Wasserstein Gradient Flows of MMD Functionals with Distance Kernel and Cauchy Problems on Quantile Functions

Richard Duong* Viktor Stein* Robert Beinert* Johannes Hertrich† Gabriele Steidl*

August 15, 2024

Abstract

We give a comprehensive description of Wasserstein gradient flows of maximum mean discrepancy (MMD) functionals $\mathcal{F}_\nu := \mathrm{MMD}_K^2(\cdot, \nu)$ towards given target measures ν on the real line, where we focus on the negative distance kernel $K(x, y) := -|x - y|$. In one dimension, the Wasserstein-2 space can be isometrically embedded into the cone $\mathcal{C}(0, 1) \subset L_2(0, 1)$ of quantile functions leading to a characterization of Wasserstein gradient flows via the solution of an associated Cauchy problem on $L_2(0, 1)$. Based on the construction of an appropriate counterpart of $\mathcal{F}_\nu$ on $L_2(0, 1)$ and its subdifferential, we provide a solution of the Cauchy problem. For discrete target measures ν, this results in a piecewise linear solution formula. We prove invariance and smoothing properties of the flow on subsets of $\mathcal{C}(0, 1)$. For certain $\mathcal{F}_\nu$-flows this implies that initial point measures instantly become absolutely continuous, and stay so over time. Finally, we illustrate the behavior of the flow by various numerical examples using an implicit Euler scheme and demonstrate differences to the explicit Euler scheme, which is easier to compute, but comes with limited convergence guarantees.

1 Introduction

Wasserstein gradient flows have been examined in stochastic analysis for a long time and recently received increasing interest in machine learning, leading to intriguing research questions that often fall outside the scope of the existing theory, see, e.g., [15, 23, 29]. In this paper, we concentrate on Wasserstein gradient flows of MMD functionals $\mathcal{F}_\nu := \mathrm{MMD}_K^2(\cdot, \nu)$ towards given target measures ν. Wasserstein gradient flows of MMDs with smooth kernels K like, e.g., the Gaussian one, preserve absolutely continuous measures as well as empirical measures, so that in the latter case, just the movement of particles has to be taken into account [2]. For non-smooth kernels like, e.g., the negative distance kernel, the properties of the measure can heavily vary during the flow. Figure 1 shows the simple example of a Wasserstein gradient flow on the line starting from an initial measure $\mu_0 = \delta_{-1}$ towards the target measure $\nu = \delta_0$, where the behavior of the flow changes completely, see also Example 6.8.

*Institute of Mathematics, Technische Universität Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany, `{duong, stein, beinert, steidl}@math.tu-berlin.de`, `http://tu.berlin/imageanalysis`.

†Department of Computer Science, University College London, `j.hertrich@ucl.ac.uk`.

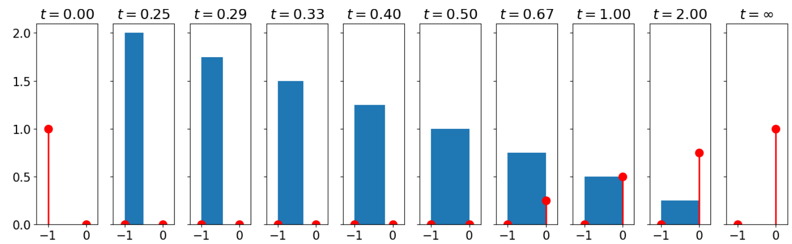

Figure 1: Wasserstein gradient flow of $\mathcal{F}_\nu$ from δ_{-1} towards $\nu = \delta_0$. The absolutely continuous part is visualized by its density in blue (area equals mass) and the atomic part by the red dotted vertical line (height equals mass). The flow changes immediately from a point measure to a uniform one with increasing support. It stays absolutely continuous until approaching the target support point 0. Then it becomes the sum of an absolutely continuous measure and a discrete one, where the weight of the latter increases in time, see [24].

In general, the characterization of Wasserstein gradient flows of MMD functionals with non-smooth kernels appears to be complicated. For the interaction energy, which is only one part of the MMD functional, we refer to [25] and in relation with potential theory to [14]; for functionals which are the sum of the interaction energy and a potential energy from $\mathcal{C}^1(\mathbb{R}^d)$ which are moreover geodesically convex, see [13]. Existence and global convergence results of MMD flows with the Coulomb kernel were proven in [8].

However, the flow examination can be simplified in one dimension, since in this case the Wasserstein spaces can be isometrically embedded into the cone $\mathcal{C}(0,1) \subset L_2(0,1)$ of quantile functions of measures. The flow of just the interaction energy, and also for a system of two interacting measures, was considered in [7, 12], see Remark 4.6. The recent preprint [20] investigates a non-local conservation law, where the interaction kernel is Lipschitz continuous and supported on the negative half line. Another special case, namely with a single point target measure was treated in [24]. Discretized versions of $\mathcal{F}_\nu$ in the one-dimensional case and their convergence properties to the continuous setting were considered in [17, 18, 22]. Our work is also different from the recent article [27], where the functional is a porous-medium type approximation of the interaction energy for the Riesz (or Coulomb) kernel $-|x-y|^r$ with $r \in (-1, 0)$, and the kernel $-|x-y|^r$ for $r > 1$ but not $r = 1$ is covered by the results in [11]. Further, connections between continuity equations and, more generally, scalar conservation laws in one dimension and their associated PDEs on quantile functions related to various Wasserstein-p distances were explored, e.g., in [10, 21, 30], but focusing on the longtime behavior of solutions. Here, we refer to the overview paper [19].

In this paper, we consider the flow of the *whole* MMD functional with attraction term *and* repulsion term for *arbitrary* starting and target measures, concentrating on the negative distance kernel $K(x, y) \coloneqq -|x-y|$. Taking the isometric embedding into account, a functional F_ν on $\mathcal{C}(0,1)$ can be associated to the MMD functional $\mathcal{F}_\nu$ on the Wasserstein space. We construct a continuous extension of this functional F_ν to the space $L_2(0,1)$ and compute its subdifferential. Then, starting with a measure μ_0 with quantile function Q_{μ_0}, we characterize the Wasserstein gradient flow of $\mathcal{F}_\nu$ via $\gamma_t \coloneqq (g(t))_\# \Lambda_{(0,1)}$, where g is the unique strong solution of the associated Cauchy problem on

$L_2(0,1)$:

$$\begin{cases} \partial_t g(t) \in -\partial F_\nu(g(t)), & t \in (0,\infty), \\ g(0) = Q_{\mu_0}. \end{cases}$$

Indeed, the strong solution of the Cauchy problem exists and stays within the cone $\mathcal{C}(0,1)$, i.e., *quantile functions remain quantile functions.* We identify further interesting invariant subset of $\mathcal{C}(0,1)$, in which the F_ν-flow remains once it has started there. In particular, we prove that the F_ν-flow preserves continuity and Lipschitz properties of the initial datum Q_{μ_0} under suitable conditions on the target Q_ν. This includes cases where point masses *cannot* form during the flow – in contrast to the example given by Figure 1. Also, the phenomenon, where initial point masses immediately explode into absolutely continuous measures, is covered by our smoothing result, see Theorem 6.5. This smoothing effect is driven by the repulsive nature of the *interaction energy* part of our MMD functional, see also [7, 12]. But note that in our case, the *potential energy* part in conjunction with a *fixed* target measure ν plays a fundamental role, whether or not this smoothing property can come into effect, and remain over time – see again Figure 1.

Moreover, we discuss the explicit pointwise solution g_s of the Cauchy problem at continuity points $s \in (0,1)$ of the quantile function Q_ν and show that the family of these functions determines the solution of the Cauchy problem via $[g(t)](s) = g_s(t)$. Furthermore, based on Euler backward and forward schemes, we illustrate the flow by some numerical examples.

Outline of the paper. In Section 2, we recall Wasserstein gradient flows, and especially flows on the line in Section 3. We take great care of the relation between quantile functions and cumulative density functions (CDFs) of probability measures and recall properties of maximal monotone operators on Hilbert spaces which we need later. Finally, we formulate our central characterization of Wasserstein gradient flows by the solution of a Cauchy problem in $L_2(0,1)$. The MMD and the functional $\mathcal{F}_\nu := \mathrm{MMD}_K^2(\cdot,\nu)$ with the distance kernel is introduced in Section 4. We determine a continuous functional F_ν on $L_2(0,1)$ whose restriction to $\mathcal{C}(0,1)$ is associated with $\mathcal{F}_\nu$, meaning that it fulfills $\mathcal{F}_\nu(\mu) = F_\nu(Q_\mu)$ for all measures μ in the Wasserstein space. We compute the subdifferential of F_ν and show that the related Cauchy problem produces indeed a flow that stays within the cone $\mathcal{C}(0,1)$. In Section 5, we provide a solution of the pointwise Cauchy problem and show that it also determines the overall solution. In particular, an explicit solution formula is given for discrete target measures ν. Section 6 deals with invariance and smoothing properties of our gradient flows. We show that certain F_ν-flows preserve (and improve) Lipschitz properties of the initial quantile Q_{μ_0} by describing the time evolution of the Lipschitz constants. Also, we prove for general targets ν that the support of the starting measure stays convex and grows monotonically. Finally, Section 7, briefly introduces Euler backward and forward schemes and illustrates properties of the flow by numerical examples. The appendix collects auxiliary and additional material.

2 Wasserstein Gradient Flows

Let $\mathcal{M}(\mathbb{R}^d)$ denote the space of σ-additive, signed Borel measures and $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$. For $\mu \in \mathcal{M}(\mathbb{R}^d)$ and a measurable map $T\colon \mathbb{R}^d \to \mathbb{R}^n$, the *push-forward* of μ via T is given by $T_\#\mu := \mu \circ T^{-1}$. We consider the *Wasserstein space* $\mathcal{P}_2(\mathbb{R}^d) := \{\mu \in$

$\mathcal{P}(\mathbb{R}^d)\colon \int_{\mathbb{R}^d} \|x\|_2^2 \,\mathrm{d}\mu(x) < \infty\}$ equipped with the *Wasserstein distance* $W_2\colon \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to [0,\infty)$,

$$W_2^2(\mu,\nu) \coloneqq \min_{\pi\in\Gamma(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|_2^2 \,\mathrm{d}\pi(x,y), \qquad \mu,\nu\in\mathcal{P}_2(\mathbb{R}^d), \tag{1}$$

where $\Gamma(\mu,\nu) \coloneqq \{\pi \in \mathcal{P}_2(\mathbb{R}^d\times\mathbb{R}^d) : (\pi_1)_\#\pi = \mu,\ (\pi_2)_\#\pi = \nu\}$ and $\pi_i(x) \coloneqq x_i,\ i = 1,2$ for $x = (x_1,x_2)$, see e.g., [41, 42]. Further, $\|\cdot\|_2$ denotes the Euclidean norm on $\mathbb{R}^d$. The set of optimal transport plans π realizing the minimum in (1) is denoted by $\Gamma^{\mathrm{opt}}(\mu,\nu)$.

A curve $\gamma\colon I \to \mathcal{P}_2(\mathbb{R}^d)$ on an interval $I \subset \mathbb{R}$, is called a *geodesic* if there exists a constant $C \ge 0$ such that

$$W_2(\gamma_{t_1},\gamma_{t_2}) = C|t_2 - t_1|, \qquad \text{for all } t_1,t_2\in I.$$

There also exists the notion of *generalized geodesics*, which in one dimension coincides with that of geodesics, so we do not introduce it here.

The Wasserstein space is a geodesic space, meaning that any two measures $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$ can be connected by a geodesic. For $\lambda\in\mathbb{R}$, a function $\mathcal{F}\colon \mathcal{P}_2(\mathbb{R}^d)\to(-\infty,\infty]$ is called *λ-convex along geodesics* if, for every $\mu,\nu\in\mathrm{dom}(\mathcal{F}) \coloneqq \{\mu\in\mathcal{P}_2(\mathbb{R}^d) : \mathcal{F}(\mu)<\infty\}$, there exists at least one geodesic $\gamma\colon[0,1]\to\mathcal{P}_2(\mathbb{R}^d)$ between μ and ν such that

$$\mathcal{F}(\gamma_t) \le (1-t)\ \mathcal{F}(\mu) + t\ \mathcal{F}(\nu) - \tfrac{\lambda}{2}\, t(1-t)\, W_2^2(\mu,\nu), \qquad t\in[0,1].$$

In the case $\lambda = 0$, we just speak about convex functions.

Let $L_{2,\mu}$ denote the Bochner space of (equivalence classes of) functions $\xi : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\xi\|^2_{L_{2,\mu}} \coloneqq \int_{\mathbb{R}^d} \|\xi(x)\|_2^2 \,\mathrm{d}\mu(x) < \infty$. The *(regular) tangent space* at $\mu\in\mathcal{P}_2(\mathbb{R}^d)$ is given by

$$\mathrm{T}_\mu\mathcal{P}_2(\mathbb{R}^d) \coloneqq \overline{\{\lambda(T-\mathrm{Id}) : (\mathrm{Id},T)_\#\mu \in \Gamma^{\mathrm{opt}}(\mu,T_\#\mu),\ \lambda>0\}}^{L_{2,\mu}}.$$

For a proper and lower semicontinuous (lsc) function $\mathcal{F}\colon \mathcal{P}_2(\mathbb{R}^d)\to(-\infty,\infty]$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$, the *reduced Fréchet subdifferential $\partial\mathcal{F}(\mu)$ at μ* consists of all $\xi\in L_{2,\mu}$ satisfying

$$\mathcal{F}(\nu)-\mathcal{F}(\mu) \ge \inf_{\pi\in\Gamma^{\mathrm{opt}}(\mu,\nu)} \int_{\mathbb{R}^{2d}} \langle \xi(x), y-x\rangle \,\mathrm{d}\pi(x,y) + o(W_2(\mu,\nu)) \tag{2}$$

for all $\nu\in\mathcal{P}_2(\mathbb{R}^d)$.

A curve $\gamma\colon I\to\mathcal{P}_2(\mathbb{R}^d)$ is (locally) p-absolutely continuous for $p\in[1,\infty]$ [1, Def. 1.1.1] if there exists a $m\in L_p(I,\mathbb{R})$ (resp. $L_{p,\mathrm{loc}}(I,\mathbb{R})$) such that

$$W_2(\gamma_t,\gamma_s) \le \int_s^t m(r)\,\mathrm{d}r \qquad \text{for all } s,t\in I, s<t,$$

and we omit the p if $p=1$. By [1, Thm. 8.3.1], if γ is *absolutely continuous*, then there exists a Borel velocity field v of functions $v_t\colon \mathbb{R}^d\to\mathbb{R}^d$ with with $\int_I \|v_t\|_{L_{2,\gamma_t}}\,\mathrm{d}t<\infty$ such that the continuity equation

$$\partial_t\gamma_t + \nabla_x\cdot(v_t\,\gamma_t) = 0 \tag{3}$$

holds on $I\times\mathbb{R}^d$ in the distributive sense, i.e., for all $\varphi\in C_c^\infty(I\times\mathbb{R}^d)$ it holds

$$\int_I\int_{\mathbb{R}^d} \partial_t\varphi(t,x) + v_t(x)\cdot\nabla_x\,\varphi(t,x)\,\mathrm{d}\gamma_t(x)\,\mathrm{d}t = 0.$$

A locally 2-absolutely continuous curve $\gamma\colon (0,\infty) \to \mathcal{P}_2(\mathbb{R}^d)$ with velocity field $v_t \in \mathrm{T}_{\gamma_t}\mathcal{P}_2(\mathbb{R}^d)$ is called *Wasserstein gradient flow with respect to* $\mathcal{F}\colon \mathcal{P}_2(\mathbb{R}^d) \to (-\infty,\infty]$ if

$$v_t \in -\partial \mathcal{F}(\gamma_t) \quad \text{for a.e. } t > 0.$$

We have the following theorem which holds also true in $\mathbb{R}^d$ when switching to so-called generalized geodesics, see [1, Thm. 11.2.1].

Theorem 2.1 (Existence and uniqueness of Wasserstein gradient flows)**.** *Let $\mathcal{F}\colon \mathcal{P}_2(\mathbb{R}) \to (-\infty,\infty]$ be bounded from below, lower semicontinuous (lsc) and λ-convex along geodesics and $\mu_0 \in \overline{\mathrm{dom}(\mathcal{F})}$. Then there exists a unique Wasserstein gradient flow $\gamma\colon (0,\infty) \to \mathcal{P}_2(\mathbb{R})$ with respect to $\mathcal{F}$ with $\gamma(0+) = \mu_0$. Furthermore, the piecewise constant curves constructed from the iterates of the minimizing movement scheme*

$$\mu_{n+1} \coloneqq \operatorname*{arg\,min}_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \Big\{ \mathcal{F}(\mu) + \frac{1}{2\tau} W_2^2(\mu_n, \mu) \Big\}, \qquad \tau > 0, \tag{4}$$

i.e., γ_τ defined by $\gamma_\tau|_{(n\tau,(n+1)\tau]} \coloneqq \mu_n$, $n = 0,1,\ldots,$ converge locally uniformly to γ as $\tau \downarrow 0$.

If $\lambda > 0$, then $\mathcal{F}$ admits a unique minimizer $\bar\mu \in \mathcal{P}_2(\mathbb{R})$ and we observe exponential convergence:

$$W_2(\gamma_t, \bar\mu) \le e^{-\lambda t} W_2(\mu_0, \bar\mu) \quad \text{and} \quad \mathcal{F}(\gamma_t) - \mathcal{F}(\bar\mu) \le e^{-2\lambda t}\big(\mathcal{F}(\mu_0) - \mathcal{F}(\bar\mu)\big).$$

If $\lambda = 0$ and $\bar\mu$ is a minimizer of $\mathcal{F}$, then we have

$$\mathcal{F}(\gamma_t) - \mathcal{F}(\bar\mu) \le \frac{1}{2t} W_2^2(\mu_0, \bar\mu).$$

3 Wasserstein Gradient Flows in 1D

In the following, we recall that $\mathcal{P}_2(\mathbb{R})$ can be isometrically embedded into $L_2(0,1)$ via so-called quantile functions. This reduces Wasserstein gradient flows in one dimension to the consideration of gradient flows in the Hilbert space $L_2(0,1)$, or more precisely, in the cone of quantile functions. We will see in the main Theorem 3.5 of this section, that we have finally to deal with a Cauchy inclusion problem.

The *cumulative distribution function* (CDF) of $\nu \in \mathcal{P}(\mathbb{R})$ is given by

$$R_\nu^+\colon \mathbb{R} \to [0,1], \quad R_\nu^+(x) \coloneqq \nu\big((-\infty,x]\big),$$

and its *quantile function* by

$$Q_\nu\colon (0,1) \to \mathbb{R}, \quad Q_\nu(s) \coloneqq \min\{x \in \mathbb{R} : R_\nu^+(x) \ge s\}.$$

Remark 3.1. *It is easy to check that Q_μ is strictly increasing if and only if R_μ^+ is continuous. Further, Q_μ is continuous if and only if R_μ^+ is strictly increasing on $\overline{(R_\mu^+)^{-1}(0,1)}$.*

We will also need the function

$$R_\nu^-\colon \mathbb{R} \to [0,1], \quad R_\nu^-(x) \coloneqq \nu\big((-\infty,x)\big).$$

Figure 2: Functions R_μ^+ (left), R_μ^- (middle) and Q_μ (right) for a discrete measure $\mu \coloneqq \sum_{k=1}^4 w_k \delta_{x_k}$ with $W_k \coloneqq \sum_{j=1}^k w_j$.

