to a gray visibly lighter than black as soon as $\rho(x) > 0$. Additionally, the support of the initial density is chosen as a ball, positioned at $(0, -2.8)$ and with radius 0.8. This way, $\rho^{(0)}$ is supported in $y \leq -2$ and the propagation of the support with lightspeed can easily be observed as the support increases its radius by 1 in each step.

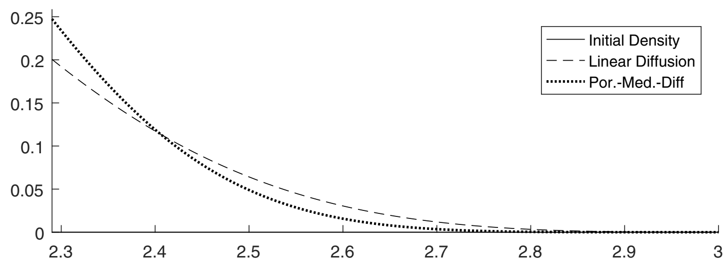
5.5.2. *Motion around obstacles.* The algorithm we used here allows for an easy implementation of impentrable obstacles in the domain. The only thing that has to be altered is the matrix ξ. There the columns and rows corresponding to a point lying within the obstacle have to be set to zero and components of ξ corresponding to a pair of points whose connecting line segment crosses the obstacle have to be recalculated (cf. (12)).

In *Figure* 5.5.2 we have realized an impenetrable box and a density flowing around it. Again we have used the step in the grayscale to illustrate the support of ρ and again we can observe the finite speed of propagation.

5.5.3. *Comparison: Linear diffusion and porous medium diffusion.* The iteration can be carried out with porous medium as well as with linear diffusion. In Figure 5.5.3 some features of the two different diffusions can be compared. The figure shows the result of iterating both with the same initial data. Note that the iteration is



(a) Comparison between linear diffusion and porous medium diffusion with parameter $m = 5$.



(b) Magnification of the comparison.

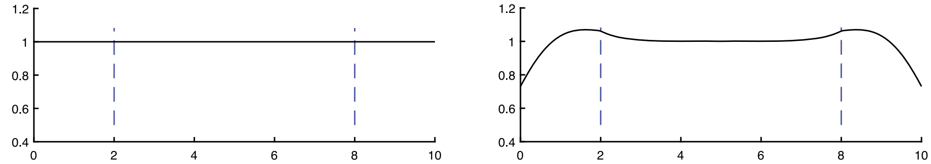
Figure 5.5.3. The iteration for the same discontinuous initial data depicted by a solid line. The iteration is performed on an equidistant grid with 1000 grid points with time-step $\tau = 0.02$ and time horizon $T = 1$ and entropic regularization parameter $\varepsilon = 0.04$.

already advanced enough that the fronts that can be expected with flux-limitation and such discontinuous initial data are already dispersed.

Porous medium diffusion disperses the mass faster than linear diffusion where there is a high density and is slower when there is low density which results in the lower density for our porous medium evolution around $x = 0$ compared to linear diffusion. On the other hand, as can clearly be seen in the magnification, linear diffusion disperses the mass faster for densities close to zero.

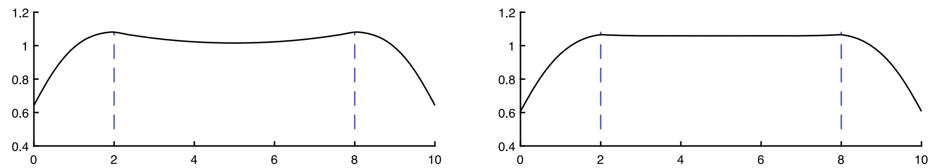
Finally, though it cannot be observed easily in the plots, the support of both the linear diffusion evolution and the porous medium evolution expands with the same velocity, which is our lightspeed.

