



HAL
open science

On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy

Siwan Boufadène, François-Xavier Vialard

► **To cite this version:**

Siwan Boufadène, François-Xavier Vialard. On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy. 2024. hal-04282762v2

HAL Id: hal-04282762

<https://hal.science/hal-04282762v2>

Preprint submitted on 26 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON THE GLOBAL CONVERGENCE OF WASSERSTEIN GRADIENT FLOW OF THE COULOMB DISCREPANCY.

SIWAN BOUFADENE[†] AND FRANÇOIS-XAVIER VIALARD[†]

ABSTRACT. In this work, we study the Wasserstein gradient flow of the Riesz energy towards a determined target measure on the space of probability measures. The Riesz energy is a quadratic functional on this space, defined using Riesz kernels, and it is in general not geodesically convex in the Wasserstein geometry. Consequently, standard arguments cannot be applied to deduce the global convergence of the Wasserstein gradient flow. Our main result is the exponential convergence of the flow to the minimizer on a closed Riemannian manifold under the condition that the logarithms of the source and target measures are bounded and Hölder continuous. To show this, we first prove that a Polyak-Lojasiewicz inequality is satisfied for sufficiently regular solutions. The key regularity result is the global-in-time existence of Hölder solutions if the initial and target data are Hölder continuous, proven either in Euclidean spaces or in closed Riemannian manifolds. We then define Lagrangian critical points and prove that such points, for the Coulomb or Energy distance discrepancies, are equal to the target everywhere except on singular sets with empty interiors. For arbitrary measures, we use flow interchange techniques to prove there are no local minima other than the global one for the Coulomb kernel. Additionally, sufficiently singular measures cannot be critical points, ensuring they are not fixed points of the discrete JKO updates.

Keywords. Calculus of variations · Optimal transportation · Coulomb kernel · Flow interchange

Mathematics Subject Classification. 49Q22, 49Q10, 90C26.

CONTENTS

1	Introduction	3
2	Polyak-Lojasiewicz Inequality and Exponential Convergence	5
2.1	Regularity and Polyak-Lojasiewicz Inequality for the Energy Functional	6
2.2	Exponential Convergence for Globally Regular Data	7
2.3	Well-Posedness in Hölder Spaces	8
2.3.1.	On the Euclidean Space	8
2.3.2.	On a Compact Riemannian Manifold	12
3	Critical Points for the Wasserstein Flow	14
3.1	Critical Points and Lagrangian Critical Points for MMD Wasserstein Gradient Flows	14
3.2	Characterization of Lagrangian Critical points for MMD Wasserstein Gradient Flows	15
4	No Local Minima in the Wasserstein Geometry	17
4.1	Heat Diffusion Perturbation	18
4.2	Estimates of Mass Transfers in the Diffusion Process	18
4.3	Proof of Theorem 4.1	19
4.4	Dimension of Measure and Critical Points	19
4.5	Extension to Riemannian Manifolds	21
	Acknowledgements	23
	References	23
	Appendix A Wasserstein Gradient Flows	26
	Appendix B Kernels, MMD and Potential Theory	28
B.1	In the Euclidean Space	29
B.2	Extension to Riemannian Manifolds	30
	Appendix C Some Proofs	30

1. INTRODUCTION

In this paper, we are interested in the Wasserstein gradient flow of some quadratic functionals defined on the set of probability measures, both on the Euclidean space and on a Riemannian manifold. These quadratic functionals are convex for the standard vertical convex structure of probability measures. However they are not geodesically convex for the Wasserstein geometry, potentially impacting the global convergence properties typically obtained in geodesically convex scenarios. The motivation for studying these gradient flows under the Wasserstein geometry comes from machine learning, more precisely from the mean-field limit of shallow neural networks [CB18; MMN18]. This line of research investigates the optimization landscape of the usual empirical risk of a single-hidden layer neural network under gradient flow. A powerful relaxation of the problem, already proposed in [Bar93], consists of embedding the space of parameters into the space of probability measures. In this context, the corresponding objective functional takes on a quadratic form and the particle gradient flow corresponds to a Wasserstein gradient flow. The Coulomb MMD energy is a particular case of quadratic functionals defined by reproducing kernels. These functionals have also raised interest in machine learning and statistics since they yield a discrepancy between probability measures and, in fact, a squared distance. These Maximum Mean Discrepancies (MMD) possess two noteworthy properties: (i) a quadratic computational complexity (or even $O(n \log(n))$ for the Energy distance kernel [Her+23]), outperforming alternatives like optimal transport, and (ii) a parametric estimation rate from empirical measures, a feature missing from standard optimal transport methods.