The functions R_ν^+, R_ν^- and Q_ν are monotonically increasing with only countably many discontinuities. They are visualized for a discrete measure in Fig. 2. For a nice overview on quantile functions in convex analysis, see also [36]. If $\nu(\{x\}) = 0$ for all $x \in \mathbb{R}$, then the functions $R_\nu \coloneqq R_\nu^- = R_\nu^+$ are continuous. In general, the points of continuity $x \in \mathbb{R}$ of R_ν^+ and R_ν^- coincide, and there it holds $R_\nu^-(x) = R_\nu^+(x)$. Generally, we have $R_\nu^- \leq R_\nu^+$ such that

$$[R_\nu^-(x), R_\nu^+(x)] \neq \emptyset \qquad \text{for all } x \in \mathbb{R}\,.$$

Both R_ν^- and Q_ν are left-continuous and, since they are increasing, also lower semicontinuous (lsc), whereas R_ν^+ is right-continuous and, since it is increasing, upper semicontinuous (usc). Further, we have for $x \in \mathbb{R}$ that

$$R_\nu^+(x) = \begin{cases} \max\{s \in (0,1) : Q_\nu(s) \leq x\}, & \text{if } x \in [\inf Q_\nu(s), \sup Q_\nu(s)), \\ 1, & \text{if } x \geq \sup Q_\nu(s), \\ 0, & \text{if } x < \inf Q_\nu(s), \end{cases} \tag{5}$$

where the infimum and supremum is taken over all $s \in (0,1)$, and

$$R_\nu^-(x) = \sup\{s \in (0,1) : Q_\nu(s) < x\} \quad \text{for} \quad x > \inf_{s \in (0,1)} Q_\nu(s). \tag{6}$$

Note that formula (5) is different from an erroneous one in [36]. Moreover, it holds the Galois inequality, which states that

$$Q_\nu(s) \leq x \quad \text{if and only if} \quad s \leq R_\nu^+(x), \tag{7}$$

The quantile functions of measures in $\mathcal{P}_2(\mathbb{R})$ form a closed, convex cone

$$\mathcal{C}(0,1) \coloneqq \{f \in L_2(0,1) : f \text{ is increasing (a.e.)}\},$$

see also Remark A.1 in the appendix.

By the following theorem, see, e.g., [41, Thm. 2.18], the mapping $\mu \mapsto Q_\mu$ is an isometric embedding of $\mathcal{P}_2(\mathbb{R})$ into $L_2(0,1)$.

Theorem 3.2. *For $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$, the quantile function $Q_\mu \in \mathcal{C}(0,1)$ satisfies $\mu = (Q_\mu)_\# \Lambda_{(0,1)}$ with the Lebesgue measure Λ on $(0,1)$ and*

$$W_2^2(\mu, \nu) = \int_0^1 |Q_\mu(s) - Q_\nu(s)|^2 \, \mathrm{d}s.$$

Thus, instead of working with $\mathcal{F} : \mathcal{P}_2(\mathbb{R}) \to (-\infty, \infty]$, we can just deal with associated functions

$$F\colon L_2(0,1) \to (-\infty, \infty] \quad \text{with} \quad F(Q_\mu) \coloneqq \mathcal{F}(\mu).$$

Note that this relation determines F only on $\mathcal{C}(0,1)$ and several extension to the whole space $L_2(0,1)$ are possible. One possibility would be to extend F outside of $\mathcal{C}(0,1)$ by ∞. Yet, later in this work we will deal with a continuous extension of a specific functional which is everywhere finite on $L_2(0,1)$.

Now, instead of the (reduced) Fréchet subdifferential (2), we will use the *regular subdifferential* of functions $F\colon L_2(0,1) \to (-\infty, \infty]$ defined by

$$\partial F(u) \coloneqq \big\{v \in L_2(0,1) : F(w) \geq F(u) + \langle v, w-u\rangle + o(\|w-u\|) \text{ for all } w \in L_2(0,1)\big\}.$$

If F is convex, then the o-term can be skipped.

The following theorem collects well-known properties of subdifferential operators on Hilbert spaces [3]. Here, the *domain* of a multivalued operator $A\colon L_2(0,1) \to 2^{L_2(0,1)}$ is denoted by $\mathrm{dom}(A) \coloneqq \{u \in L_2(0,1) : Au \neq \emptyset\}$.

Theorem 3.3. *Let $F : L_2(0,1) \to (-\infty,\infty]$ be proper and convex. Then $\partial F : L_2(0,1) \to 2^{L_2(0,1)}$ is a monotone operator, i.e. for every u_i and $v_i \in \partial F(u_i)$, $i = 1,2$ it holds*

$$\langle u_1 - u_2, v_1 - v_2\rangle \geq 0.$$

If F is in addition lsc, then ∂F is maximal monotone and thus for any $\varepsilon > 0$,

$$\mathrm{Range}(I + \varepsilon\partial F) = \bigcup_{u \in L_2(0,1)} (I + \varepsilon \partial F)(u) = L_2(0,1).$$

Moreover, ∂F is a closed operator, $\overline{\mathrm{dom}(F)} = \overline{\mathrm{dom}(\partial F)}$ and the resolvent

$$J_\varepsilon^{\partial F} \coloneqq (I + \varepsilon\partial F)^{-1} : L_2(0,1) \to L_2(0,1)$$

is single-valued.

Concerning the λ-convexity and lower semicontinuity of $\mathcal{F}$, we have the following proposition, whose proof is given in the appendix.

Proposition 3.4. *If $F : \mathcal{C}(0,1) \to (-\infty,\infty]$ is λ-convex, then $\mathcal{F}(\mu) \coloneqq F(Q_\mu)$, $\mu \in \mathcal{P}_2(\mathbb{R})$, is λ-convex along geodesics. If F is lsc, then the same holds true for $\mathcal{F}$.*

The next theorem is central for our further considerations. It characterizes Wasserstein gradient flows by the solution of a Cauchy problem, where we have to ensure that the solution remains in the cone $\mathcal{C}(0,1)$. To this end, recall that given an operator $A : \mathrm{dom}(A) \subseteq L_2(0,1) \to 2^{L_2(0,1)}$ and an initial function $g_0 \in L_2(0,1)$, a *strong solution* $g : [0,\infty) \to L_2(0,1)$ of the Cauchy problem

$$\begin{cases} \partial_t g(t) + Ag(t) \ \ni \ 0, \quad t \in (0,\infty), \\ g(0) \ = \ g_0, \end{cases} \tag{8}$$

is a function $g \in W_1^1\big((0,T]; L_2(0,1)\big) \cap C\big([0,\infty); L_2(0,1)\big)$ for any $T > 0$ which meets the initial condition and solves the differential inclusion in (8) pointwise for a.e. $t > 0$, where $\partial_t g(t)$ denotes the strong derivative of g at t.

Theorem 3.5. *Let $F\colon L_2(0,1) \to (-\infty,\infty]$ be a proper, convex and lsc function. Assume that for all (small) $\varepsilon > 0$, the resolvent $J_\varepsilon^{\partial F}$ maps $\mathcal{C}(0,1)$ into itself. Then, for any initial datum $g_0 \in \mathcal{C}(0,1) \cap \operatorname{dom}(\partial F)$, the Cauchy problem*

$$\begin{cases} \partial_t g(t) + \partial F(g(t)) \ \ni \ 0, \quad t \in (0,\infty), \\ g(0) \ = \ g_0, \end{cases} \tag{9}$$

has a unique strong solution $g : [0,\infty) \to \mathcal{C}(0,1)$ which is expressible by the exponential formula

$$g(t) = e^{-t\partial F} g_0 \coloneqq \lim_{n\to\infty} \left(I + \frac{t}{n}\partial F \right)^{-n} g_0. \tag{10}$$

The curve $\gamma_t \coloneqq (g(t))_\# \Lambda_{(0,1)}$ has quantile functions $Q_{\gamma_t} = g(t)$ and is a Wasserstein gradient flow of $\mathcal{F}$ with $\gamma(0+) = (g_0)_\# \Lambda_{(0,1)}$.

Proof. 1. By assumption, ∂F is the subdifferential of a proper, convex and lsc function, and hence maximal monotone by Theorem 3.3. By standard results of semigroup theory, see e.g., Theorem A.2, there exists a strong solution $g : [0,\infty) \to L_2(0,1)$ of (9) satisfying the exponential formula (10). It remains to show that starting in $g_0 \in \mathcal{C}(0,1)$, the solution remains in the cone $\mathcal{C}(0,1)$. Indeed, since $J_\varepsilon^{\partial F}$ maps $\mathcal{C}(0,1)$ into itself, we can conclude that $\left(I + \frac{t}{n}\partial F\right)^{-n} g_0 \in \mathcal{C}(0,1)$ for all $n \in \mathbb{N}$. Since $\mathcal{C}(0,1)$ is closed in $L_2(0,1)$, this also holds true when the limit in (10) is taken, hence $g(t) \in \mathcal{C}(0,1)$ for all $t \geq 0$.

2. Let $\gamma_t \coloneqq (g(t))_\# \Lambda_{(0,1)}$. First, we show that γ is locally 2-absolutely continuous. By Theorem A.2, the function g is Lipschitz continuous, so there exists a $L > 0$ such that for all $s, t \geq 0$ with $s < t$ we have

$$W_2(\gamma_t, \gamma_s) = \|g(t) - g(s)\|_{L_2(0,1)} \leq L(t-s) = \int_s^t L \,\mathrm{d}\tau,$$

where the first equality is due to Theorem 3.2. Since the constant function $m \equiv L$ is in $L_\infty([0,\infty))$, the curve $\gamma\colon [0,\infty) \to (\mathcal{P}_2(\mathbb{R}), W_2)$ is locally 2-absolutely continuous.

Next, we show that the velocity field v_t of γ_t from (3) fulfills $v_t \in -\partial\, \mathcal{F}(\gamma_t)$. To calculate v_t, we exploit [1, Prop 8.4.6] stating that, for a.e. $t \in (0,\infty)$, the velocity field satisfies

$$\begin{aligned} 0 &= \lim_{\tau\downarrow 0} \frac{W_2(\gamma_{t+\tau}, (\mathrm{Id} + \tau\, v_t)_\# \gamma_t)}{\tau} \\ &= \lim_{\tau\downarrow 0} \frac{W_2(g(t+\tau)_\# \Lambda_{(0,1)}, \big(g(t) + \tau\,(v_t \circ g(t))\big)_\# \Lambda_{(0,1)})}{\tau} \end{aligned}$$

Since $v_t \in \mathrm{T}_{\gamma_t} \mathcal{P}_2(\mathbb{R}^d)$, there exists a $\tilde\tau > 0$ such that $(\mathrm{id}, \mathrm{id} + \tilde\tau v_t)_\# \gamma_t$ becomes an optimal transport plan. Moreover, [1, Lem 7.2.1] implies that the transport plans remain optimal for all $0 < \tau \leq \tilde\tau$. For small $\tau > 0$, the mappings $\mathrm{id} + \tau v_t$ are thus optimal and, especially, monotonically increasing. Consequently, the functions $g(t) + \tau(v_t \circ g(t))$ are also monotonically increasing, and their left-continuous representatives are quantile functions. Employing the isometry to $L_2(0,1)$, we hence obtain

$$0 = \lim_{\tau\downarrow 0} \left\| \frac{g(t+\tau) - g(t)}{\tau} - v_t \circ g(t) \right\|_{L_2(0,1)} = \|\partial_t g(t) - v_t \circ g(t)\|_{L_2(0,1)}.$$

Thus, by construction (9) of g, we see that $v_t \circ g(t) \in -\partial F\left(g(t)\right)$ for a.e. t. In particular, for any $\mu \in \mathcal{P}_2(\mathbb{R})$, we obtain

$$\begin{aligned}0 &\leq F(Q_\mu) - F\left(g(t)\right) + \int_0^1 \left(v_t \circ g(t)\right)(s)\left(Q_\mu(s) - \left(g(t)\right)(s)\right) \mathrm{d}s \\ &= \mathcal{F}(\mu) - \mathcal{F}(\gamma_t) + \int_{\mathbb{R}\times\mathbb{R}} v_t(x)\,(y-x)\,\mathrm{d}\pi(x,y),\end{aligned}$$

where $\pi \coloneqq (g(t), Q_\mu)_\# \Lambda_{(0,1)}$. By Theorem 3.2, the plan π is optimal between γ_t and μ, see also [41, Thm. 2.18]. By [31, Thm. 16.1(i),(ii)], we also know that π is unique, so (2) yields that $v_t \in -\partial\,\mathcal{F}(\gamma_t)$. □

By the same arguments as in the proof of Theorem 3.5, we have the following corollary concerning invariant subsets of $\mathcal{C}(0,1)$ of F-flows.

Corollary 3.6. *Let $D \subset \mathcal{C}(0,1)$ be a closed subset and let $F\colon L_2(0,1) \to (-\infty,\infty]$ be a proper, convex and lsc function. Assume that for all (small) $\varepsilon > 0$, the resolvent $J_\varepsilon^{\partial F}$ maps D into itself. Then, the solution of the Cauchy problem* (9) *starting in $g_0 \in D$ fulfills $g(t) \in D$ for all $t \geq 0$.*

4 Flows of MMD with Distance Kernel

In this paper, we are mainly interested in Wasserstein gradient flows of MMD functionals $\mathcal{F}_\nu : \mathcal{P}_2(\mathbb{R}) \to [0,\infty)$ for the negative distance kernel. After introducing these functionals, we will define an associated functional $F_\nu : L_2(0,1) \to \mathbb{R}$ with $F_\nu(Q_\mu) = \mathcal{F}_\nu(\mu)$. More precisely, we will extend F_ν from $\mathcal{C}(0,1)$ to $L_2(0,1)$ such that its subdifferential can be easily computed.

MMDs are defined with respect to kernels $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$. In this paper, we are interested in the negative distance kernel

$$K(x,y) \coloneqq -|x-y|, \tag{11}$$

which is symmetric and *conditionally positive definite* of order one. Then we define

$$\begin{aligned}\mathrm{MMD}_K^2(\mu,\nu) \coloneqq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} K(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} K(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y) \\ + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} K(x,y)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y).\end{aligned} \tag{12}$$

The square root of the above formula defines a distance on $\mathcal{P}_2(\mathbb{R}^d)$ for many kernels of interest including the negative distance kernel (11). In particular, we have that $\mathrm{MMD}_K(\mu,\nu) \geq 0$ with equality if and only if $\mu = \nu$. Fixing the target measure ν, the third summand becomes a constant and we may consider the *MMD functional*

$$\mathcal{F}_\nu(\mu) \coloneqq \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} K(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) - \int_{\mathbb{R}^d\times\mathbb{R}^d} K(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(y).$$

The first summand is known as *interaction energy*, while the second one is called *potential energy* of $V_\nu \coloneqq \int_{\mathbb{R}^d} K(\cdot,y)\,\mathrm{d}\nu(y)$.

In the following, we are exclusively interested in $d = 1$ and the negative distance kernel (11). Note that the MMD of the negative distance kernel is also known as energy distance [39, 40] and that in one dimension we have a relation to the Cramer distance

$$\mathrm{MMD}_K^2(\mu,\nu) = \int_{-\infty}^{\infty} (R_\mu(x) - R_\nu(x))^2 \, \mathrm{d}x,$$

see [38, 26]. More precisely, we will deal with Wasserstein gradient flows of

$$\mathcal{F}_\nu(\mu) = -\frac{1}{2}\int_{\mathbb{R}\times\mathbb{R}} |x-y| \, \mathrm{d}\mu(x) \, \mathrm{d}\mu(y) + \int_{\mathbb{R}\times\mathbb{R}} |x-y| \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y). \tag{13}$$

In dimensions $d \geq 2$, neither the interaction energy nor the whole MMD_K^2 functional with the negative distance kernel are λ-convex along geodesics, see [25], so that Theorem 2.1 does not apply. We will see that this is different on the real line. Note that in 1D, but not in higher dimensions, λ-convexity along geodesics implies the stronger property of λ-convexity along so-called generalized geodesics.

Next, we propose an associated functional of (13) on $L_2(0,1)$, which is determined on $\mathcal{C}(0,1)$ by $\mathcal{F}_\nu(\mu) = F_\nu(Q_\mu)$. We consider the functional $F_\nu \colon L_2(0,1) \to \mathbb{R}$ defined by

$$\begin{aligned} F_\nu(u) &:= \int_0^1 \left((1-2s)\big(u(s) - Q_\nu(s)\big) + \int_0^1 |u(s) - Q_\nu(t)| \, \mathrm{d}t \right) \mathrm{d}s \\ &= \int_0^1 f\left(s, u(s)\right) \mathrm{d}s, \end{aligned} \tag{14}$$

where $f \colon (0,1) \times \mathbb{R} \to \mathbb{R}$ is given by

$$f(s,u) := (1-2s)\big(u - Q_\nu(s)\big) + H(u), \quad H(u) := \int_0^1 |u - Q_\nu(t)| \, \mathrm{d}t. \tag{15}$$

Indeed, by the following lemma, whose proof can be found in [24], the functional F_ν is associated to $\mathcal{F}_\nu$.

Lemma 4.1. *For $\mathcal{F}_\nu$ in (13), the functional $F_\nu \colon L_2(0,1) \to \mathbb{R}$ defined by (14) fulfills $F_\nu(Q_\mu) = \mathcal{F}_\nu(\mu)$ for all $\mu \in \mathcal{P}_2(\mathbb{R})$.*

By the next lemma, whose proof is given in the appendix, the functional F_ν has further desirable properties.

Lemma 4.2. *The functional $F_\nu \colon L_2(0,1) \to \mathbb{R}$ in (14) is convex and continuous.*

The subdifferential of F_ν can be computed explicitly.

Lemma 4.3 (Subdifferential of F_ν)**.** *For $u \in L_2(0,1)$, it holds*

$$\begin{aligned} \partial F_\nu(u) = \big\{ f \in L_2(0,1) : \\ f(s) \in 2\big[R_\nu^-\big(u(s)\big), R_\nu^+\big(u(s)\big)\big] - 2s \text{ for a.e. } s \in (0,1) \big\}. \end{aligned} \tag{16}$$

In particular, we have $\mathrm{dom}(\partial F_\nu) = L_2(0,1)$.

Proof. It holds $h \in \partial F_\nu(u)$ if and only if

$$h(s) \in \partial\big[f(s,\cdot)\big]\big(u(s)\big) \quad \text{for a.e. } s \in (0,1),$$

see, e.g., [35, Cor. 1B] or [37, Thm. 10.39]. Now,

$$\partial\big[f(s,\cdot)\big](u) = (1-2s) + \partial H(u), \tag{17}$$

so that it remains to consider the second summand. Recall that for the convex, one-dimensional function H, we have $\partial H(u) = [D_- H(u), D_+ H(u)]$ with the one-sided derivatives

$$D_\pm H(u) \coloneqq \lim_{\lambda \downarrow 0} \frac{H(u \pm \lambda) - H(u)}{\lambda}.$$

Let $h_t(u) \coloneqq |u - Q_\nu(t)|$ in the definition (15) of H. Then we conclude by Lebesgue's dominated convergence theorem that

$$\begin{aligned} D_\pm H(u) &= \lim_{\lambda \downarrow 0} \frac{H(u \pm \lambda) - H(u)}{\lambda} = \lim_{\lambda \downarrow 0} \int_0^1 \frac{h_t(u \pm \lambda) - h_t(u)}{\lambda} \,\mathrm{d}t \\ &= \int_0^1 \lim_{\lambda \downarrow 0} \frac{h_t(u \pm \lambda) - h_t(u)}{\lambda} \,\mathrm{d}t = \int_0^1 D_\pm h_t(u) \,\mathrm{d}t, \end{aligned}$$

and consequently

$$\partial H(u) = \Big[\int_0^1 D_- h_t(u)\,\mathrm{d}t, \int_0^1 D_+ h_t(u)\,\mathrm{d}t\Big].$$

Next, we have

$$\begin{aligned} &D_- h_t(u) = D_+ h_t(u) = 1 && \text{if } u > Q_\nu(t), \\ &D_- h_t(u) = D_+ h_t(u) = -1 && \text{if } u < Q_\nu(t), \\ &D_- h_t(u) = -1,\ D_+ h_t(u) = 1 && \text{if } u = Q_\nu(t), \end{aligned}$$

so that we obtain by (5) and (6) the value

$$\int_0^1 D_- h_t(u)\,\mathrm{d}t = \int_0^{R_\nu^-(u)} 1\,\mathrm{d}t + \int_{R_\nu^-(u)}^{R_\nu^+(u)} -1\,\mathrm{d}t + \int_{R_\nu(u)^+}^1 -1\,\mathrm{d}t = 2R_\nu^-(u) - 1$$

and similarly $\int_0^1 D_+ h_t(u)\,\mathrm{d}t = 2R_\nu^+(u) - 1$. This implies

$$\partial H(u) = 2[R_\nu^-(u), R_\nu^+(u)] - 1$$

and by (17) finally

$$\partial\big[f(s,\cdot)\big](u) = 2\big[R_\nu^-\big(u(s)\big), R_\nu^+\big(u(s)\big)\big] - 2s. \qquad \square$$

By the following lemma, $J_\varepsilon^{\partial F_\nu}$ fulfills the invariance condition from Theorem 3.5.