5.5.4. *Edge effect.* Our last experiment is posed on a one-dimensional interval $[0, 10]$, which is discretized with 1000 intervals of equal length. The result in Figure 5.5.4 highlights an undesired effect at the edges: although we initialize with a uniform distribution (which corresponds to a stationary solution), the density very quickly becomes nonhomogeneous near the boundary points. In the first order, the solution represents the second marginal of the matrix ξ; since the matrix is "cut off" at the boundary, there is a lack of mass near the end points. The energy introduces a second order effect, which tries to compensate the primary effect by transporting mass from the bulk to the edges.



(a) The initial probability density $\rho^{(0)}$.

(b) The density after the first iteration $\rho^{(1)}$.



(c) The density after the fourth iteration $\rho^{(4)}$.

(d) The steady state corresponding to this initial vector.

FIGURE 5.5.4. The edge effect caused by the blurring with the Gibbs kernel. Iteration performed on a 1000 grid point equidistantly distributed on $[0, 10]$ with parameters $\tau = 2$, $\varepsilon = 2$, $m = 2$. As an initial distribution we used $\rho_i^0 = 1$. The horizontal, dotted lines are drawn at $x = 0 + \tau$ and $10 - \tau$ and mark the width of the edge effect.

This effect is that the stronger it is, the larger the entropic regularization parameter $\varepsilon > 0$; the pictures have been produced for a "huge" value $\varepsilon = 2$.

REFERENCES

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations **10** (2005), no. 3, 309–360. MR2123134

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600

[3] F. Andreu, V. Caselles, and J. M. Mazón, *Some regularity results on the 'relativistic' heat equation*, J. Differential Equations **245** (2008), no. 12, 3639–3663, DOI 10.1016/j.jde.2008.06.024. MR2462697

[4] F. Andreu, V. Caselles, J. M. Mazón, and S. Moll, *The Dirichlet problem associated to the relativistic heat equation*, Math. Ann. **347** (2010), no. 1, 135–199, DOI 10.1007/s00208-009-0428-3. MR2593288

[5] J.-D. Benamou, G. Carlier, Q. Mérigot, and É. Oudet, *Discretization of functionals involving the Monge-Ampère operator*, Numer. Math. **134** (2016), no. 3, 611–636, DOI 10.1007/s00211-015-0781-y. MR3555350

[6] J. Bertrand and M. Puel, *The optimal mass transport problem for relativistic costs*, Calc. Var. Partial Differential Equations **46** (2013), no. 1-2, 353–374, DOI 10.1007/s00526-011-0485-9. MR3016512

[7] A. Blanchet, V. Calvez, and J. A. Carrillo, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model*, SIAM J. Numer. Anal. **46** (2008), no. 2, 691–721, DOI 10.1137/070683337. MR2383208

[8] Y. Brenier, *Extended Monge-Kantorovich Theory*, Optimal transportation and applications (Martina Franca, 2001), Lecture Notes in Math., vol. 1813, Springer, Berlin, 2003, pp. 91–121, DOI 10.1007/978-3-540-44857-0_4. MR2006306

[9] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal. **49** (2017), no. 2, 1385–1418, DOI 10.1137/15M1050264. MR3635459

[10] J. A. Carrillo, V. Caselles, and S. Moll, *On the relativistic heat equation in one space dimension*, Proc. Lond. Math. Soc. (3) **107** (2013), no. 6, 1395–1423, DOI 10.1112/plms/pdt015. MR3149851

[11] J. A. Carrillo, K. Craig, and F. S. Patacchini, *A blob method for diffusion*, Calc. Var. Partial Differential Equations **58** (2019), no. 2, Art. 53, 53, DOI 10.1007/s00526-019-1486-3. MR3913840

[12] J. A. Carrillo, B. Düring, D. Matthes, and D. S. McCormick, *A Lagrangian scheme for the solution of nonlinear diffusion equations using moving simplex meshes*, J. Sci. Comput. **75** (2018), no. 3, 1463–1499, DOI 10.1007/s10915-017-0594-5. MR3798109