Let G be a (conditionally positive) kernel, such as the Gaussian kernel, on the Euclidean space \mathbb{R}^d . The MMD between μ and ν two probability measures is the energy

$$E_\nu(\mu) = \iint (\mu(x) - \nu(x))G(x, y)(\mu(y) - \nu(y)).$$

Such a functional is nonnegative and strictly convex on the space of probability measures if the kernel G is conditionally positive. In our work, we are interested in a fixed target measure ν and a time-dependent μ_t , optimized through the action of velocity fields. More precisely, we are interested in the Wasserstein gradient flow of the functional E_ν with respect to μ . Our primary concern is the question of global convergence towards the unique minimizer, which is ν . For example, when the kernel is smooth enough, empirical measures remain preserved by the Wasserstein gradient flow. Consequently, the Wasserstein gradient flow of this energy with a finite empirical measure as the source and a density as the target cannot give convergence. This fact motivates exploring non-smooth kernels, such as the energy distance $(x, y) \mapsto -\|x - y\|$ on the Euclidean space, which is not C^1 . Consequently, one can hope for global convergence of the Wasserstein gradient flow even when the source measure and the target measure are mutually singular. Indeed, in dimension one, the corresponding functional is geodesically convex in the Wasserstein geometry, implying global convergence of the solution μ_t towards ν . This specific kernel has been studied in the context of Wasserstein gradient flows in [Her+24] and [Hag+23] for applications in machine learning and imaging. The authors explicitly leave as open the question of global convergence.

The question we address in this paper is the extension of this one-dimensional result to higher dimensions. There are at least two different directions for generalizing this result. Firstly, the Energy distance kernel $-\|x - y\|$ remains conditionally positive definite on \mathbb{R}^d for $d \geq 1$. Secondly, in dimension one, $-|x - y|$ is the Coulomb kernel, proportional to the inverse of the Laplacian operator. The inverse of the Laplacian can also be defined in higher dimensions, for instance on \mathbb{R}^3 , it is given by $\frac{1}{\|x - y\|}$. Both kernels belong to the family of Riesz kernels. One motivation for using these kernels also stems from numerical experiments, where the energy distance notably stands out. Indeed, the energy distance kernel is easy to implement. An efficient and fast method for calculating the discrepancy associated with it, as described in [Her+23], allows for large-scale applications. Furthermore, it exhibits favorable behavior compared to other kernels such as the Gaussian kernel. More precisely, global convergence is observed. Conversely, the Coulomb kernel is more intricate to implement due to its blow-up along the diagonal. This makes drawing conclusions from numerical experiments more delicate. Yet, from a theoretical point of view, the Coulomb kernel has been studied intensively in the mathematical

literature, in particular due to its physical significance [Ser16; Ser15]. Recent results presented in [JW18] and [CRS23] study large stochastic systems of interacting particles under Coulomb interaction in the Euclidean space, proving propagation of chaos and convergence to the limit continuity equation using relative entropy methods. In [CRS23] this work is done on the torus manifold.

For the Coulomb kernel, given smoothness assumptions on μ a time-dependent density and ν a fixed density, the Wasserstein gradient flow associated with E_ν reads, on \mathbb{R}^d or on a Riemannian manifold:

$$\partial_t \mu = -\nabla \cdot (\mu \nabla \varphi), \quad \Delta \varphi = \mu - \nu. \quad (1.1)$$