Lemma 4.4. *Let F_ν be defined by* (14)*. Then, $J_\varepsilon^{\partial F_\nu}$ maps $\mathcal{C}(0,1)$ into itself for all $\varepsilon > 0$.*

Proof. Let $\varepsilon > 0$ be arbitrarily fixed and $h \in \mathcal{C}(0,1)$. By Theorem 3.3 and Lemma 4.2, we have that $\operatorname{Range}(I + \frac{\varepsilon}{2}\partial F_\nu) = L_2(0,1)$, so we obtain the existence of $u \in L_2(0,1)$ fulfilling $h \in \left(I + \frac{\varepsilon}{2}\partial F_\nu\right) u$. The explicit representation of ∂F_ν in Lemma 4.3 yields

$$u(s) + \varepsilon[R_\nu^-(u(s)), R_\nu^+(u(s))] \ \ni \ h(s) + \varepsilon s \quad \text{for a.e. } s \in (0,1). \tag{18}$$

We have to show that $u \in \mathcal{C}(0,1)$. Since $h + \varepsilon(\cdot) \in \mathcal{C}(0,1)$, there exists a null set $\mathcal{N} \subset (0,1)$ such that outside of $\mathcal{N}$, $h + \varepsilon(\cdot)$ is increasing and (18) holds. Assume that $u \notin \mathcal{C}(0,1)$. Then, there exist $s_1, s_2 \in (0,1)$ outside of $\mathcal{N}$ with $s_1 < s_2$ and $u(s_1) > u(s_2)$. But since $R_\nu^-(u(s_1)) - R_\nu^+(u(s_2)) = \nu((u(s_2), u(s_1)) \geq 0$, it follows that

$$h(s_1) + \varepsilon s_1 \geq u(s_1) + \varepsilon R_\nu^-(u(s_1)) > u(s_2) + \varepsilon R_\nu^+(u(s_2)) \geq h(s_2) + \varepsilon s_2,$$

contradicting that $h + \varepsilon(\cdot)$ is increasing outside of $\mathcal{N}$, and the proof is finished. □

Combining the results from Lemmas 4.2 and 4.4, we can apply Theorem 3.5 to F_ν. By Proposition 3.4, the properties of F_ν carry over to $\mathcal{F}_\nu$, thus Theorem 2.1 applies to $\mathcal{F}_\nu$. Together, we obtain the following theorem.

Theorem 4.5. *Let $\mathcal{F}_\nu$ and F_ν be defined by* (13) *and* (14)*, respectively. Then the Cauchy problem*

$$\begin{cases} \partial_t g(t) \in -\partial F_\nu(g(t)), & t \in (0,\infty), \\ g(0) = Q_{\mu_0}, \end{cases} \tag{19}$$

has a unique strong solution g, and the associated curve $\gamma_t \coloneqq (g(t))_\# \Lambda_{(0,1)}$ is the unique Wasserstein gradient flow of $\mathcal{F}_\nu$.

Finally, let us briefly have a look at the results of the paper [7] (based on [6]) concerning only the interaction energy part of the MMD functional.

Remark 4.6. *The authors of [7] showed a similar result as Theorem 4.5, but only for the interaction energy part $\mathcal{E}(\mu) = \frac{1}{2}\int_{\mathbb{R}^d \times \mathbb{R}^d} K(x,y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$ of the MMD* (12)*. More precisely, they derived two equivalent criteria for a curve $\gamma\colon (0,\infty) \to \mathcal{P}_2(\mathbb{R})$ to be a Wasserstein gradient flow with respect to $\mathcal{E}$. The first characterization is that the distributions functions $R_t \coloneqq R^+_{\gamma(t)}$ solve*

$$\partial_t R_t(x) + \partial_x(R_t^2(x) - R_t(x)) = 0$$

subject to some minimum entropy condition. Second, this can be characterized via quantile functions. That is, γ is a Wasserstein gradient flow with respect to $\mathcal{E}$ if and only if its quantile functions $Q_t \coloneqq Q_{\gamma(t)}$ solve the L_2-subgradient inclusion

$$-\frac{\mathrm{d}}{\mathrm{d}t} Q_t \in \partial E(Q_t), \quad \textit{where} \quad E(g) = \begin{cases} \int_0^1 (1-2s) g(s)\,\mathrm{d}s, & \textit{if } g \in \mathcal{C}(0,1), \\ \infty, & \textit{otherwise}, \end{cases}$$

for almost every $t > 0$. Based on this representation, they can explicitly compute the Wasserstein gradient flow with respect to $\mathcal{E}$ via the quantile function $Q_{\gamma(t)}(s) = Q_{\gamma(0)}(s) + t(2s-1)$ and prove that $\gamma(t)$ is absolutely continuous for all $t > 0$. These results were extended in [12], where the authors consider gradient flows for the functional $\mathrm{MMD}_K^2\colon \mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ and the related PDE system. In contrast to our setting, these gradient flows are defined on $\mathcal{P}_2(\mathbb{R}) \times \mathcal{P}_2(\mathbb{R})$, i.e., they consider two *interacting measures (species), while on our side, the second entry is fixed as the target measure ν. Using again the embedding of the Wasserstein space into $L^2(0,1)$, they prove that gradient flows exist and that they remain absolutely continuous whenever the initial measures are absolutely continuous. Stability of stationary states of singular interaction energies and differentiable potentials on $\mathbb{R}$ with bounded, compactly supported initial data is studied in [21].*

5 Explicit Solution of the Cauchy Problem

Having the subdifferential of F_ν in (16) in mind, we first aim to find a solution of the pointwise Cauchy problem, i.e., for fixed $s \in (0,1)$ we are interested in

$$\begin{cases} \partial_t g_s(t) \in 2s - 2[R_\nu^-(g_s(t)), R_\nu^+(g_s(t))], & \text{for a.e. } \ t > 0, \\ g_s(0) = Q_{\mu_0}(s), \end{cases} \tag{20}$$

satisfying $g_s(t) = [g(t)](s)$. Since the proposed method works in the same way for the left- and right-continuous version of the CDF, we use $R_\nu^\pm$ to indicate one of them.

Lemma 5.1 (Pointwise solution)**.** *Let $\nu, \mu_0 \in \mathcal{P}_2(\mathbb{R})$. For any fixed continuity point $s \in (0,1)$ of Q_ν, the curve*

$$g_s(t) \coloneqq \begin{cases} \Phi_{\nu,s}^{-1}(2t), & t < T_s, \\ Q_\nu(s), & \textit{otherwise}, \end{cases} \tag{21}$$

is a strong solution of (20)*, where $T_s \coloneqq \frac{1}{2}\Phi_{\nu,s}\left(Q_\nu(s)\right)$ and*

$$\Phi_{\nu,s}(x) \coloneqq \int_{Q_{\mu_0}(s)}^{x} \left(s - R_\nu^\pm(z)\right)^{-1} \mathrm{d}z$$

with $x \in [\min\{Q_{\mu_0}(s), Q_\nu(s)\}, \max\{Q_{\mu_0}(s), Q_\nu(s)\}]$. In particular, the inclusion in (20) *holds true for every $t \in (0,\infty)$, where $\dot g_s(t)$ exists.*

Proof. 1. First, assume that $s \in (0,1)$ is *any* point such that $Q_{\mu_0}(s) < Q_\nu(s)$. Since $R^- \le R^+$ and by the Galois inequality (7), we have for any $z < Q_\nu(s)$ that $R_\nu^\pm(z) < s$. Hence $\left(s - R_\nu^\pm(z)\right)^{-1}$ exists and $\Phi_{\nu,s}(x)$ is finite for all $x < Q_\nu(s)$. For $x = Q_\nu(s)$ we distinguish two cases.

Case 1: $\Phi_{\nu,s}(Q_\nu(s)) = \infty$. Fix any $\hat x < Q_\nu(s)$. Then, $\Phi_{\nu,s} : [Q_{\mu_0}(s), \hat x] \to [0,\infty)$ is absolutely continuous, and its derivative exists a.e. and is given by

$$\Phi'_{\nu,s}(x) = \left(s - R_\nu^\pm(x)\right)^{-1} > 0, \quad \text{for a.e. } x \in [Q_{\mu_0}(s), \hat x].$$

By a result of Zareckii [4, 33], the function $\Phi_{\nu,s} : [Q_{\mu_0}(s), \hat x] \to [0, \Phi_{\nu,s}(\hat x)]$ is invertible, and its inverse is also absolutely continuous with derivative

$$(\Phi_{\nu,s}^{-1})'(t) = \frac{1}{\Phi'_{\nu,s}(\Phi_{\nu,s}^{-1}(t))} = s - R_\nu^\pm(\Phi_{\nu,s}^{-1}(t)), \quad \text{for a.e. } t \in [0, \Phi_{\nu,s}(\hat x)].$$

Hence, by definition (21), the function g_s is absolutely continuous on $[0, \frac{1}{2}\Phi_{\nu,s}(\hat x))$, and it holds

$$\dot g_s(t) = 2s - 2R_\nu^\pm(g_s(t)), \quad \text{for a.e. } t \in [0, \frac{1}{2}\Phi_{\nu,s}(\hat x)).$$

Considering the above arguments, we notice that g_s is differentiable at t if and only if $g_s(t)$ is a continuity point of $R_\nu^\pm$, i.e. $R_\nu^+(g_s(t)) = R_\nu^-(g_s(t))$. Since $\Phi_{\nu,s}(\hat{x}) \uparrow \infty$ for $\hat{x} \uparrow Q_\nu(s)$, we see that g_s is a strong solution of (20).

Case 2: $\Phi_{\nu,s}(Q_\nu(s)) < \infty$. Then above arguments with $\hat{x} = Q_\nu(s)$ show that g_s is absolutely continuous on $[0, T_s)$, and it holds

$$\dot{g}_s(t) = 2s - 2R_\nu^\pm(g_s(t)), \quad \text{for a.e. } t \in [0, T_s).$$

By the continuity of $\Phi_{\nu,s}^{-1} : [0, 2T_s] \to [Q_{\mu_0}(s), Q_\nu(s)]$ and by construction of g_s, we conclude that g_s is continuous in T_s. Since $g_s \equiv Q_\nu(s)$ on $[T_s, \infty)$, it is absolutely continuous on the whole half-line $[0, \infty)$. On the one hand, (7) immediately yields $s \le R_\nu^+(Q_\nu(s))$, and on the other hand, its negation $s > R_\nu^+(x) \ge R_\nu^-(x)$ for $x < Q_\nu(s)$ implies $s \ge R_\nu^-(Q_\nu(s))$ due to the left-continuity of R_ν^-. For this reason, we finally have

$$\dot{g}_s(t) = 0 \in 2s - 2[R_\nu^-(Q_\nu(s)), R_\nu^+(Q_\nu(s))] \quad \text{for all } t \in [T_s, \infty),$$

which means that g_s is a strong solution of (20).

2. Next, assume that $s \in (0,1)$ is a *continuity* point of Q_ν such that $Q_{\mu_0}(s) > Q_\nu(s)$. Then Galois inequality (7) together with the fact that s is a continuity point of Q_ν implies for $z > Q_\nu(s)$ that $R_\nu^\pm(z) > s$. Now, the rest follows analogously as in the first part.

3. For the case $Q_{\mu_0}(s) = Q_\nu(s)$, we clearly get the constant solution, and the proof is done. $\square$

Next we show that the curve $g : [0, \infty) \to \mathcal{C}(0,1)$ is differentiable almost everywhere and solves the original L_2 problem (19).

Theorem 5.2 (L_2 solution)**.** *The family*

$$\{g_s \textit{ in } (21) \colon \; s \in (0,1) \;\; \textit{is a continuity point of} \;\; Q_\nu\}$$

strongly solves (19) *via* $[g(t)](s) \coloneqq g_s(t)$.

Proof. Since (21) is a strong solution of (20), it holds for a.e. $t > 0$ that

$$|\dot{g}_s(t)| \le |2s - 2R_\nu^\pm(g_s(t))| \le 4.$$

Since g_s is absolutely continuous in t, this implies that $g_s : [0, \infty) \to \mathbb{R}$ is Lipschitz continuous for a.e. fixed $s \in (0,1)$, and therefore, $g : [0, \infty) \to L_2(0,1)$ is Lipschitz continuous. In particular, g is differentiable at a.e. $t > 0$, i.e., for any differentiability point $\hat{t} > 0$ outside a null set, the sequence of functions

$$\partial_t g(\hat{t}) \coloneqq \lim_{h \to 0} \frac{g(\hat{t} + h) - g(\hat{t})}{h}$$

converges in $L_2(0,1)$. In particular, this holds true for any *positive* zero sequence $(h_n)_{n \in \mathbb{N}}$. Then, there exists a subsequence $(h_{n_k})_k$ and a null set $N_{\hat{t}} \subset (0,1)$ such that

$$[\partial_t g(\hat{t})](s) = \lim_{k \to \infty} \frac{g_s(\hat{t} + h_{n_k}) - g_s(\hat{t})}{h_{n_k}} \quad \text{for all } s \in (0,1) \setminus \mathcal{N}_{\hat{t}}\,.$$

Now, fix $s \in (0,1) \setminus N_{\hat{t}}$ which is also a continuity point of Q_ν. By Theorem A.2, the strong solution g_s of (20) solves

$$\frac{d^+}{dt} g_s(t) \in 2s - 2[R_\nu^-(g_s(t)), R_\nu^+(g_s(t))]$$

	$Q_{\mu_0}(s) \leq Q_\nu(s)$	$Q_{\mu_0}(s) \geq Q_\nu(s)$	
ℓ_s	$W_{\ell_s-1} < s < W_{\ell_s}$	$W_{\ell_s-1} < s < W_{\ell_s}$	
k_s	$x_{k_s} \leq Q_{\mu_0}(s) < x_{k_s+1}$	$x_{k_s-1} < Q_{\mu_0}(s) \leq x_{k_s}$	
$x_{s,j}$	x_{k_s+j}	x_{k_s-j}	$j = 1, \dots, \lvert \ell_s - k_s \rvert$
$R_{s,j}$	W_{k_s+j}	W_{k_s-j-1}	$j = 0, \dots, \lvert \ell_s - k_s \rvert - 1$

Table 1: Quantities from Corollary 5.3, where $W_k \coloneqq \sum_{j=1}^k w_j$ and $x_0 \coloneqq -\infty$, $x_{n+1} \coloneqq \infty$.

even for *every* $t > 0$, where $\frac{d^+}{dt}$ denotes the right derivative. Altogether, for the choice $t = \hat{t}$, it follows

$$\begin{aligned}
[\partial_t g(\hat{t})](s) &= \lim_{k\to\infty} \frac{g_s(\hat{t} + h_{n_k}) - g_s(\hat{t})}{h_{n_k}} \\
&= \frac{d^+}{dt} g_s(\hat{t}) \ \in \ 2s - 2[R_\nu^-(g_s(\hat{t})), R_\nu^+(g_s(\hat{t}))].
\end{aligned}$$

This proves that g is a strong solution of (19), and we are done. □

Next, we apply the explicit solution formula (21) to describe the flow from an arbitrary starting measure $\mu_0 \in \mathcal{P}_2(\mathbb{R})$ to a discrete measure ν.

Corollary 5.3 (Point measure target)**.** *Let*

$$\nu \coloneqq \sum_{j=1}^n w_j \delta_{x_j}$$

with weights $0 < w_j \leq 1$ *fulfilling* $\sum_{j=1}^n w_j = 1$ *and* $x_1 < x_2 < \cdots < x_n$*. Then the strong solution of* (19) *is given by*

$$[g(t)](s) \coloneqq \begin{cases} Q_{\mu_0}(s) + 2\,(s - R_{s,0})\,t, & t \in [t_{s,0}, t_{s,1}), \\ x_{s,j} + 2\,(s - R_{s,j})\,(t - t_{s,j}), & t \in [t_{s,j}, t_{s,j+1}), \\ Q_\nu(s), & t \geq t_{s,|\ell_s - k_s|}, \end{cases}$$

where

$$t_{s,0} \coloneqq 0, \quad t_{s,1} \coloneqq \tfrac{x_{s,1} - Q_{\mu_0}(s)}{2(s - R_{s,0})}, \quad t_{s,j+1} \coloneqq t_{s,j} + \frac{x_{s,j+1} - x_{s,j}}{2(s - R_{s,j})},$$

for $j \in \{1, \dots, |\ell_s - k_s| - 1\}$*, and the quantities* $\ell_s, k_s, x_{s,j}, R_{s,j}$ *are given in Table 1.*

Proof. Recall that g_s is either monotonically increasing or monotonically decreasing until it hits the target quantile function $Q_\nu(s)$. The values $R_{s,j}$ in Table 1 denote the values of the constant plateaus, which the solution passes. The values $x_{s,j}$ denote the locations where the derivative jumps. Solving (20) piecewise gives the stated explicit form. □

In Figure 3 we illustrate that from a pointwise view, the flow g_s changes only linearly over time. However, since all quantities in Table 1, especially the time points $t_{s,j}$ corresponding to the discontinuities of the derivative, depend non-linearly on $s \in (0, 1)$, the actual evolution of the quantile function itself, becomes highly non-linear. This is illustrated in the following example.

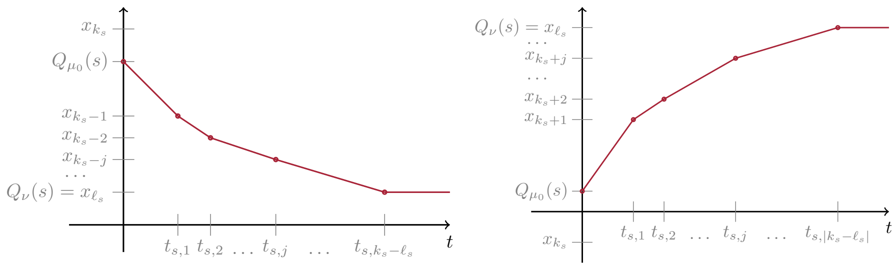

Figure 3: The map $[0,\infty) \to \mathbb{R}$, $t \mapsto [g(t)](s)$ for a $s \in (0,1)$ with $Q_\mu(s) < Q_\nu(s)$ **(left)** and $Q_\mu(s) > Q_\nu(s)$ **(right)**. The slopes of the affine linear pieces, determined by $s - R_{s,j}$, decrease for increasing t, that is, for increasing j.

Example 5.4. *Let*

$$\mu_0 \coloneqq \tfrac{1}{3}\delta_{-1} + \tfrac{1}{3}\delta_{1/2} + \tfrac{1}{3}\delta_2 \quad \textit{and} \quad \nu \coloneqq \tfrac{1}{4}\delta_0 + \tfrac{3}{4}\delta_1.$$

The evolution of the quantile function can be computed as

$$[g(t)](s) = \begin{cases} \min\{2st - 1, 0\}, & s \in (0, \frac{1}{4}),\\ (22), & s \in (\frac{1}{4}, \frac{1}{3}),\\ \min\{(\frac{1}{2} - \frac{t}{4}) + 2st, 1\}, & s \in (\frac{1}{3}, \frac{2}{3}),\\ \max\{(2 - 2t) - 2st, 1\} & s \in (\frac{2}{3}, 1), \end{cases}$$

where

$$[g(t)]\big|_{(\frac{1}{4},\frac{1}{3})}(s) = \begin{cases} 2st - 1, & s \le t_1^{-1}(t),\\ \frac{1}{4s} - (\frac{t}{2} + 1) + 2st, & t_1^{-1}(t) \le s \le t_2^{-1}(t),\\ 1, & t_2^{-1}(t) \le s \end{cases} \tag{22}$$

with the monotonically decreasing, continuous functions $t_1(s) \coloneqq t_{s,1}$ and $t_2(s) \coloneqq t_{s,2}$. Except for $s \in (1/4, 1/3)$, the quantile function is piecewise linear, which corresponds to uniform measures and point measures. For $s \in (1/4, 1/3)$, it becomes non-linear. The flow $g(t)_\# \Lambda_{(0,1)}$ is illustrated in Figure 4.