[13] J. A. Carrillo and J. S. Moll, *Numerical simulation of diffusive and aggregation phenomena in nonlinear continuity equations by evolving diffeomorphisms*, SIAM J. Sci. Comput. **31** (2009/10), no. 6, 4305–4329, DOI 10.1137/080739574. MR2566595

[14] V. Caselles, *Convergence of the 'relativistic' heat equation to the heat equation as* $c \to \infty$, Publ. Mat. **51** (2007), no. 1, 121–142, DOI 10.5565/PUBLMAT_51107_06. MR2307149

[15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling algorithms for unbalanced optimal transport problems*, Math. Comp. **87** (2018), no. 314, 2563–2609, DOI 10.1090/mcom/3303. MR3834678

[16] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport.* In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[17] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17, DOI 10.1137/S0036141096303359. MR1617171

[18] O. Junge, D. Matthes, and H. Osberger, *A fully discrete variational scheme for solving nonlinear Fokker-Planck equations in multiple space dimensions*, SIAM J. Numer. Anal. **55** (2017), no. 1, 419–443, DOI 10.1137/16M1056560. MR3614682

[19] O. Junge and B. Söllner, *A convergent Lagrangian discretization of p-Laplace equations in one space dimension*, Archive preprint arxiv:1906.01321, 2019.

[20] P. Laurençot and B.-V. Matioc, *A gradient flow approach to a thin film approximation of the Muskat problem*, Calc. Var. Partial Differential Equations **47** (2013), no. 1-2, 319–341, DOI 10.1007/s00526-012-0520-5. MR3044141

[21] J. Maas and D. Matthes, *Long-time behavior of a finite volume discretization for a fourth order diffusion equation*, Nonlinearity **29** (2016), no. 7, 1992–2023, DOI 10.1088/0951-7715/29/7/1992. MR3521637

[22] D. Matthes and H. Osberger, *Convergence of a variational Lagrangian scheme for a nonlinear drift diffusion equation*, ESAIM Math. Model. Numer. Anal. **48** (2014), no. 3, 697–726, DOI 10.1051/m2an/2013126. MR3177862

[23] D. Matthes and H. Osberger, *A convergent Lagrangian discretization for a nonlinear fourth-order equation*, Found. Comput. Math. **17** (2017), no. 1, 73–126, DOI 10.1007/s10208-015-9284-6. MR3600850

[24] R. J. McCann and M. Puel, *Constructing a relativistic heat flow by transport time steps* (English, with English and French summaries), Ann. Inst. H. Poincaré Anal. Non Linéaire **26** (2009), no. 6, 2539–2580, DOI 10.1016/j.anihpc.2009.06.006. MR2569908

[25] F. Otto, *Lubrication approximation with prescribed nonzero contact angle*, Comm. Partial Differential Equations **23** (1998), no. 11-12, 2077–2164, DOI 10.1080/03605309808821411. MR1662172

[26] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174, DOI 10.1081/PDE-100002243. MR1842429

[27] G. Peyré, *Entropic approximation of Wasserstein gradient flows*, SIAM J. Imaging Sci. **8** (2015), no. 4, 2323–2351, DOI 10.1137/15M1010087. MR3413589

[28] G. Peyré and M. Cuturi, *Computational optimal transport*, to appear in Foundations and Trends in Machine Learning, 2018.

[29] P. Rosenau, *Tempered diffusion: A transport process with propagating fronts and inertial delay*, Physical Review A, **46**:R7371, (1992), no. 12.

Technische Universität München, Zentrum Mathematik Boltzmannstr. 3 D-85748 Garching, Germany
Email address: matthes@ma.tum.de

Technische Universität München, Zentrum Mathematik Boltzmannstr. 3 D-85748 Garching, Germany
Email address: soellneb@ma.tum.de