The potential φ is a solution to the Poisson equation with source term $\mu - \nu$. The potential φ can be expressed, up to a positive constant, as $\varphi = -G \star (\mu - \nu)$, giving an example of non-linear non-local interactions. Non-local interaction energy systems associated with a radial kernel $W(x) = w(\|x\|)$ are solutions to the equation:

$$\frac{\partial \mu}{\partial t} = -\nabla \cdot (\mu (\nabla W \star \mu)).$$

Confinement results for these dynamics have been established, depending on the choice of W . In many instances, W is assumed to be λ -convex, preserving particles as discussed in [Car+11], allowing for the use of mean-field techniques. Attractive potentials—e.g. $w'(r) \geq 0$ everywhere—are the simplest ones, where in some cases, the total mass aggregates at the potential's mass center. Some other potentials, including swarming systems models [CCH14], Morse potentials or characteristic function of sets [BCY14; Car+12; Car+11] are said to be attractive-repulsive. Various hypotheses are taken to establish confinement [Car+11; BCY14]. Some potentials with a singularity at 0 have been treated, for example in [BCY14] with $W(x) = G(x) + W_a(x)$ where G is the Coulomb kernel $G(x) = \|x\|^{-d+2}$ and W_a is an attractive potential that satisfies $\lim_{r \rightarrow \infty} W'_a(r)r^{1/d} = +\infty$. However, these results do not apply to our case of study, since our functional is not λ -convex and has diffusive properties. Additionally, the confining part, which depends on a target measure ν , is weaker compared to previously cited papers and, as of now, we were not able to prove mass confinement.

Main contributions. This paper presents two primary results focusing on the Coulomb kernel.

The first main result is the global convergence of the solution of the continuity equation 2.4 towards the target ν , on a closed Riemannian manifold within a smooth initial setting. At the beginning of Section 2, assuming the solution is sufficiently smooth and exists at all times, a simple calculation shows that the Polyak-Lojasiewicz inequality holds at all times on a closed Riemannian manifold. Notably, this observation reduces the question of (exponential) global convergence to a regularity inquiry: whether the solutions exist for all time in adequately smooth functional spaces. Therefore, we first study the gradient flow of the Coulomb discrepancy in a smooth setting, i.e. the corresponding continuity PDE under certain regularity assumptions. Assuming Hölder continuity for initial and target measures, we establish that solutions exist at all times, with propagation of Hölder regularity. This proves global convergence with an exponential rate of convergence in the case of closed Riemannian manifolds.

The second main result is Theorem 4.1, which relates to the landscape of the Coulomb discrepancy on both the Euclidean space and closed Riemannian manifolds. It states that, in the Wasserstein geometry, the energy functional has no local minima apart from the global one. More precisely, we prove that if the current measure differs from the target, there always exists a measure curve (starting at the current one) that is 1/2-Hölder in Wasserstein along which the energy is strictly decreasing. This is done through the use of flow interchange techniques, specifically leveraging properties of the Boltzmann entropy along the flow.

Among other results, we prove in Section 3 that any Lagrangian critical point for the Coulomb and Energy distance kernels is equal to the target measure everywhere except on singular sets with empty interiors. In a similar direction, we prove that if the difference between the current measure and the target has a Minkowski dimension smaller than the ambient one, then it cannot be a critical point of the Wasserstein gradient flow.

Perspectives. Left open by our work is the question of global convergence of the flow in a closed Riemannian manifold for all source and target measures. Although the result seems highly plausible on a closed Riemannian manifold, extending it to a non-compact setting such as the Euclidean space would require addressing the confinement issue or changing completely the proof strategy. Indeed, there is a competition between the repulsive behavior of the Coulomb kernel and the attraction of the target measure. The repulsive part can lead to mass spreading to infinity in the Euclidean space, which makes the analysis more difficult in our opinion. Note also that in a finite-dimensional setting, knowing that there is no local minima but the global one guarantees the global convergence for almost every initial condition under some assumptions on the objective functional. Obtaining similar results in our infinite-dimensional case would be of interest.

Notations.

- If A is a subset of \mathbb{R}^d , $A^c := \mathbb{R}^d \setminus A$ denotes its complement.
- $\mu^{\otimes 2}$ is the product measure on $\mathbb{R}^d \times \mathbb{R}^d$: for any Borel sets A, B , $\mu^{\otimes 2}(A \times B) := \mu(A)\mu(B)$.
- C_c^∞ is the space of test functions, i.e. infinitely differentiable functions with compact support.
- Convolution is denoted by \star .
- If (M, g) is a Riemannian manifold, $d_M(x, y)$ denotes the geodesic distance between $x, y \in M$.