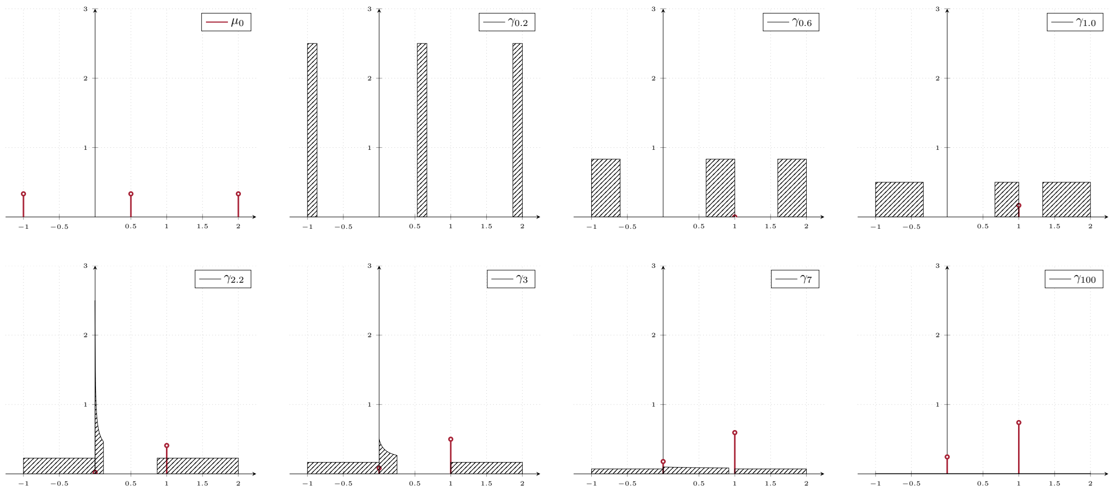

Figure 4: The Wasserstein gradient flow between the point measures in Example 5.4. The gray regions illustrate the density of the absolute continuous part of the flow, whereas the red spikes illustrate the discrete part. In the lower row, the mass from δ_{-1} split up, and a portion moves towards δ_1. The movement speed of this portion, however, significantly decreases due to its attraction by δ_0. Figuratively, the mass sticks to δ_0 and can escape only slowly. For the corresponding densities see Figure 14.

6 Invariant Subsets and Smoothing Properties of F_ν-Flows

In this section, we are interested in *invariant subsets of* $\mathcal{C}(0,1)$, i.e., subsets in which F_ν-flows remain once starting there. We also prove a *smoothing* result, where the F_ν-flow immediately becomes more regular than the initial datum.

Note that [9, p. 131, Prop. 4.5] generally characterizes conditions for closed subsets to be invariant. However, we take a more refined approach involving the resolvent $J_\varepsilon^{\partial F_\nu}$, the exponential formula (10), and our explicit calculation (16) of ∂F_ν. This approach will yield more precise results – and is not limited to closed subsets.

As a starting point, recall that $J_\varepsilon^{\partial F_\nu}$ maps $\mathcal{C}(0,1)$ into itself by Lemma 4.4. By (16) this means: For all $\varepsilon > 0$ and any $h \in \mathcal{C}(0,1)$, there exists $u \in \mathcal{C}(0,1)$ such that

$$u(s) + \varepsilon[R_\nu^-(u(s)), R_\nu^+(u(s))] \;\ni\; h(s) + \varepsilon s \quad \text{for a.e. } s \in (0,1). \tag{23}$$

To simplify the following arguments and notations, we identify an (equivalence class) $v \in \mathcal{C}(0,1)$ with its unique left-continuous and increasing version, i.e. $v := Q_{\mu_v}$, where $\mu_v := v_\# \Lambda_{(0,1)}$, such that v is uniquely defined *everywhere* on $(0,1)$ (and not only up to a null set) [1].

With the above identification of $v \in \mathcal{C}(0,1)$ with its quantile function, the following lemma shows that (23) holds *everywhere* on $(0,1)$.

[1] Further, we make the agreement to always exclude denominators of (difference) quotients being zero, implicitly.

Lemma 6.1. *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ be arbitrary. Then, for any $h \in \mathcal{C}(0,1)$ there exists $u := J^{\partial F_\nu}_{\varepsilon/2} h \in \mathcal{C}(0,1)$ such that*

$$u(s) + \varepsilon[R^-_\nu(u(s)), R^+_\nu(u(s))] \ \ni\ h(s) + \varepsilon s \quad \textit{for all} \quad s \in (0,1). \tag{24}$$

The proof is given in the appendix.

Now, let $-\infty \le a \le b \le \infty$ and $L > 0$. We will study the following closed subsets of $\mathcal{C}(0,1)$:

I) bounded quantile functions:

$$D_{[a,b]} := \{Q_\mu \in \mathcal{C}(0,1) : \overline{Q_\mu(0,1)} \subseteq [a,b]\} = \{Q_\mu \in \mathcal{C}(0,1) : \operatorname{supp}\mu \subseteq [a,b]\},$$

II) quantile functions admitting a lower L-Lipschitz condition:

$$D^-_L := \{Q_\mu \in \mathcal{C}(0,1) : L \le \frac{Q_\mu(s_1) - Q_\mu(s_2)}{s_1 - s_2} \ \text{ for all } s_1, s_2 \in (0,1)\},$$

III) L-Lipschitz continuous quantile functions:

$$D^+_L := \{Q_\mu \in \mathcal{C}(0,1) : \frac{Q_\mu(s_1) - Q_\mu(s_2)}{s_1 - s_2} \le L \ \text{ for all } s_1, s_2 \in (0,1)\}.$$

We will also consider the following subset which is not closed in $\mathcal{C}(0,1)$:

IV) continuous quantile functions:

$$\begin{aligned} D_c &:= \{Q_\mu \in \mathcal{C}(0,1) : Q_\mu \text{ is continuous}\} \\ &= \{Q_\mu \in \mathcal{C}(0,1) : R^+_\mu \text{ is strictly increasing on } \overline{(R^+_\mu)^{-1}(0,1)}\} \\ &= \{Q_\mu \in \mathcal{C}(0,1) : \operatorname{supp}\mu \text{ is convex}\}. \end{aligned}$$

Note that the subset of quantile functions of absolutely continuous measures with respect to the Lebesgue measure is also not closed in $\mathcal{C}(0,1)$.

Next, we provide some examples of probability measures $\nu \in \mathcal{P}_2(\mathbb{R})$ and discuss to which of the sets from above their quantile functions belong.

Example 6.2. *i) For any atomic measure ν with a finite number of atoms we have $Q_\nu \in D_{[a,b]}$ for some $a, b \in \mathbb{R}$ and $Q_\nu \notin D^-_L$ for any $L > 0$.*

ii) If ν defines a uniform distribution on $[a,b]$, then its quantile function is $s \mapsto a + s(b-a)$. Hence $Q_\nu \in D^+_{b-a} \cap D^-_{b-a} \cap D_{[a,b]} \cap D_c$.

iii) For the normal distribution $\nu \sim \mathcal{N}(\mu, \sigma^2)$, we have $\operatorname{supp}\nu = \mathbb{R}$, so $Q_\nu \in D_c$, and that $Q_\nu(s) = \mu + \sqrt{2}\sigma \operatorname{erf}^{-1}(2s-1)$ is not Lipschitz continuous. However, $Q_\nu \in D^-_L$ with $L := (2\pi\sigma^2)^{\frac{1}{2}}$. The same is true for the Laplace distribution or a mixture of two normal distributions.

iv) The exponential, Pareto and folded norm distributions, see (33), have a support that is unbounded in only one direction, so their quantile functions belong to $D_c \cap D_{[a,\infty)}$ for some $a \in \mathbb{R}$. As above, these quantile functions do not belong to D^+_L, but to D^-_L.

Note that if $\mu \in \mathcal{P}_2(\mathbb{R})$ is absolutely continuous and $Q_\mu \notin D_{[a,b]}$ for any $a, b \in \mathbb{R}$, then we can not have $Q_\mu \in D^+_L$ for any $L > 0$.

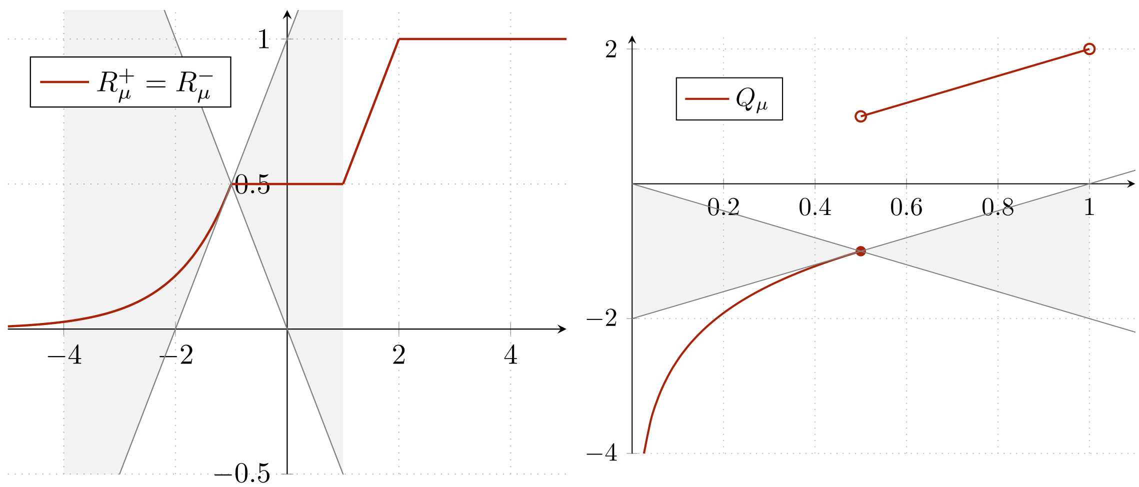

Figure 5: **Left:** R^+_μ is $\frac{1}{2}$-Lipschitz continuous, since there is a "double cone" (white) with extremal rays of slope $\frac{1}{2}$ such that if it is centered at any point of the graph of R^+_μ, then the graph has no points within this double cone. **Right:** Q_μ admits a lower 2-Lipschitz condition, since the white double cone with extremal rays of slope 2 contains the whole graph of Q_μ.

By the following Proposition 6.3, the above sets $D^\pm_L$ can be described in terms of CDFs instead of quantile functions, see Figure 5.

Proposition 6.3. *Let $Q_\mu \in \mathcal{C}(0,1)$. Then the following holds true:*

i) $Q_\mu \in D^-_L$ if and only if $R_\mu = R^+_\mu$ is $\frac{1}{L}$-Lipschitz continuous, i.e.,

$$\frac{R_\mu(x) - R_\mu(y)}{x - y} \leq \frac{1}{L} \quad \textit{for all } x, y \in \mathbb{R}\,.$$

In this case, μ is absolutely continuous, and the above holds iff its density f_μ fulfills

$$f_\mu \leq \frac{1}{L} \quad \textit{a.e. on } \mathbb{R}\,.$$

ii) $Q_\mu \in D^+_L$ if and only if R^+_μ admits a lower $\frac{1}{L}$-Lipschitz condition on $\overline{(R^+_\mu)^{-1}(0,1)}$, i.e.,

$$\frac{1}{L} \leq \frac{R^+_\mu(x) - R^+_\mu(y)}{x - y} \quad \textit{for all } x, y \in \overline{(R^+_\mu)^{-1}(0,1)}.$$

If μ is absolutely continuous, then the above holds iff its density f_μ fulfills

$$\frac{1}{L} \leq f_\mu \quad \textit{a.e. on } \operatorname{conv}(\operatorname{supp}\mu).$$

The proof is given in the appendix.

The following proposition says that the support of flows cannot escape the convex hull of the support of the target measure ν once they start there.

Proposition 6.4 (Invariance of $D_{[a,b]}$)**.** *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ be such that $\overline{\operatorname{conv}}(\operatorname{supp}\nu) \subseteq [a,b]$. Then, $J_\varepsilon^{\partial F_\nu}$ maps $D_{[a,b]}$ into itself for all $\varepsilon > 0$. Hence, the solution of the Cauchy problem* (19) *starting in $g_0 \in D_{[a,b]}$ fulfills $g(t) \in D_{[a,b]}$ for all $t \geq 0$.*

Proof. Let $h \in D_{[a,b]}$ and consider a solution $u \in \mathcal{C}(0,1)$ of (24). W.l.o.g., let a, b be finite (otherwise there is nothing to show). Assume there exists $s^* \in (0,1)$ such that

$$u(s^*) < a \leq \min(\operatorname{supp}\nu).$$

Then, it holds $R_\nu^-(u(s^*)) = R_\nu^+(u(s^*)) = 0$ and by (24) that

$$a > u(s^*) = h(s^*) + \varepsilon s^* \geq a + \varepsilon s^*,$$

which is a contradiction. Now, assume there is $s^* \in (0,1)$ such that

$$u(s^*) > b \geq \max(\operatorname{supp}\nu).$$

Then, we obtain $R_\nu^-(u(s^*)) = R_\nu^+(u(s^*)) = 1$ and by (24) further

$$b < u(s^*) = h(s^*) + \varepsilon(s^* - 1) \leq b + \varepsilon(s^* - 1),$$

again a contradiction. Together, we have proved $u(s) \in [a,b]$ for *every* $s \in (0,1)$, which shows $u \in D_{[a,b]}$. The remaining claim follows by Corollary 3.6. □

In order to prove invariance results for D_L^- and D_L^+, we define[2] for given $Q_\mu \in \mathcal{C}(0,1)$ the largest lower Lipschitz constant

$$L_{\text{low}}(Q_\mu) := \max\{L \geq 0 : \frac{Q_\mu(s_1) - Q_\mu(s_2)}{s_1 - s_2} \geq L \ \text{ for all } s_1, s_2 \in (0,1)\} \geq 0,$$

where $L_{\text{low}}(Q_\mu) = 0$ is explicitly allowed. Note that by Proposition 6.3, it holds $L_{\text{low}}(Q_\mu) > 0$ if and only if R_μ is Lipschitz continuous with constant $L_{\text{low}}(Q_\mu)^{-1}$. Further, consider the usual smallest Lipschitz constant

$$\operatorname{Lip}(Q_\mu) := \min\{L \geq 0 : \frac{Q_\mu(s_1) - Q_\mu(s_2)}{s_1 - s_2} \leq L \ \text{ for all } s_1, s_2 \in (0,1)\} \leq \infty.$$

By Proposition 6.3, it holds $0 < \operatorname{Lip}(Q_\mu) < \infty$ if and only if R_μ^+ admits a lower $\operatorname{Lip}(Q_\mu)^{-1}$-Lipschitz condition on $\overline{(R_\mu^+)^{-1}(0,1)}$.

Now, we can formulate an invariance and smoothing property of F_ν-flows. In addition, we can accurately describe how the Lipschitz constants of the F_ν-flow evolve over time. We start with the lower constant.

Theorem 6.5. *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ with $L_{\text{low}}(Q_\nu) > 0$. Consider any initial value $g_0 = Q_{\mu_0} \in \mathcal{C}(0,1)$, where $L_{\text{low}}(g_0) = 0$ is explicitly allowed. Then, the strong solution $g : [0,\infty) \to \mathcal{C}(0,1)$ of the Cauchy problem* (19) *enjoys for all $t > 0$ the smoothing property*

$$L_{\text{low}}(g(t)) \geq L_{\text{low}}(g_0) \cdot e^{-\frac{2t}{L_{\text{low}}(Q_\nu)}} + L_{\text{low}}(Q_\nu) \cdot (1 - e^{-\frac{2t}{L_{\text{low}}(Q_\nu)}}) > 0. \tag{25}$$

[2]Notice that in the definitions of $L_{\text{low}}(Q_\mu)$ and $\operatorname{Lip}(Q_\mu)$, *"for all"* can be replaced by *"for almost all"*, yielding the same constants since Q_μ is left-continuous.

In particular, the CDF R_{γ_t} of the associated Wasserstein gradient flow γ_t is Lipschitz continuous for any $t > 0$ with Lipschitz constant

$$\mathrm{Lip}(R_{\gamma_t}) \le \Big(L_{\mathrm{low}}(g_0) \cdot e^{-\frac{2t}{L_{\mathrm{low}}(Q_\nu)}} + L_{\mathrm{low}}(Q_\nu) \cdot (1 - e^{-\frac{2t}{L_{\mathrm{low}}(Q_\nu)}})\Big)^{-1} < \infty. \tag{26}$$

Proof. i) First, let $h \in \mathcal{C}(0,1)$ and $\varepsilon > 0$. By (24) and since $R_\nu^+ = R_\nu^- = R_\nu$ is continuous, there exists a solution $u \in \mathcal{C}(0,1)$ of

$$h(s) + \varepsilon s = u(s) + \varepsilon R_\nu(u(s)) \quad \text{for all } s \in (0,1).$$

By the Lipschitz continuity of R_ν, it holds for all $s_1, s_2 \in (0,1)$, $s_1 > s_2$, that

$$\begin{aligned}\frac{u(s_1) - u(s_2)}{s_1 - s_2} &= \frac{h(s_1) - h(s_2)}{s_1 - s_2} + \varepsilon - \varepsilon \frac{R_\nu(u(s_1)) - R_\nu(u(s_2))}{s_1 - s_2} \\ &\ge L_{\mathrm{low}}(h) + \varepsilon - \varepsilon \frac{1}{L_{\mathrm{low}}(Q_\nu)} \cdot \frac{u(s_1) - u(s_2)}{s_1 - s_2},\end{aligned}$$

which yields

$$\frac{u(s_1) - u(s_2)}{s_1 - s_2} \ge \frac{L_{\mathrm{low}}(h) + \varepsilon}{q}, \quad q \coloneqq 1 + \frac{\varepsilon}{L_{\mathrm{low}}(Q_\nu)}.$$

This means $L_{\mathrm{low}}(u) \ge \frac{L_{\mathrm{low}}(h)+\varepsilon}{q} > 0$. (Notice the improvement $L_{\mathrm{low}}(u) > 0$, even if $L_{\mathrm{low}}(h) = 0$.) We have proved for any $h \in \mathcal{C}(0,1)$ and any $\varepsilon > 0$ that

$$L_{\mathrm{low}}((I + \frac{\varepsilon}{2}\partial F_\nu)^{-1} h) \ge \frac{L_{\mathrm{low}}(h) + \varepsilon}{q}.$$

Now, fix $n \in \mathbb{N}$ and observe inductively that

$$\begin{aligned}L_{\mathrm{low}}((I + \frac{\varepsilon}{2}\partial F_\nu)^{-n} g_0) &\ge \frac{L_{\mathrm{low}}((I + \frac{\varepsilon}{2}\partial F_\nu)^{-(n-1)} g_0)}{q} + \frac{\varepsilon}{q} \\ &\ge \frac{L_{\mathrm{low}}((I + \frac{\varepsilon}{2}\partial F_\nu)^{-(n-2)} g_0)}{q^2} + \frac{\varepsilon}{q^2} + \frac{\varepsilon}{q} \ge \dots \\ &\ge \frac{L_{\mathrm{low}}(g_0)}{q^n} + \varepsilon \sum_{k=1}^{n} \frac{1}{q^k} \\ &= \frac{L_{\mathrm{low}}(g_0)}{q^n} + \frac{\varepsilon}{q} \cdot \frac{1 - q^{-n}}{1 - q^{-1}} \\ &= \frac{L_{\mathrm{low}}(g_0)}{q^n} + L_{\mathrm{low}}(Q_\nu)(1 - q^{-n}).\end{aligned}$$

Now, for $t > 0$ and choosing $\varepsilon \coloneqq \frac{2t}{n}$, it follows

$$\begin{aligned}L_{\mathrm{low}}((I + \frac{t}{n}\partial F_\nu)^{-n} g_0) &\ge \frac{L_{\mathrm{low}}(g_0)}{(1 + \frac{2t}{nL_{\mathrm{low}}(Q_\nu)})^n} + L_{\mathrm{low}}(Q_\nu)(1 - (1 + \frac{2t}{nL_{\mathrm{low}}(Q_\nu)})^{-n}) \\ &\xrightarrow{n\uparrow\infty} L_{\mathrm{low}}(g_0) \cdot e^{-\frac{2t}{L_{\mathrm{low}}(Q_\nu)}} + L_{\mathrm{low}}(Q_\nu) \cdot (1 - e^{-\frac{2t}{L_{\mathrm{low}}(Q_\nu)}}).\end{aligned}$$

Finally, the solution $g : [0,\infty) \to \mathcal{C}(0,1)$ of the Cauchy problem (19) is given by the exponential formula (and L_2-limit) (10) which implies

$$L_{\text{low}}(g(t)) \geq L_{\text{low}}(g_0) \cdot e^{-\frac{2t}{L_{\text{low}}(Q_\nu)}} + L_{\text{low}}(Q_\nu) \cdot (1 - e^{-\frac{2t}{L_{\text{low}}(Q_\nu)}}) > 0$$

for all $t > 0$, which concludes the proof. $\square$

Note that even if the initial measure satisfies $L_{\text{low}}(g_0) = 0$ (e.g., atomic measures), the Lipschitz continuity of the target CDF, i.e., $L_{\text{low}}(Q_\nu) > 0$, is enough to force the flow's CDFs to immediately become Lipschitz continuous for any $t > 0$, and the Lipschitz constants exponentially improve for $t \to \infty$ towards the target Lipschitz constant $L_{\text{low}}(Q_\nu)^{-1}$ as described in (26).

The following example demonstrates that the bound (26) of the Lipschitz constants in Theorem 6.5 is sharp.

Example 6.6. *Consider as target measure ν the uniform distribution on $[0,1]$, and as initial measure $\mu_0 \coloneqq \delta_0$ the Dirac measure at 0. Then, for the corresponding CDFs, we get that*

$$R_\nu(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x < 1, \\ 1, & x \geq 1 \end{cases}$$

is 1-Lipschitz continuous, whereas

$$R^+_{\mu_0}(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

is a jump function. Still, the $\mathcal{F}_\nu$-Wasserstein gradient flow γ_t immediately regularizes to uniform distributions for $t > 0$ given by

$$R_{\gamma_t}(x) = \begin{cases} 0, & x \leq 0, \\ (1 - e^{-2t})^{-1} x, & 0 < x < 1 - e^{-2t}, \\ 1, & x \geq 1 - e^{-2t}, \end{cases}$$

which are $(1 - e^{-2t})^{-1}$-Lipschitz continuous. Thus, the upper bound of the Lipschitz constants (26) is sharp.

Next, we deal with the upper Lipschitz constant.

Theorem 6.7. *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ such that $\operatorname{supp} \nu = [a,b]$ and $\operatorname{Lip}(Q_\nu) < \infty$. Consider an initial value $g_0 = Q_{\mu_0} \in \mathcal{C}(0,1)$ with convex support $\operatorname{supp} \mu_0 \subseteq [a,b]$ and $\operatorname{Lip}(g_0) < \infty$. Then, the strong solution $g : [0,\infty) \to \mathcal{C}(0,1)$ of the Cauchy problem (19) fulfills for all $t > 0$ the invariance property*

$$\operatorname{Lip}(g(t)) \leq \operatorname{Lip}(g_0) \cdot e^{-\frac{2t}{\operatorname{Lip}(Q_\nu)}} + \operatorname{Lip}(Q_\nu) \cdot (1 - e^{-\frac{2t}{\operatorname{Lip}(Q_\nu)}}) < \infty. \tag{27}$$

In particular, the CDF $R^+_{\gamma_t}$ of the associated Wasserstein gradient flow γ_t admits a lower Lipschitz constant on $\overline{(R^+_{\gamma_t})^{-1}(0,1)}$ for all $t > 0$, and the constant is

$$L_{\text{low}}(R^+_{\gamma_t}) \geq \Big(\operatorname{Lip}(g_0) \cdot e^{-\frac{2t}{\operatorname{Lip}(Q_\nu)}} + \operatorname{Lip}(Q_\nu) \cdot (1 - e^{-\frac{2t}{\operatorname{Lip}(Q_\nu)}})\Big)^{-1} > 0.$$

Proof. If $\mathrm{Lip}(Q_\nu) = 0$, then by assumption, it holds $\operatorname{supp}\mu_0 \subseteq \operatorname{supp}\nu = \{a\}$. Hence, the solution of (19) is given by $g(t) \equiv a$ for all $t \geq 0$, and trivially satisfies (27). So, we can assume that $\mathrm{Lip}(Q_\nu) > 0$. Now, let $h \in D_{[a,b]}$ with $\mathrm{Lip}(h) < \infty$, and $\varepsilon > 0$. By Proposition 6.4, there exists a solution $u \in D_{[a,b]}$ of (24). Let $s_1, s_2 \in (0,1)$, $s_1 > s_2$. W.l.o.g., we can assume that $u(s_2) < u(s_1)$. Hence, we can choose $\kappa > 0$ small enough such that $a \leq u(s_2) < u(s_1) - \kappa \leq b$. Now, by (24), it holds

$$u(s_1) \leq h(s_1) + \varepsilon s_1 - \varepsilon R_\nu^-(u(s_1))$$

and

$$u(s_2) \geq h(s_2) + \varepsilon s_2 - \varepsilon R_\nu^+(u(s_2)).$$

By Proposition 6.3, R_ν^+ admits a lower $\mathrm{Lip}(Q_\nu)^{-1}$-Lipschitz condition on $\overline{(R_\nu^+)^{-1}(0,1)} = \operatorname{supp}\nu = [a,b]$, so that

$$\begin{aligned}\frac{u(s_1)-u(s_2)}{s_1-s_2} &\leq \frac{h(s_1)-h(s_2)}{s_1-s_2} + \varepsilon - \varepsilon\frac{R_\nu^-(u(s_1)) - R_\nu^+(u(s_2))}{s_1-s_2}\\ &\leq \frac{h(s_1)-h(s_2)}{s_1-s_2} + \varepsilon - \varepsilon\frac{R_\nu^+(u(s_1)-\kappa) - R_\nu^+(u(s_2))}{s_1-s_2}\\ &\leq \frac{h(s_1)-h(s_2)}{s_1-s_2} + \varepsilon - \varepsilon\frac{1}{\mathrm{Lip}(Q_\nu)}\cdot\frac{(u(s_1)-\kappa)-u(s_2)}{s_1-s_2}.\end{aligned}$$

Letting $\kappa \downarrow 0$ leads to

$$\frac{u(s_1)-u(s_2)}{s_1-s_2} \leq \mathrm{Lip}(h) + \varepsilon - \frac{\varepsilon}{\mathrm{Lip}(Q_\nu)}\cdot\frac{u(s_1)-u(s_2)}{s_1-s_2},$$

which yields

$$\frac{u(s_1)-u(s_2)}{s_1-s_2} \leq \frac{\mathrm{Lip}(h)+\varepsilon}{q}, \quad q := 1 + \frac{\varepsilon}{\mathrm{Lip}(Q_\nu)}.$$

This shows $\mathrm{Lip}(u) \leq \frac{\mathrm{Lip}(h)+\varepsilon}{q} < \infty$. In other words, for any $h \in D_{[a,b]}$ with $\mathrm{Lip}(h) < \infty$ and $\varepsilon > 0$, it holds

$$\mathrm{Lip}((I + \frac{\varepsilon}{2}\partial F_\nu)^{-1}h) \leq \frac{\mathrm{Lip}(h)+\varepsilon}{q}.$$

Now, setting $h := g_0$, we obtain the desired estimate (27) in the same lines as in the previous proof, and we are done. ☐

We mention that formally, the conditions $\mathrm{Lip}(Q_\nu) < \infty$ and $\mathrm{Lip}(Q_{\mu_0}) = \infty$ bring no improvement in the estimate (27) for $\mathrm{Lip}(g(t))$. Further, by the following example, the support assumption in Theorem 6.7 cannot be skipped.

Example 6.8. *In our example in the introduction, we considered the Wasserstein gradient flow from $\mu_0 = \delta_{-1}$ to $\nu = \delta_0$ with $\mathrm{Lip}(Q_{\mu_0}) = \mathrm{Lip}(Q_\nu) = 0$. By Corollary 5.3, the quantile functions given by*

$$[g(t)](s) = \min\{2st - 1, 0\}. \tag{28}$$

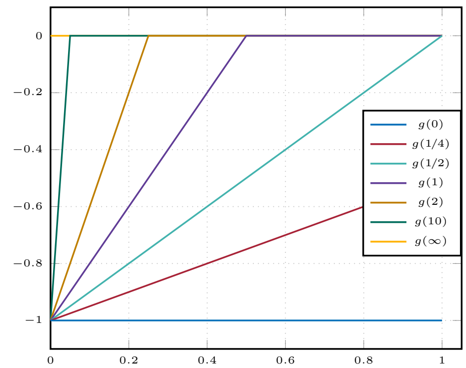

Figure 6: The functions $g(t)$ from (28) for $t \in \{0, \frac{1}{4}, \frac{1}{2}, 1, 2, 10\}$.

are piecewise linear and sharply contained in D^+_{2t}, see Figure 6. In particular, the Lipschitz constants cannot be bounded. The corresponding Wasserstein gradient flow reads as

$$\gamma_t = \begin{cases} \delta_{-1}, & t = 0, \\ \frac{1}{2t}\Lambda_{[-1,-1+2t]}, & 0 \le t \le \frac{1}{2}, \\ \frac{1}{2t}\Lambda_{[-1,0]} + \left(1 - \frac{1}{2t}\right)\delta_0, & \frac{1}{2} < t, \end{cases}$$

and was already depicted in Figure 1.

As a consequence of Theorems 6.5, 6.7 and their proofs, we immediately obtain the following invariance properties of D^-_L and $D_{[a,b]} \cap D^+_L$ with respect to the F_ν-flow.

Corollary 6.9 (Invariance of D^-_L and $D^+_L \cap D_{[a,b]}$)**.** *Let $\nu \in \mathcal{P}_2(\mathbb{R})$.*

i) Assume $Q_\nu \in D^-_L$. Then, the solution of the Cauchy problem (19) *starting in $g_0 \in D^-_L$ fulfills $g(t) \in D^-_L$ for all $t \ge 0$. More precisely, it holds* (25)*. Moreover, $J^{\partial F_\nu}_\varepsilon$ maps D^-_L into itself for all $\varepsilon > 0$.*

In particular, if the initial measure μ_0 has a L_0-Lipschitz continuous CDF R_{μ_0}, then the flow's CDF R_{γ_t} remains Lipschitz (and thus absolutely) continuous with constant $\le \max\{L_0, L^{-1}\}$ for all $t \ge 0$.

ii) Assume $\operatorname{supp}\nu = [a, b]$ and $Q_\nu \in D^+_L$. Then, the solution of the Cauchy problem (19) *starting in $g_0 \in D_{[a,b]} \cap D^+_L$ fulfills $g(t) \in D_{[a,b]} \cap D^+_L$ for all $t \ge 0$. More precisely, it holds* (27)*. Moreover, $J^{\partial F_\nu}_\varepsilon$ maps $D_{[a,b]} \cap D^+_L$ into itself for all $\varepsilon > 0$.*

In particular, if the initial measure μ_0 has convex support $\operatorname{supp}\mu_0 \subseteq [a, b]$ and is in D^+_L, then γ_t has convex support for all $t \ge 0$ and no 'gaps' of empty mass can form[3]*.*

Note that in part i), no point masses can arise during the flow – given that the target measure is Lipschitz continuous – as opposed to the example of Figure 1.

Finally, we study the set D_c of continuous quantile functions. By the following lemma, it is also invariant with respect to the resolvent $J^{\partial F_\nu}_\varepsilon$.

[3]This particular invariance property of F_ν-flows concerning the convexity of supports, we will generalize to arbitrary target measures $\nu \in \mathcal{P}_2(\mathbb{R})$ further below.

Lemma 6.10. *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ be arbitrary. Then, $J_\varepsilon^{\partial F_\nu}$ maps D_c into itself for all $\varepsilon > 0$.*

Proof. For $h \in D_c$, we consider a solution $u \in \mathcal{C}(0,1)$ of (24). Suppose that u is not continuous at some $s_0 \in (0,1)$. Since u is left-continuous and increasing, we have $\lim_{t \downarrow s_0} u(t) > u(s_0)$. But then, (24) and the continuity of h imply

$$\begin{aligned} h(s_0) + \varepsilon s_0 = \lim_{t \downarrow s_0} (h(t) + \varepsilon t) &\geq \lim_{t \downarrow s_0} \big(u(t) + \varepsilon R_\nu^-(u(t))\big) \\ &> u(s_0) + \varepsilon R_\nu^+(u(s_0)) \geq h(s_0) + \varepsilon s_0, \end{aligned}$$

a contradiction. Thus, $u \in D_c$. □

Since D_c is not closed with respect to the L_2-norm, Corollary 3.6 cannot be applied directly. Still, we have the following invariance and monotonicity result. Note that there are no restrictions on the target measure $\nu \in \mathcal{P}_2(\mathbb{R})$.

Theorem 6.11 (Invariance of D_c & monotonicity of the support)**.** *Let $\nu \in \mathcal{P}_2(\mathbb{R})$ and $g_0 \in D_c$. Then the following holds true:*

i) The solution g of the Cauchy problem (19) starting in $g_0 \in D_c$ fulfills $g(t) \in D_c$ for all $t \geq 0$.

ii) The ranges fulfill $\overline{g(t_1)(0,1)} \subseteq \overline{g(t_2)(0,1)}$ for all $0 \leq t_1 \leq t_2$.

In other words the theorem says: if the initial measure μ_0 has convex support, then $\operatorname{supp} \gamma_t$ stays convex for all $t \geq 0$, and we have the monotonicity $\operatorname{supp} \gamma_{t_1} \subseteq \operatorname{supp} \gamma_{t_2}$ for all $0 \leq t_1 \leq t_2$.

Proof. i) Fix $t > 0$. For the initial datum $g_0 \in D_c$, Lemma 6.10 yields

$$g_n \coloneqq (I + \frac{t}{n} \partial F_\nu)^{-n} g_0 \in D_c \quad \text{for all } n \in \mathbb{N}.$$

We verify that $(g_n)_{n \in \mathbb{N}}$ fulfills the assumptions of the Arzelà-Ascoli theorem.
1) Fix $s \in (0,1)$ and $n \in \mathbb{N}$. By (24), it holds for $u \coloneqq (I + \frac{t}{n} \partial F_\nu)^{-1} g_0$ that

$$\begin{aligned} u(s) &\leq g_0(s) + \frac{2t}{n} s - \frac{2t}{n} R_\nu^-(u(s)) \leq g_0(s) + \frac{2t}{n} s, \\ u(s) &\geq g_0(s) + \frac{2t}{n} s - \frac{2t}{n} R_\nu^+(u(s)) \geq g_0(s) + \frac{2t}{n} s - \frac{2t}{n}, \end{aligned}$$

so that

$$|u(s)| \leq |g_0(s)| + \frac{2t}{n}(s+1).$$

Iteratively, it follows

$$|g_n(s)| \leq |g_0(s)| + 2t(s+1).$$

Since $n \in \mathbb{N}$ was arbitrary, $(g_n)_n$ is pointwise bounded.
2) Fix $0 < s_2 < s_1 < 1$ and $n \in \mathbb{N}$. W.l.o.g. we can assume for u as above that $u(s_2) < u(s_1)$. Since $R_\nu^-(u(s_1)) - R_\nu^+(u(s_2)) \geq 0$, it follows by (24) that

$$\begin{aligned} |u(s_1) - u(s_2)| &\leq g_0(s_1) - g_0(s_2) + \frac{2t}{n}(s_1 - s_2) - \frac{2t}{n}\big(R_\nu^-(u(s_1)) - R_\nu^+(u(s_2))\big) \\ &\leq |g_0(s_1) - g_0(s_2)| + \frac{2t}{n}|s_1 - s_2|. \end{aligned}$$

Again, it follows inductively that

$$|g_n(s_1) - g_n(s_2)| \leq |g_0(s_1) - g_0(s_2)| + 2t|s_1 - s_2|.$$

Since $n \in \mathbb{N}$ was arbitrary, $(g_n)_{n\in\mathbb{N}}$ is equicontinuous.

Now let $K \subset (0,1)$ be compact. By Arzelà-Ascoli's theorem, (for a subsequence) (g_n) converges uniformly on K to a continuous function $\tilde{g}\colon K \to \mathbb{R}$. By the exponential formula (10), the L_2-limit of (g_n) is already given by $g(t)$. By the uniqueness of the limit, it follows that $g(t)|_K \equiv \tilde{g}$ is continuous on K. Since K was an arbitrary compact subset of $(0,1)$, we obtain that $g(t)$ is continuous on $(0,1)$, which means $g(t) \in D_c$ as claimed.

ii) For the monotonicity claim, let $0 \leq t_1 \leq t_2 < \infty$ be arbitrary fixed. The strong solution g of the Cauchy problem (19) has the Bochner integral representation

$$g(t_2) = \int_{t_1}^{t_2} \partial_t g(t)\,\mathrm{d}t + g(t_1),$$

which allows a pointwise evaluation

$$g(t_2)(s) = \int_{t_1}^{t_2} \partial_t g(t)(s)\,\mathrm{d}t + g(t_1)(s) \quad \text{for a.e. } s \in (0,1).$$

Since $\partial_t g(t) \in -\partial F_\nu(g(t))$ for a.e. $t > 0$, and noting that $0 \leq R_\nu^- \leq R_\nu^+ \leq 1$, it follows for a.e. $s \in (0,1)$ that

$$\begin{aligned} g(t_2)(s) &\leq \int_{t_1}^{t_2} 2s - 2R_\nu^-\big(g(t)(s)\big)\,\mathrm{d}t + g(t_1)(s) \\ &\leq 2s(t_2 - t_1) + g(t_1)(s). \end{aligned}$$

This yields

$$\lim_{s\downarrow 0} g(t_2)(s) \leq \lim_{s\downarrow 0} g(t_1)(s). \tag{29}$$

Analogously, it holds for a.e. $s \in (0,1)$ that

$$\begin{aligned} g(t_2)(s) &\geq \int_{t_1}^{t_2} \Big(2s - 2R_\nu^+\big(g(t)(s)\big)\Big)\,\mathrm{d}t + g(t_1)(s) \\ &\geq 2(s-1)(t_2 - t_1) + g(t_1)(s), \end{aligned}$$

which implies

$$\lim_{s\uparrow 1} g(t_2)(s) \geq \lim_{s\uparrow 1} g(t_1)(s). \tag{30}$$

Combining (29) and (30), we have proved that

$$\lim_{s\downarrow 0} g(t_2)(s) \leq \lim_{s\downarrow 0} g(t_1)(s) \leq \lim_{s\uparrow 1} g(t_1)(s) \leq \lim_{s\uparrow 1} g(t_2)(s).$$

By part i), we know that $g(t_2) \in D_c$. The intermediate value theorem finally implies that $\overline{g(t_1)(0,1)} \subseteq \overline{g(t_2)(0,1)}$, and we are done. □

Remark 6.12. *Theorem 6.11 states that the support of the flow γ_t monotonically increases with the time t, given that the initial support* $\operatorname{supp}\mu_0$ *is convex. We leave it as an open problem, whether this still holds in general without the convexity assumption on* $\operatorname{supp}\mu_0$.

In Appendix A.4, we highlight some statements of this section from a different point of view.

7 Numerical Experiments

In this section, we first discuss a backward and a forward Euler scheme for the numerical computation of the one-dimensional MMD flow with the negative distance kernel and give some examples afterwards.

7.1 Euler Schemes

Implicit (backward) Euler scheme. The minimizing movement scheme (4), see also JKO scheme [28],

$$\mu_{n+1} := \operatorname*{arg\,min}_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \Big\{ \mathcal{F}_\nu(\mu) + \frac{1}{2\tau} W_2^2(\mu_n, \mu) \Big\}, \qquad \tau > 0,$$

can be rewritten by Theorem 3.2 in terms of quantile functions as

$$g_{n+1} = \operatorname*{arg\,min}_{g \in \mathcal{C}(0,1)} \Big\{ F_\nu(g) + \frac{1}{2\tau} \int_0^1 |g - g_n|^2 \, \mathrm{d}s \Big\}, \qquad \tau > 0.$$

Because F_ν is proper, convex and lsc, the solution of this problem is given by (also see (10))

$$g_{n+1} = (I + \tau \partial F_\nu)^{-1}(g_n) \tag{31}$$

which is by Lemma 4.3 and Lemma 6.1 equivalent to

$$g_n(s) + 2\tau s \in g_{n+1}(s) + 2\tau[R_\nu^-(g_{n+1}(s)), R_\nu^+(g_{n+1}(s))] \qquad \text{for all } s \in (0,1). \tag{32}$$

Note that by Lemma 6.1 the functions g_n and g_{n+1} are quantile functions and thus increasing, so that $g_n + 2\tau I$ is strictly increasing. Figure 7 gives a visual intuition for solving (32).

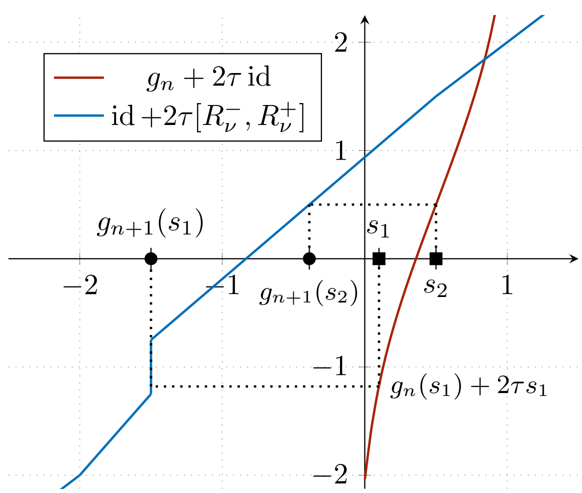

Figure 7: Implicit Euler step visualized for $g_n = Q_{\gamma_n}$ with $\gamma_n \sim \mathcal{N}(0,1)$ and $\tau = \frac{1}{2}$ and ν being a mixture of two uniform distributions. Starting from any point $s_1 \in (0,1)$ and $y := (g_n + 2\tau I)(s_1)$, we are searching for $x := g_{n+1}(s) \in \mathbb{R}$ such that $y \in x + 2\tau[R_\nu^-(x), R_\nu^+(x)]$. This is also illustrated for a point $s_2 \in (0,1)$.

For fixed $\tau > 0$, we perform $n = 1, 2, \dots$ implicit Euler steps (31) via solving (32). By Theorem 2.1, we see that $(g_n)_\# \Lambda_{(0,1)}$ can be considered as an approximation of the Wasserstein gradient flow in the time interval $(n\tau, (n+1)\tau]$. Further, for fixed τ and $n \to \infty$, the sequence $(g_n)_{n\in\mathbb{N}}$ converges weakly to Q_ν in $L_2(0,1)$, see [3, 34], and the corresponding measures $\mu_n \coloneqq (g_n)_\# \Lambda_{(0,1)}$ converge narrowly to the minimizer ν of $\mathcal{F}_\nu$, see [32, Thm 6.7]. The scheme (31) resembles a proximal point algorithm, and for convergence results for more general so-called quasi α-firmly nonexpansive mappings (instead of just proximal mappings), we refer to [5].

Explicit (forward) Euler scheme. For constant step size $\tau > 0$, we will also consider the explicit Euler discretization

$$g_{n+1} = g_n - \tau \nabla F_\nu(g_n),$$

which is only available if $R_\nu^+ = R_\nu^- =: R_\nu$ is continuous, so that

$$\nabla F_\nu(g) = 2R_\nu \circ g - 2I, \qquad g \in L_2(0,1).$$

The explicit Euler scheme has the advantage that we do not have to solve an inclusion in each step. However, this method comes with weaker convergence guarantees. Moreover, it might not preserve $\mathcal{C}(0,1)$, that is, the iterates might not be monotone. In all our numerical examples we chose $\tau = \frac{1}{100}$, which preserved monotonicity. However, monotonicity was not preserved for large step sizes, e.g., $\tau = 2$.

7.2 Numerical Experiments

Next, we compare the implicit and explicit Euler schemes for various combinations of absolutely continuous initial and target measures[4]. The following examples are covered by Theorem 6.5, or Corollary 6.9 i): since $L_{\text{low}}(Q_\nu) > 0$ in each figure, we have $L_{\text{low}}(Q_{\gamma_t}) > 0$ for all $t > 0$.

Flow between two Gaussians, resp., Laplacians. In Figure 8 we plot the MMD flow, where μ_0 and ν are both Gaussians with different means, but equal variance. Although the qualitative behaviors of the implicit and explicit schemes are similar, the explicit scheme "is ahead of" the implicit one. In particular, we observe that the shape of the normal distribution is not at all preserved during the flow and that instead the densities first spread out and become more flat and then form a peak again, when they meet the mean of the target distribution.

In Figure 9, this behavior can also be observed when the initial and the target measure are both Laplacians. Here we can also see that the non-differentiability of the density at its peak is smoothed out and then re-created during the flow.

In Figure 10 we plot the MMD flow, where μ_0 and ν are both Gaussians with different variance but equal mean. Again, the behavior of both discretizations is similar, just shifted in time, and the measures do not stay Gaussian. In particular, as predicted by Theorem 6.11 the visible support of γ_t is strictly increasing and matches the tails of the target distribution only exponentially slowly.

[4]The python code recreating these plots can be found at `https://github.com/ViktorAJStein/MMD_Wasserstein_gradient_flow_on_the_line/tree/main`.

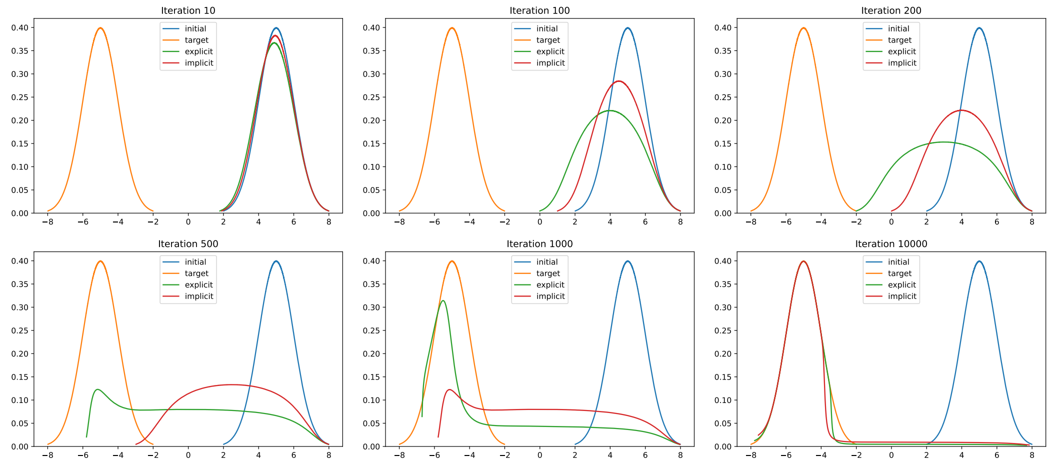

Figure 8: Comparison of implicit (red) and explicit (green) Euler schemes between two Gaussians $\mu_0 \sim \mathcal{N}(5,1)$ and $\nu \sim \mathcal{N}(-5,1)$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 15.

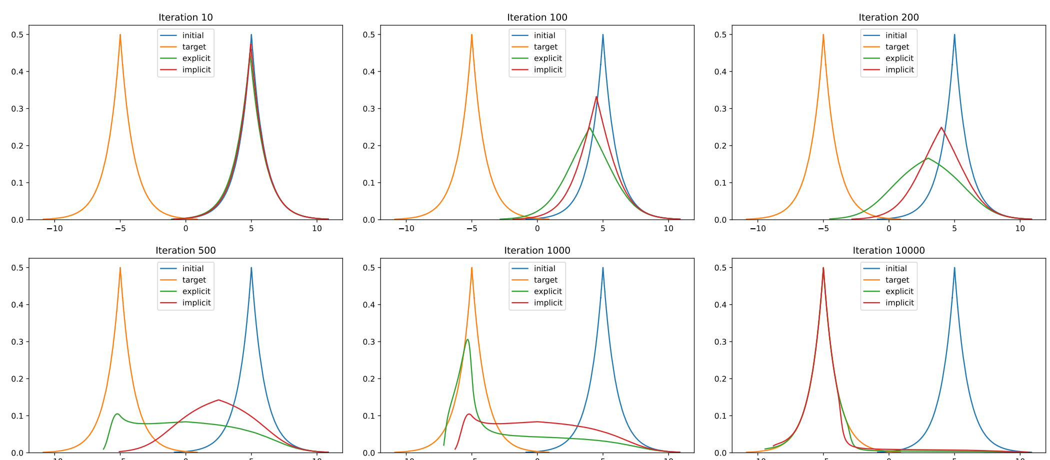

Figure 9: Comparison of implicit (red) and explicit (green) Euler schemes between two Laplacians $\mu_0 \sim \mathcal{L}(5,1)$ and $\nu \sim \mathcal{L}(-5,1)$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 16.

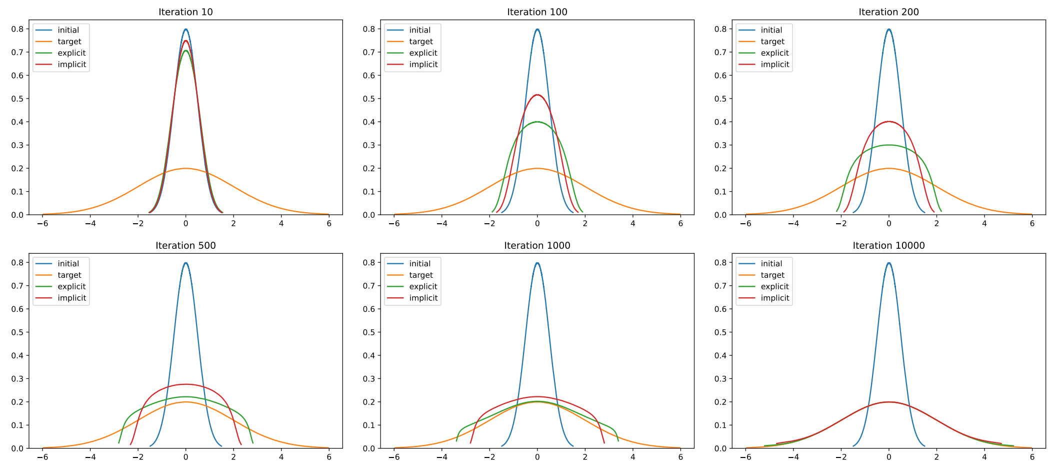

Figure 10: Comparison of implicit (red) and explicit (green) Euler schemes between two Gaussians $\mu_0 \sim \mathcal{N}(0, \sqrt{2})$ and $\nu \sim \mathcal{N}(0, \frac{1}{\sqrt{2}})$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 17.

Flow between measures with different number of modes. Consider now the case where the initial measure μ_0 is a Gaussian mixture model, which is symmetric with respect to the origin, and ν is a standard Gaussian. This case is plotted in Figure 11. We observe a similar relation between implicit and explicit scheme as in the previous example. Interestingly, we see that when the two parts of the visible support of γ_t collide at the origin, it takes a lot of time to recover the peak of the target measure, that is, the region of high density is recovered after regions of lower density.

In Figure 12 we switch initial and target measure. Here, we observe a similar behavior as in Figure 8: first the densities spread out and become more flat, and the two peaks of the target are developed not until the density meets the modes of the target.

Flow between measures without full support Our simulations can also handle initial measures and targets which are not everywhere supported. Consider for example the folded normal distribution $\mathcal{FN}(\mu, \sigma^2)$ with mean μ and scale $\sigma^2 > 0$, which has the density

$$\mathbb{R} \to \mathbb{R}, \qquad x \mapsto \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{x^2 + \mu^2}{2\sigma^2}\right) \cosh\left(\frac{\mu x}{\sigma^2}\right). \tag{33}$$

We consider the case of the target and initial measure being folded Gaussians in Figure 13. We again observe that the structure of the distribution is not preserved by the flow, the support grows monotonically, and that the tail of the target distribution is matched exponentially slow.

Acknowledgments. GS acknowledges funding by the German Research Foundation (DFG) within the project STE 571/16-1. JH acknowledges funding by the EPSRC programme grant *The Mathematics of Deep Learning* with reference EP/V026259/1.

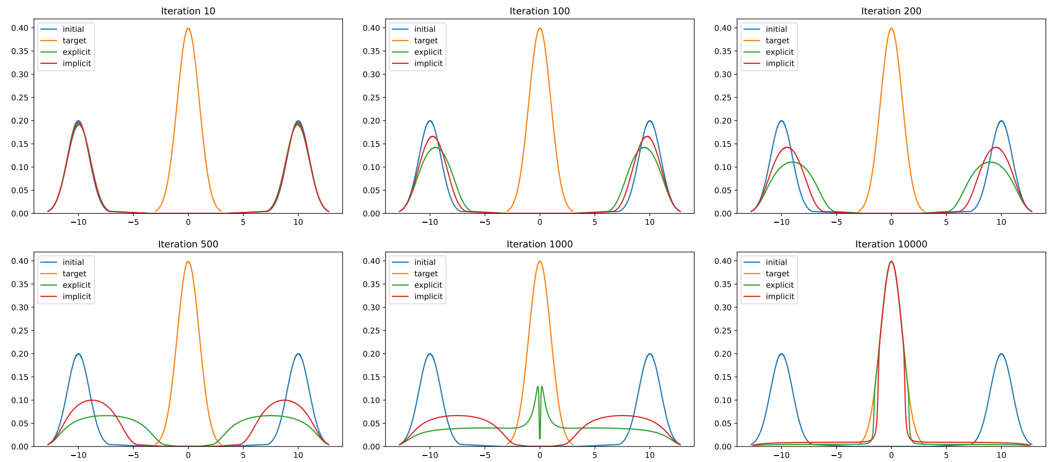

Figure 11: Comparison of implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \frac{1}{2}\mathcal{N}(-10,1) + \frac{1}{2}\mathcal{N}(10,1)$ and $\nu \sim \mathcal{N}(0,1)$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 18.

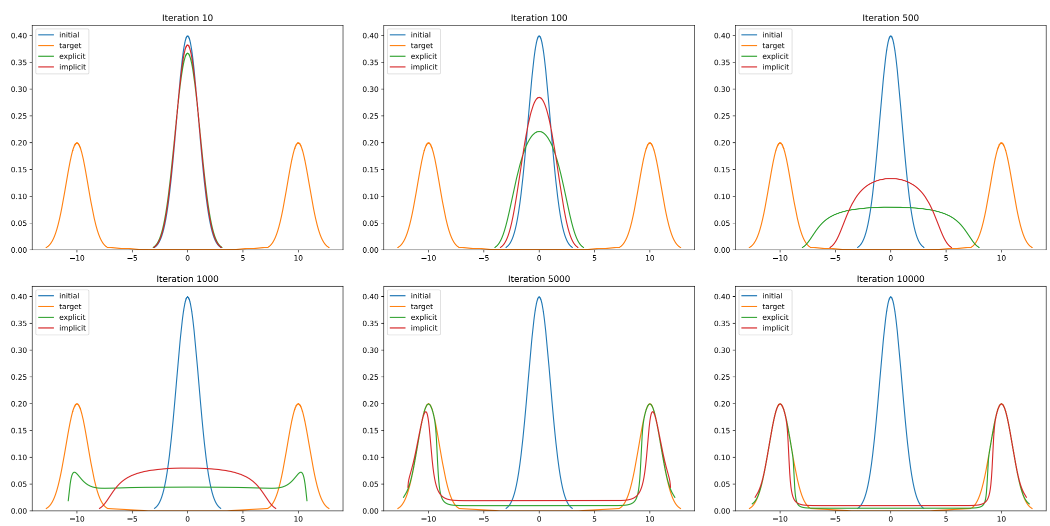

Figure 12: Comparison of implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \mathcal{N}(0,1)$ and $\nu \sim \frac{1}{2}\mathcal{N}(-10,1) + \frac{1}{2}\mathcal{N}(10,1)$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 19.

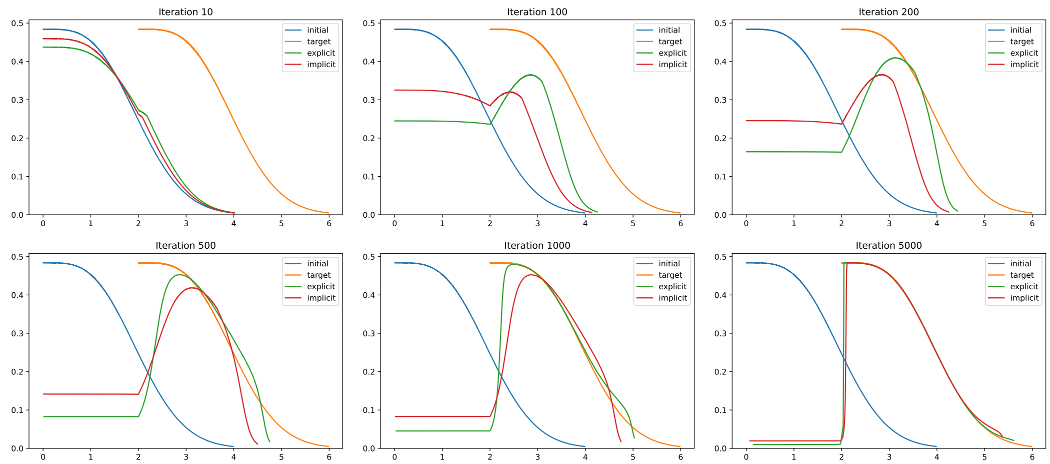

Figure 13: Comparison of implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \mathcal{FN}(0,1)$ and $\nu \sim \mathcal{FN}(2,1)$ with $\tau = \frac{1}{100}$. For the corresponding quantile functions, see Figure 20.

References

[1] L. Ambrosio, N. Gigli, and G. Savare. *Gradient Flows.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, 2nd edition, 2008.

[2] M. Arbel, A. Korba, A. Salim, and A. Gretton. Maximum mean discrepancy gradient flow. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[3] H. Bauschke and P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics. Springer New York, 2011.

[4] J. Benedetto and W. Czaja. *Integration and Modern Analysis.* Birkhäuser, Boston, 2009.

[5] A. Berdellima and G. Steidl. Quasi-alpha firmly nonexpansive mappings in Wasserstein spaces. *Fixed Point Theory, arXiv:2203.04851*, 2024.

[6] G. A. Bonaschi. Gradient flows driven by a non-smooth repulsive interaction potential. Master's thesis, University of Pavia, 2013.

[7] G. A. Bonaschi, J. A. Carrillo, M. Di Francesco, and M. A. Peletier. Equivalence of gradient flows and entropy solutions for singular nonlocal interaction equations in 1d. *ESAIM: Control, Optimisation and Calculus of Variations*, 21(2):414–441, 2015.

[8] S. Boufadène and F.-X. Vialard. On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy. *arXiv preprint arXiv:2312.00800*, 2023.

[9] H. Brezis. *Operateurs Maximaux Monotones.* North-Holland Mathematics Studies, 1973.

[10] M. Burger and M. Di Francesco. Large time behavior of nonlocal aggregation models with nonlinear diffusion. *Networks and Heterogeneous Media*, 3(4):749–785, 2008.

[11] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: Mean-field limit and Wasserstein distances. In A. Muntean and F. Toschi, editors, *Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation*, pages 1–46. Springer Vienna, Vienna, 2014.

[12] J. A. Carrillo, M. Di Francesco, A. Esposito, S. Fagioli, and M. Schmidtchen. Measure solutions to a system of continuity equations driven by Newtonian nonlocal interactions. *Discrete and Continuous Dynamical Systems*, 40(2):1191–1231, 2020.

[13] J. A. Carrillo, D. Slepčev, and L. Wu. Nonlocal-interaction equations on uniformly prox-regular sets. *Discrete and Continuous Dynamical Systems*, 36(3):1209–1247, 2016.

[14] D. Chafaï, R. Matzke, E. Saff, M. Vu, and R. Womersley. Riesz energy with a radial external field: When is the equilibrium support a sphere? *arXiv preprint arXiv:2405.00120*, 2024.

[15] Y. Chen, D. Z. Huang, J. Huang, S. Reich, and A. M. Stuart. Sampling via gradient flows in the space of probability measures. *arXiv preprint arXiv:2310.03597*, 2023.

[16] M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.

[17] H. Daneshmand and F. Bach. Polynomial-time sparse measure recovery: From mean field theory to algorithm design. *arXiv preprint arXiv:2204.07879*, 2022.

[18] H. Daneshmand, J. D. Lee, and C. Jin. Efficient displacement convex optimization with particle gradient descent. In *International Conference on Machine Learning*, pages 6836–6854. PMLR, 2023.

[19] M. Di Francesco. Scalar conservation laws seen as gradient flows: known results and new perspectives. *ESAIM: Proceedings and Surveys*, 54:18–44, 2016.

[20] M. Di Francesco, S. Fagioli, and E. Radici. Measure solutions, smoothing effect, and deterministic particle approximation for a conservation law with nonlocal flux. arXiv preprint arXiv:2406.03837, 2024.

[21] K. Fellner and G. Raoul. Stability of stationary states of non-local equations with singular interaction potentials. *Mathematical and Computer Modelling*, 53(7):1436–1450, 2011.

[22] M. Fornasier, J. Haskovec, and G. Steidl. Consistency of variational continuous-domain quantization via kinetic theory. *Applicable Analysis*, 92(6):1283–1298, 2013.

[23] P. Hagemann, J. Hertrich, F. Altekrüger, R. Beinert, J. Chemseddine, and G. Steidl. Posterior sampling based on gradient flows of the MMD with negative distance kernel. In *International Conference on Learning Representations*, 2024.

[24] J. Hertrich, R. Beinert, M. Gräf, and G. Steidl. Wasserstein gradient flows of the discrepancy with distance kernel on the line. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 431–443. Springer, 2023.

[25] J. Hertrich, M. Gräf, R. Beinert, and G. Steidl. Wasserstein steepest descent flows of discrepancies with Riesz kernels. *Journal of Mathematical Analysis and Applications*, 531(1):127829, 2024.

[26] J. Hertrich, C. Wald, F. Altekrüger, and P. Hagemann. Generative sliced MMD flows with Riesz kernels. In *International Conference on Learning Representations*, 2024.

[27] Y. Huang, E. Mainini, J. L. Vázquez, and B. Volzone. Nonlinear aggregation-diffusion equations with Riesz potentials. *Journal of Functional Analysis*, 287(2):110465, 2024.

[28] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker–Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[29] R. Laumont, V. Bortoli, A. Almansa, J. Delon, A. Durmus, and M. Pereyra. Bayesian imaging using plug & play priors: when Langevin meets Tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737, 2022.

[30] H. Li and G. Toscani. Long-time asymptotics of kinetic models of granular flows. *Archive for Rational Mechanics and Analysis*, 172:407–428, 2004.

[31] F. Maggi. *Optimal Mass Transport on Euclidean Spaces.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2023.

[32] E. Naldi and G. Savaré. Weak topology and Opial property in Wasserstein spaces with applications to gradient flows and proximal point algorithms of geodesically convex functionals. *Rendiconti Lincei Matematica e Applicazioni*, 32(4):725–750, 2022.

[33] P. I. Natanson. *Theory of Functions of a Real Variable.* Ungar, New York, 1955.

[34] J. Peypouquet and S. Sorin. Evolution equations for maximal monotone operators: Asymptotic analysis in continuous and discrete time. *Journal of Convex Analysis*, 17(3&4):1113–1163, 2010.

[35] R. T. Rockafellar. Integrals which are convex functionals. II. *Pacific Journal of Mathematics*, 39(2):439–469, 1971.

[36] R. T. Rockafellar and J. O. Royset. Random variables, monotone relations, and convex analysis. *Mathematical Programming*, 148:297–331, 2014.

[37] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167. Springer, 2009.

[38] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263 – 2291, 2013.

[39] G. Székely. E-statistics: The energy of statistical samples. *Techical Report, Bowling Green University*, 2002.

[40] G. J. Székely and M. L. Rizzo. Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272, 2013.

[41] C. Villani. *Topics in Optimal Transportation.* Number 58 in Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.

[42] C. Villani. *Optimal Transport.* Springer, Berlin, 2009.

A Additional Material

A.1 Supplement to Section 3

Remark A.1. *To be completely accurate, the set of quantile functions is the set of increasing and left-continuous functions (not equivalence classes) in $\mathcal{L}_2(0,1)$. However, each almost everywhere increasing function $f \in L_2(0,1)$ has a unique left-continuous representative: consider a function $\tilde{f} \in \mathcal{L}_2(0,1)$ in the equivalence class f, which is everywhere increasing. Then $\tilde{f}$ has at most countably many jump discontinuities $(s_k)_{k\in\mathbb{N}}$. Define $f_0\colon (0,1) \to \mathbb{R}$ via $s \mapsto \lim_{t\uparrow s} \tilde{f}(t)$. Then f_0 differs from $\tilde{f}$ only at possibly $(s_k)_{k\in\mathbb{N}}$, and f_0 is increasing everywhere and left-continuous.*

Indeed, $\mathcal{C}(0,1)$ is closed, since for any sequence $(f_n)_{n\in\mathbb{N}} \subset \mathcal{C}(0,1)$ converging to f in $L_2(0,1)$ as $n \to \infty$, we have $f_n(s) \to f(s)$ along a subsequence for almost all $s \in (0,1)$. Since the pointwise limit of increasing functions is increasing, we see that $f \in \mathcal{C}(0,1)$.

Proof of Proposition 3.4 Let $\lambda \in \mathbb{R}$ and $F\colon L_2(0,1) \to \mathbb{R}$ be λ-convex, that is,

$$F\left(tu + (1-t)v\right) \le tF(u) + (1-t)F(v) - \frac{1}{2}\lambda t(1-t)\|u-v\|^2$$

for all $u, v \in \operatorname{dom}(F) \cap \mathcal{C}(0,1)$ and all $t \in (0,1)$. Let $\gamma\colon [0,1] \to \mathcal{P}_2(\mathbb{R})$ be any geodesic with $\gamma_0 = \mu$ and $\gamma_1 = \nu$. Since $\mu \mapsto Q_\mu$ is an isometry by Theorem 3.2, the curve $t \mapsto Q_{\gamma_t}$ is a geodesic in $L_2(0,1)$. Since $L_2(0,1)$ is a linear space, the only geodesics are straight line segments, so we obtain

$$Q_{\gamma_t} = (1-t)Q_\mu + tQ_\nu.$$

Finally, we conclude

$$\begin{aligned}
\mathcal{F}(\gamma_t) &= F(Q_{\gamma_t}) = F\big((1-t)Q_\mu + tQ_\nu\big) \\
&\le tF(Q_\mu) + (1-t)F(Q_\nu) - \frac{1}{2}\lambda t(1-t)\|Q_\mu - Q_\nu\|^2_{L_2(0,1)} \\
&= t\,\mathcal{F}(\mu) + (1-t)\,\mathcal{F}(\nu) - \frac{1}{2}\lambda t(1-t)W_2^2(\mu,\nu).
\end{aligned}$$

The lower semicontinuity of $\mathcal{F}$ follows directly from the lower semicontinuity of F using the isometric embedding, see Theorem 3.2. □

Theorem A.2 (Existence and regularity of strong solutions to (9) [9, Thm. 3.1, p. 54], [16])**.** *Let H be a Hilbert space and $A\colon H \to 2^H$ be a maximal monotone operator. For all $g_0 \in \operatorname{dom}(A)$, there exists a unique function $g\colon [0,\infty) \to H$ such that*

1. *$g(t) \in \operatorname{dom}(A)$ for all $t > 0$,*
2. *g is Lipschitz continuous on $[0,\infty)$, that is, $\frac{\mathrm{d}g}{\mathrm{d}t} \in L_\infty((0,\infty), H)$ (in the sense of distributions and in the strong sense a.e.) and*

$$\left\|\frac{\mathrm{d}g}{\mathrm{d}t}\right\|_{L_\infty((0,\infty),H)} \le \|A^\circ g_0\|,$$

where $A^\circ z \coloneqq \arg\min\{\|y\| : y \in Az\}$ denotes the minimal norm selection, *3.* $\frac{\mathrm{d}g}{\mathrm{d}t}(t) \in -Ag(t)$ *for almost all* $t > 0$,

4. $g(0) = g_0$.

Furthermore, g *satisfies the following properties:*

5. g *admits a right derivative for all* $t > 0$ *and* $\frac{\mathrm{d}^+ g}{\mathrm{d}t}(t) + A^\circ g(t) = 0$ *for all* $t > 0$,

6. $t \mapsto A^\circ g(t)$ *is right-continuous and* $t \mapsto \|A^\circ g(t)\|$ *is decreasing.*

7. g *is given by the exponential formula*

$$g(t) = \lim_{n\to\infty} \left(I + \frac{t}{n}A\right)^{-n} g_0,$$

uniformly in t *on compact intervals.*

A.2 Supplement to Section 4

Proof of Proposition 4.2 The functional F_ν is convex, since for $u, v \in L_2(0,1)$ and $t \in (0,1)$ we have

$$\begin{aligned} F_\nu\big(tu + (1-t)v\big) &= \int_0^1 \Bigg((1-2s)\big(tu(s) + (1-t)v(s) - Q_\nu(s)\big) \\ &\quad + \int_0^1 |tu(s) + (1-t)v(s) - Q_\nu(t)|\,\mathrm{d}t\Bigg)\,\mathrm{d}s \\ &\le t\int_0^1 \left((1-2s)\big(u(s) - Q_\nu(s)\big) + \int_0^1 |u(s) - Q_\nu(t)|\,\mathrm{d}t\right)\mathrm{d}s \\ &\quad + (1-t)\int_0^1 \left((1-2s)\big(v(s) - Q_\nu(s)\big) + \int_0^1 |v(s) - Q_\nu(t)|\,\mathrm{d}t\right)\mathrm{d}s \\ &= tF_\nu(u) + (1-t)F_\nu(v). \end{aligned}$$

To show that F_ν is everywhere finite, notice that for all $u \in L_2(0,1)$,

$$\begin{aligned} |F_\nu(u)| &\le \int_0^1 |1-2s||u(s) - Q_\nu(s)| + |u(s)| + \|Q_\nu\|_{L_1(0,1)}\,\mathrm{d}s \\ &\le 2(\|u\|_{L_2(0,1)} + \|Q_\nu\|_{L_2(0,1)}) < \infty. \end{aligned}$$

Lastly, we will show the continuity of F_ν. Suppose that $(u_n)_{n\in\mathbb{N}} \subset L_2(0,1)$ converges to $u \in L_2(0,1)$. Then there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ such that $u_{n_k}(s) \to u(s)$ for $k \to \infty$ for a.e. $s \in (0,1)$, and there exists a $m \in L_1(0,1)$ such that $|u_{n_k}(s)| \le m(s)$ for a.e. $s \in (0,1)$.

Hence applying Lebesgue's dominated convergence theorem twice yields

$$\begin{aligned} \lim_{k\to\infty} F_\nu(u_{n_k}) &= \lim_{k\to\infty} \int_0^1 \left((1-2s)\big(u_{n_k}(s) - Q_\nu(s)\big) + \int_0^1 |u_{n_k}(s) - Q_\nu(t)|\,\mathrm{d}t\right)\mathrm{d}s \\ &= F_\nu(u). \end{aligned}$$

We can apply the dominated convergence theorem because in both cases the integrands are bounded above by an integrable function as follows: for almost all $s \in (0,1)$ we have by applying the triangle inequality many times that

$$\begin{aligned}
&\left|(1-2s)\big(u_{n_k}(s) - Q_\nu(s)\big) + \int_0^1 |u_{n_k}(s) - Q_\nu(t)|\,\mathrm{d}t\right| \\
&\le \underbrace{|1-2s|}_{\le 1}\,|u_{n_k}(s) - Q_\nu(s)| + \int_0^1 |u_{n_k}(s) - Q_\nu(t)|\,\mathrm{d}t \\
&\le |u_{n_k}(s)| + |Q_\nu(s)| + |u_{n_k}(s)| + \int_0^1 |Q_\nu(t)|\,\mathrm{d}t \\
&\le 2m(s) + |Q_\nu(s)| + \|Q_\nu\|_{L_1(0,1)}.
\end{aligned}$$

This function is integrable because $m \in L_1(0,1)$ and $Q_\nu \in L_2(0,1) \subset L_1(0,1)$. Analogously, for any $s \in (0,1)$, the integrand $(0,1) \ni t \mapsto |u_{n_k}(s) - Q_\nu(t)|$ is bounded by the integrable function $t \mapsto |u_{n_k}(s)| + |Q_\nu(t)|$. ☐

A.3 Supplement to Section 6

Proof of Lemma 6.1 Let $s_0 \in (0,1)$ be arbitrary and take a sequence $(s_n) \subset (0,1)$ of values s where (23) holds, and such that $s_n \uparrow s_0$. Then, it holds

$$h(s_0) + \varepsilon s_0 = \lim_{n\to\infty} h(s_n) + \varepsilon s_n \ge \lim_{n\to\infty} u(s_n) + \varepsilon R_\nu^-(u(s_n)) = u(s_0) + \varepsilon R_\nu^-(u(s_0))$$

by the left-continuity of h, u and R_ν^-, and since u is increasing. Also, one has

$$h(s_0) + \varepsilon s_0 = \lim_{n\to\infty} h(s_n) + \varepsilon s_n \le \lim_{n\to\infty} u(s_n) + \varepsilon R_\nu^+(u(s_n)) \le u(s_0) + \varepsilon R_\nu^+(u(s_0)),$$

since R_ν^+ is upper semicontinuous. Since $s_0 \in (0,1)$ was arbitrary, this shows the claim. ☐

Proof of Proposition 6.3 Let us write $R_\mu \coloneqq R_\mu^+$.
i) *'Only if'*: Let $x, y \in R_\mu^{-1}((0,1))$ with $x > y$. W.l.o.g. suppose $R_\mu(x) > R_\mu(y)$ and let $\varepsilon > 0$ such that $R_\mu(x) > R_\mu(y) + \varepsilon$. Then, it holds $Q_\mu(R_\mu(x)) \le x$ and $Q_\mu(R_\mu(y)+\varepsilon) > y$ by the Galois inequalities (7). Now, using that $Q_\mu \in D_{L,\mathrm{low}}$, it follows that

$$\begin{aligned}
L|R_\mu(x) - (R_\mu(y)+\varepsilon)| &\le |Q_\mu(R_\mu(x)) - Q_\mu(R_\mu(y)+\varepsilon)| \\
&= Q_\mu(R_\mu(x)) - Q_\mu(R_\mu(y)+\varepsilon) \\
&\le x - y = |x-y|.
\end{aligned}$$

Letting $\varepsilon \downarrow 0$ shows

$$|R_\mu(x) - R_\mu(y)| \le \frac{1}{L}|x-y| \quad \text{for all } x, y \in R_\mu^{-1}((0,1)).$$

Since $R_\mu \equiv 0$ on $R_\mu^{-1}((-\infty,0])$, $R_\mu \equiv 1$ on $R_\mu^{-1}([1,\infty))$, and since R_μ is continuous (on $\mathbb{R}$) by Remark 3.1, R_μ is Lipschitz continuous on $\mathbb{R}$ with constant $\le L^{-1}$, which proves the 'only-if' part.

'If': Since R_μ is (Lipschitz) continuous, one has $R_\mu(Q_\mu(s)) = s$ for all $s \in (0,1)$. It immediately follows

$$|Q_\mu(s_1) - Q_\mu(s_2)| \geq L|R_\mu(Q_\mu(s_1)) - R_\mu(Q_\mu(s_2))| = L|s_1 - s_2|$$

for all $s_1, s_2 \in (0,1)$, which concludes the proof of part i).

ii) *'Only if'*: Since Q_μ is (Lipschitz) continuous, R_μ is strictly increasing on $R_\mu^{-1}((0,1))$ by Remark 3.1, and hence, $Q_\mu(R_\mu(x)) = x$ for all $x \in R_\mu^{-1}((0,1))$. It immediately follows

$$|R_\mu(x) - R_\mu(y)| \geq \frac{1}{L}|Q_\mu(R_\mu(x)) - Q_\mu(R_\mu(y))| = \frac{1}{L}|x - y|$$

for all $x, y \in R_\mu^{-1}((0,1))$. This inequality directly extends to all $x, y \in \overline{R_\mu^{-1}((0,1))}$, since R_μ is increasing.

'If': Let $s_1, s_2 \in (0,1)$ with $s_1 > s_2$. W.l.o.g. suppose $Q_\mu(s_1) > Q_\mu(s_2)$ and let $\varepsilon > 0$ such that $Q_\mu(s_1) - \varepsilon > Q_\mu(s_2)$. Then, it holds $R_\mu(Q_\mu(s_2)) \geq s_2$ and $R_\mu(Q_\mu(s_1) - \varepsilon) < s_1$ by the Galois inequalities (7). Now, using the lower Lipschitz bound of R_μ, it follows

$$\begin{aligned}
\frac{1}{L}|(Q_\mu(s_1) - \varepsilon) - Q_\mu(s_2)| &\leq |R_\mu(Q_\mu(s_1) - \varepsilon) - R_\mu(Q_\mu(s_2))| \\
&= R_\mu(Q_\mu(s_1) - \varepsilon) - R_\mu(Q_\mu(s_2)) \\
&\leq s_1 - s_2 = |s_1 - s_2|.
\end{aligned}$$

Letting $\varepsilon \downarrow 0$ shows $Q_\mu \in D_L$, which completes the proof of ii).

Finally, assume that μ is absolutely continuous. Then the bounds for its density follow by applying the fundamental theorem of calculus for absolutely continuous R_μ, i.e., for all $x \geq y$,

$$|R_\mu(x) - R_\mu(y)| = R_\mu(x) - R_\mu(y) = \int_y^x f_\mu(\xi)\, \mathrm{d}\xi. \qquad \square$$

A.4 Smoothness Properties: Different Approach

In the rest of this section, we highlight the statements from Section 6 from a different point of view. Instead of working with the exponential formula (10), we deal with the pointwise differential inclusion (20), allowing for a more direct approach. [5]

The following theorem is analogous to Theorem 6.7. Let us remark that both theorems also hold true in a local sense, i.e., on subintervals $I \subseteq (0,1)$.

Theorem A.3. *Let $I := (a,b)$ be such that $Q_{\mu_0}(I) \subset \operatorname{supp} \nu$. If Q_{μ_0} is locally upper Lipschitz on I with constant $L_{\mu_0} \geq 0$, and if $R_\nu^\pm$ is locally lower Lipschitz on $\operatorname{conv}(Q_{\mu_0}(I) \cup Q_\nu(I))$ with constant $L_\nu^{-1} > 0$, then $g(t)$ is locally upper Lipschitz on I with*

$$\operatorname{Lip}(g(t)) \leq L_{\mu_0}\, e^{-2tL_\nu^{-1}} + L_\nu\, (1 - e^{-2tL_\nu^{-1}}).$$

Proof. W.l.o.g. we consider $s_1 < s_2$ in I with $Q_{\mu_0}(s) < Q_\nu(s)$ for all $s_1 \leq s \leq s_2$. Due to the local lower Lipschitz property of R_ν^+, the derivatives of g_{s_1} are bounded by

$$\dot{g}_{s_1}(t) = 2s_1 - 2R_\nu^+(g_{s_1}(t)) \geq 2L_\nu^{-1}(Q_\nu(s_1) - g_{s_1}(t))$$

[5] But note that the exponential approach of Section 6 also yields information about the implicit Euler steps.

for almost every $t \geq 0$. Solving the differential equation with the lower bound and the initial value $Q_{\mu_0}(s_1)$, we obtain

$$g_{s_1}(t) \geq Q_\nu(s_1) + (Q_{\mu_0}(s_1) - Q_\nu(s_1))\, e^{-2tL_\nu^{-1}}.$$

A similar procedure for g_{s_2} yields

$$\dot{g}_{s_2}(t) = 2s_2 - 2R_\nu^+(g_{s_2}(t)) \leq 2s_2 - 2R_\nu^+(Q_{\mu_0}(s_2) + 2L_\nu^{-1}(Q_{\mu_0}(s_2) - g_{s_2}(t))$$

and

$$g_{s_2}(t) \leq L_\nu(s_2 - R_\nu^+(Q_{\mu_0}(s_2))) + Q_{\mu_0}(s_2) - L_\nu(s_2 - R_\nu^+(Q_{\mu_0}(s_2)))\, e^{-2tL_\nu^{-1}}.$$

Taking the difference and exploiting that

$$\begin{aligned} Q_{\mu_0}(s_2) - Q_\nu(s_1) &\leq L_\nu(R_\nu^+(Q_{\mu_0}(s_2)) - R_\nu^+(Q_\nu(s_1))) \\ &\leq L_\nu(R_\nu^+(Q_{\mu_0}(s_2)) - s_1) \end{aligned}$$

by the lower Lipschitz property of R_ν^+, we obtain

$$\begin{aligned} |g_{s_2}(t) - g_{s_1}(t)| &\leq (Q_{\mu_0}(s_2) - Q_{\mu_0}(s_1))\, e^{-2tL_\nu^{-1}} \\ &\quad + (Q_{\mu_0}(s_2) - Q_\nu(s_1) + L_\nu s_2 - L_\nu R_\nu^+(Q_{\mu_0}(s_2))\,(1 - e^{-2tL_\nu^{-1}}) \\ &\leq (L_{\mu_0}\, e^{-2tL_\nu^{-1}} + L_\nu\,(1 - e^{-2tL_\nu^{-1}}))\,(s_2 - s_1). \end{aligned}$$

□

We can analogously argue for the lower Lipschitz property.

The following theorem is a counterpart to Theorem 6.11, but with a subtle difference in its nature: it states an invariance of the continuity in a *single point* $s \in (0,1)$, assuming in addition that s is a continuity point of the target Q_ν. On the other hand, Theorem 6.11 states an invariance of continuity on *neighborhoods* of s, without any further assumptions on the target.

Theorem A.4. *Let $s \in (0,1)$ be a continuity point of Q_ν and Q_{μ_0}, then $g(t)$ is continuous at s for all $t \geq 0$.*

Proof. Since Q_ν is continuous in s, for all $x < Q_\nu(s)$, there exists $\delta_x > 0$ such that $\Phi_{\nu,s'}(x) < \infty$ for all $s' \in (s - \delta_x, s + \delta_x)$. The integrand of $\Phi_{\nu,s}$ is monotone in s such that $\Phi_{\nu,s}(x)$ is continuous in s due to Lebesgue's dominated convergence theorem. To show that $g(t)$ is continuous in s, we use the ϵ–δ criterion. First, we discuss the case $2t = \Phi_{\nu,s}(x)$ with $x < Q_\nu(s)$. For arbitrary $\epsilon > 0$, we find $x_1 \in (x - \epsilon, x)$ and $\delta_1 > 0$ such that $\Phi_{\nu,s'}(x_1) < 2t$ for all $s' \in (s - \delta_1, s)$. Due to the monotonicity of $\Phi_{\nu,s}(x)$ in s, we have $\Phi_{\nu,s}^{-1}(2t) \in (x - \epsilon, x)$ for all $s' \in (s - \delta_1, s)$. Analogously, we find $x_2 \in (x, x + \epsilon)$ and $\delta_2 > 0$ such that $\Phi_{\nu,s'}(x_2)$ is finite and $\Phi_{\nu,s'}(x_2) > 2t$ for all $s' \in (s, s + \delta_2)$. Again due to the monotonicity, we have $\Phi_{\nu,s}^{-1}(2t) \in (x, x + \epsilon)$ for all $s' \in (s, s + \delta_2)$. Taking $\delta \coloneqq \min\{\delta_1, \delta_2\}$, we finally have

$$[g(t)](s') \in ([g(t)](s) - \epsilon, [g(t)](s) + \epsilon) \quad \text{for all} \quad s' \in (s - \delta, s + \delta).$$

Second, for the case $2t \geq \Phi_{\nu,s}(Q_\nu(s))$, we have to argue slightly differently. Here, we find $x_1 \in (Q_\nu(s) - \epsilon, Q_\nu(s))$ and $\delta_1 > 0$ such that $\Phi_{\nu,s'}(x_1) < 2t$ and thus $\Phi_{\nu,s'}^{-1}(2t) > Q_\nu(s) - \epsilon$ for all $s' \in (s - \delta_1, s + \delta_1)$. Due to the continuity of the target Q_ν in s, we find $\delta_2 > 0$ such that $Q_\nu(s') < Q_\nu(s) + \epsilon)$ for all $s' \in (s - \delta_2, s + \delta_2)$. Taking $\delta \coloneqq \min\{\delta_1, \delta_2\}$, we again have

$$[g(t)](s') \in ([g(t)](s) - \epsilon, [g(t)](s) + \epsilon) \quad \text{for all} \quad s' \in (s - \delta, s + \delta).$$

□

A.5 Quantile functions plots for the experiments

In this supplementary section we plot the quantile functions belonging to the densities in figures 4 and 8 – 13 from the main text.

Here, one can nicely verify the results from the previous subsection, in particular, that a pointwise view on the Cauchy problem on quantile functions is insightful.

In the fully discrete case, see Figure 14, we have $Q_{\mu_0} \notin D_c$, and so we can not apply Theorem 6.11 i). In fact, $Q_{\gamma_t} \notin D_c$ for all $t \geq 0$.

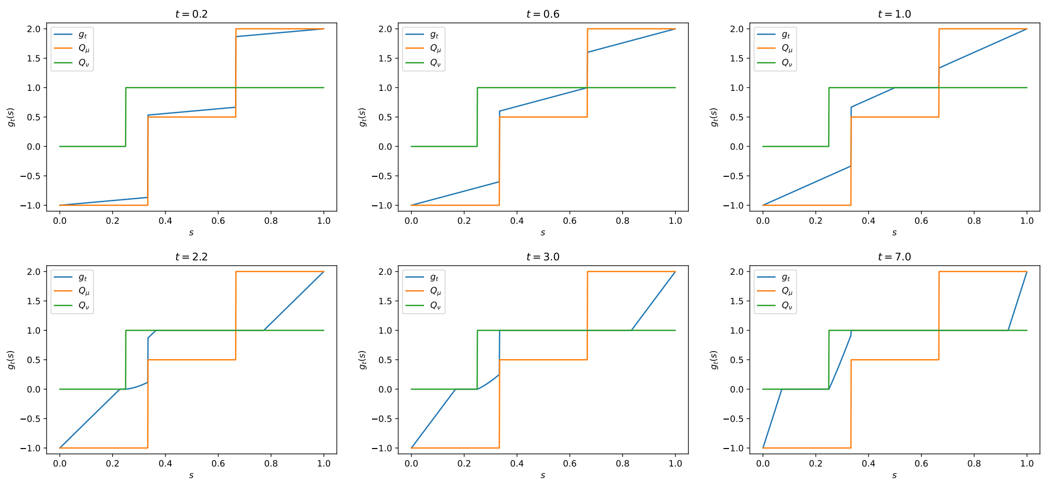

Figure 14: The quantile functions of the Wasserstein gradient flow between the point measures in Example 5.4. For the corresponding densities, see Figure 4.

Concerning the examples with absolutely continuous target, plotted in figures 15 – 20, we observe that the quantile functions differ mostly at the boundary, that is, for values of s close to zero or close to one. Furthermore, we observe that Proposition 6.4 holds true, i.e., the range of the quantile functions Q_{γ_t} does not leave the range of Q_{μ_0} and Q_ν, and we can verify the monotonicity of the ranges of Q_{γ_t} predicted by Theorem 6.11 ii). We also see that Theorem 6.11 holds true: the quantile functions stay continuous because $g_0 = Q_{\mu_0}$ is continuous. Lastly, we have $\mathrm{Lip}(Q_{\mu_0}) = \infty$ in all these examples, so Theorem 6.7 does not apply. Indeed, $Q_{\gamma_t} \notin D_L^+$ for all $L > 0$.

A.6 Implicit Euler Scheme for the Flow Towards δ_0 Starting in a Uniform or Gaussian Measure

Let $\nu = \delta_0$. Then the multi-valued operator $I + 2\tau[R_\nu^-, R_\nu^+]$ appearing on the left hand side of (32) is given by

$$\mathbb{R} \ni x \mapsto x + 2\tau[R_\nu^-(x), R_\nu^+(x)] = \begin{cases} x, & \text{if } x < 0, \\ [0, 2\tau], & \text{if } x = 0, \\ x + 2\tau, & \text{if } x > 0, \end{cases}$$

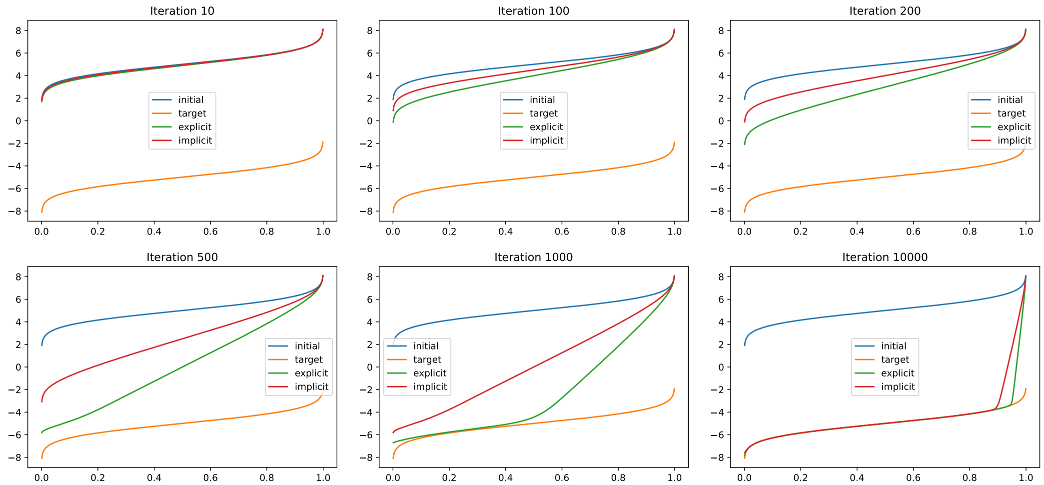

Figure 15: Comparison of the quantile functions belonging to the implicit (red) and explicit (green) Euler schemes between two Gaussians $\mu_0 \sim \mathcal{N}(5,1)$ and $\nu \sim \mathcal{N}(-5,1)$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 8.

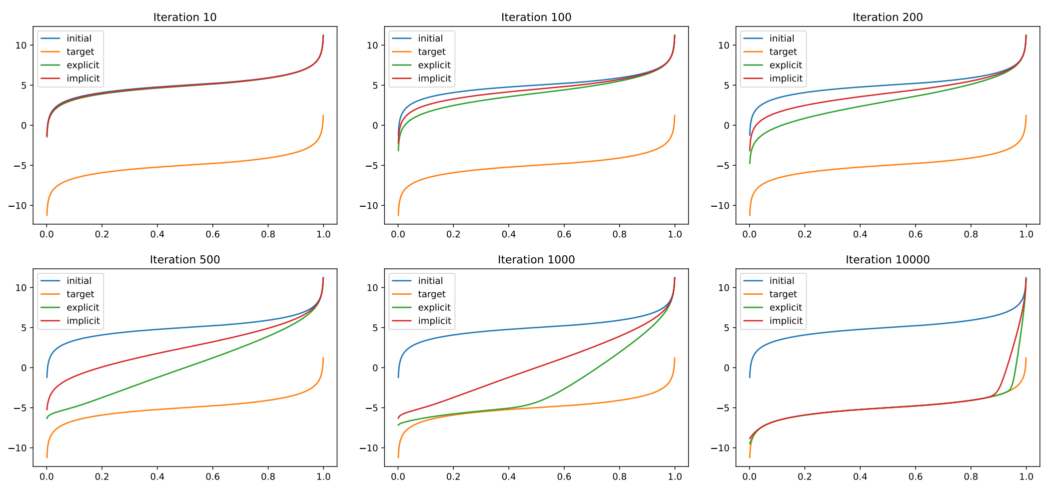

Figure 16: Comparison of the quantile functions belonging to the implicit (red) and explicit (green) Euler schemes between two Laplacians $\mu_0 \sim \mathcal{L}(5,1)$ and $\nu \sim \mathcal{L}(-5,1)$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 9.

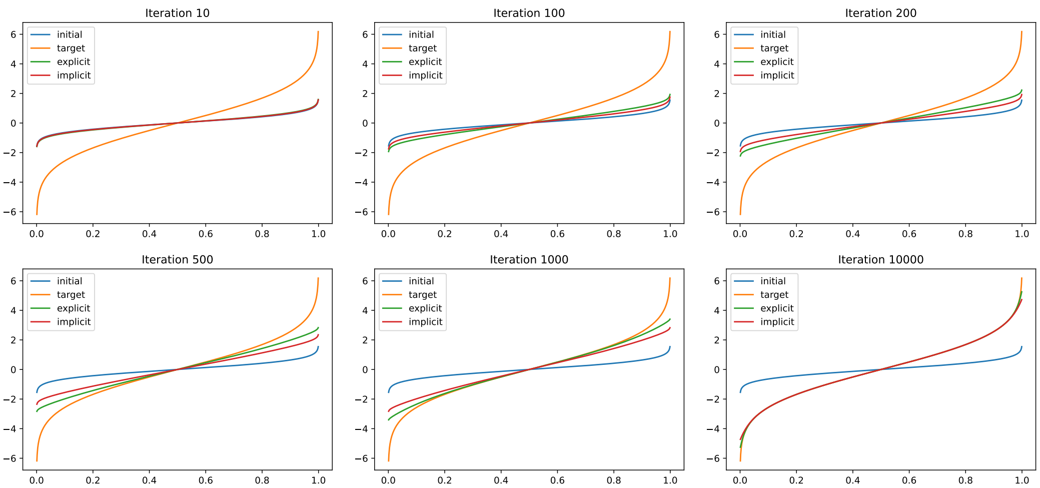

Figure 17: Comparison of the quantile functions belonging to the implicit (red) and explicit (green) Euler schemes between two Gaussians $\mu_0 \sim \mathcal{N}(0, \sqrt{2})$ and $\nu \sim \mathcal{N}(0, \frac{1}{\sqrt{2}})$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 10.

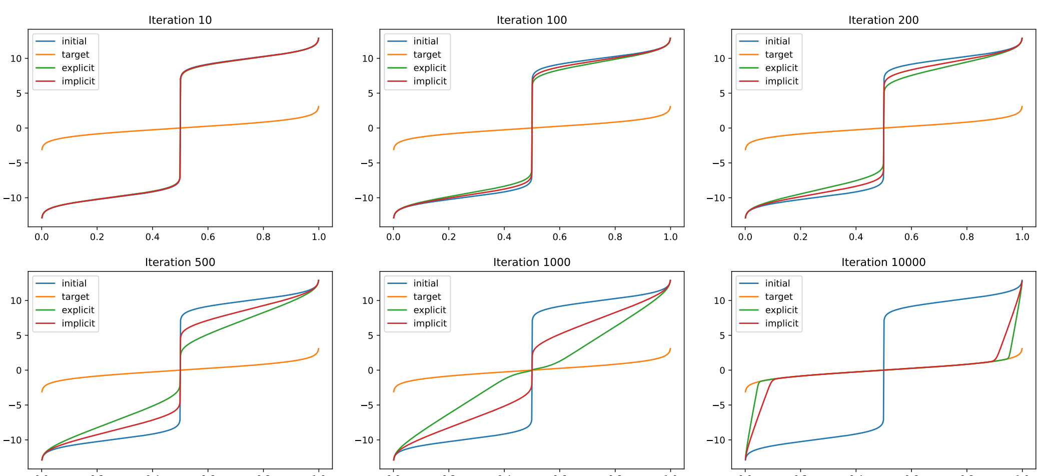

Figure 18: Comparison of of the quantile functions belonging to the implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \frac{1}{2}\mathcal{N}(-10, 1) + \frac{1}{2}\mathcal{N}(10, 1)$ and $\nu \sim \mathcal{N}(0, 1)$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 11.

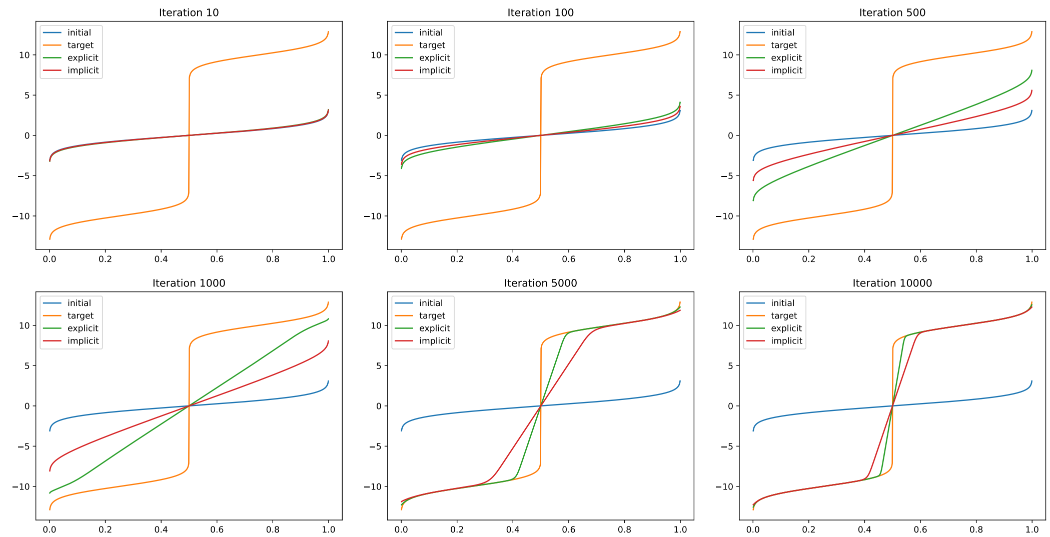

Figure 19: Comparison of implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \mathcal{N}(0,1)$ and $\nu \sim \frac{1}{2}\mathcal{N}(-10,1) + \frac{1}{2}\mathcal{N}(10,1)$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 12.

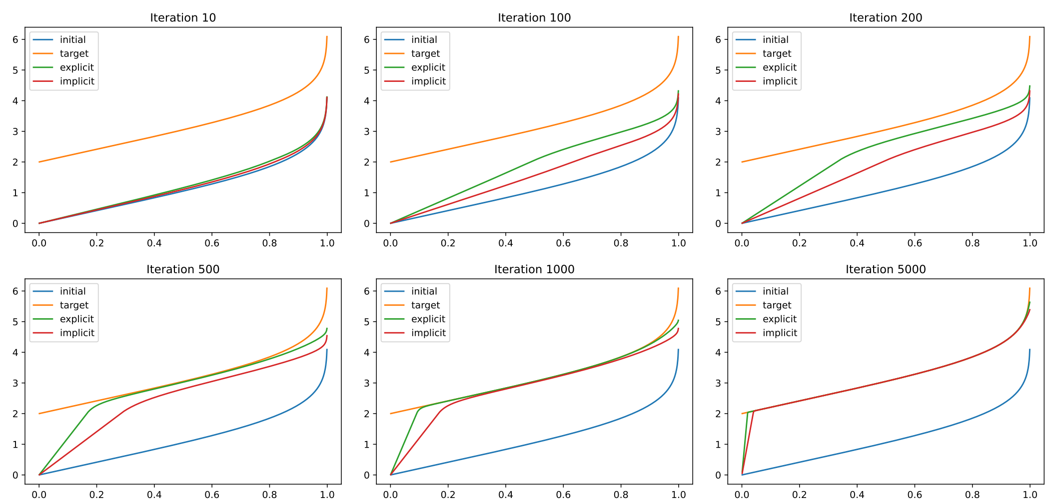

Figure 20: Comparison of the quantile functions belonging to the implicit (red) and explicit (green) Euler schemes between $\mu_0 \sim \mathcal{FN}(0,1)$ and $\nu \sim \mathcal{FN}(2,1)$ with $\tau = \frac{1}{100}$. For the corresponding densities, see Figure 13.

and its inverse is given by $S_\tau(\cdot - \tau)$, where

$$S_\tau\colon\ \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \begin{cases} x+\tau, & \text{if } x < -\tau, \\ 0, & \text{if } -\tau \le x \le \tau, \\ x-\tau, & \text{if } x > \tau, \end{cases}$$

is the soft-shrinkage operator with threshold $\tau > 0$.

Thus the implicit Euler scheme on quantile functions is given by the simple update

$$g_{n+1}(s) = S_\tau(g_n(s) + 2\tau s - \tau).$$

For a uniform initial distribution and for a Gaussian initial distribution, these iterates are displayed in Figure 21.

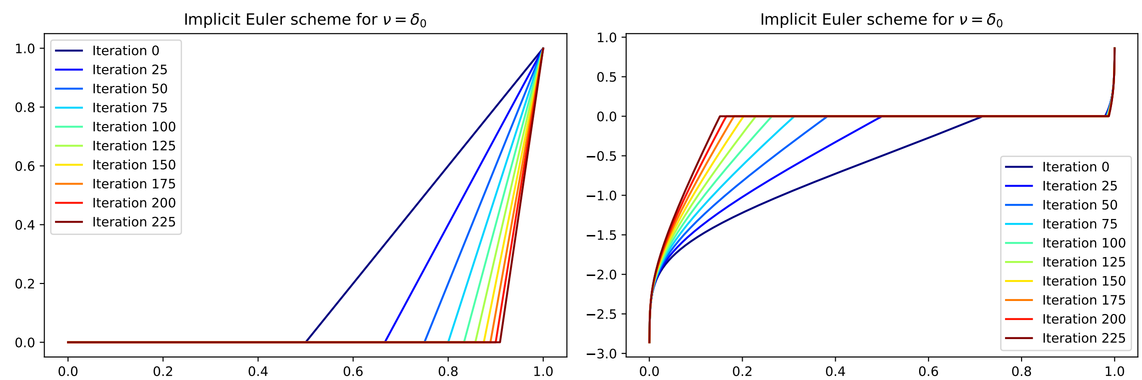

Figure 21: Iterates (quantile functions) of the implicit Euler scheme (9) with $\nu = \delta_0$ and $\tau = \frac{1}{100}$ starting in $\mu_0 \sim \mathcal{U}[0,1]$ (left) and $\mu_0 \sim \mathcal{N}(-1, 0.5)$ (right).