2. POLYAK-LOJASIEWICZ INEQUALITY AND EXPONENTIAL CONVERGENCE

Our goal is to prove that a Wasserstein gradient flow curve μ_t of the energy E_ν converges to the target ν . In finite dimensions, a standard condition for convergence is the Polyak-Lojasiewicz inequality.

Definition 2.1 (Polyak-Lojasiewicz condition). Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function. It is said to satisfy the Polyak-Lojasiewicz condition (PL condition) with parameter $\lambda > 0$ if the following inequality holds for all $x \in \mathbb{R}^d$:

$$\frac{1}{2} |\nabla f(x)|^2 \geq \lambda(f(x) - f^*).$$

This condition is weaker than many other classical conditions, including strong convexity, weak strong convexity, or the restricted secant inequality, see [KNS16] for a review. With this condition, an exponential convergence rate to the global minimum can be proven. Moreover, we can use a weaker dynamical version assumption. When examining the convergence of a gradient flow curve x_t in the Euclidean space, defined by $\dot{x} = -\nabla f(x)$, it is sufficient for this inequality to hold along the curve. In the subsequent discussion, we focus on localized PL inequalities along specific curves, as opposed to global ones.

For gradient flows on measures, this type of inequality is also known under the name of *entropy-entropy production* inequality in the context of gradient flow [KMV+16]. In general, such inequalities are functional inequalities. For instance, the log-Sobolev inequality is a PL inequality for the entropy under the Wasserstein geometry. Similarly, in [KMV+16], the authors use such a generalized Beckner inequality to obtain PL. To define an analog inequality in Wasserstein spaces we need the following chain rule [AGS05, Proposition 10.3.18].

Proposition 2.2. *Let \mathcal{F} be a proper lower semicontinuous functional and μ_t be an absolutely continuous curve with tangent velocity v_t . We suppose $\mathcal{F} \circ \mu$ is approximately differentiable in time almost everywhere and that for all t the set $\partial\mathcal{F}(\mu_t)$ of vector field subdifferentials $\xi \in L_2(\mu_t)$ (see Definition A.5) is non-empty. Then for any $\xi_t \in \partial\mathcal{F}(\mu_t)$ we have:*

$$\frac{d}{dt} \mathcal{F}(\mu_t) = \int v_t \cdot \xi_t d\mu_t.$$

For functionals that are regular enough, we have $v_t = -\nabla \frac{\delta\mathcal{F}}{\delta\mu}(\mu_t)$ and $\nabla \frac{\delta\mathcal{F}}{\delta\mu}(\mu_t) \in \partial\mathcal{F}(\mu_t)$ so that:

$$\frac{d}{dt} \mathcal{F}(\mu_t) = - \left\| \nabla \frac{\delta\mathcal{F}}{\delta\mu}(\mu_t) \right\|_{L_2(\mu_t)}^2. \quad (2.1)$$

This motivates our definition of the Polyak-Lojasiewicz in Wasserstein spaces.

Definition 2.3. Let \mathcal{F} be a functional as in Proposition 2.2. Suppose that its global minimum is equal to 0. Let μ_t be an absolutely continuous curve. The functional \mathcal{F} is said to satisfy a Polyak-Lojasiewicz inequality with parameter $\lambda > 0$ along the curve μ_t if for all $t > 0$:

$$\left\| \nabla \frac{\delta \mathcal{F}}{\delta \mu}(\mu_t) \right\|_{L_2(\mu_t)}^2 \geq \lambda \mathcal{F}(\mu_t). \quad (2.2)$$

If this inequality holds, we get, for regular enough functionals, thanks to equation (2.1) and Gronwall's lemma:

$$\mathcal{F}(\mu_t) \leq \mathcal{F}(\mu_0) e^{-\lambda t},$$

proving global exponential convergence.

2.1. Regularity and Polyak-Lojasiewicz Inequality for the Energy Functional. In this section, we prove a *partial local* Polyak-Lojasiewicz inequality, in the sense that the constant λ in 2.2 does depend on μ . It is notably weaker than a global PL inequality. The issue is that this inequality is only *partial and local*, in the sense that it depends on a positive lower bound on μ that may not exist. In finite dimensions, obtaining a local Polyak-Lojasiewicz inequality (with a λ that may depend on a given neighborhood) is still informative on the landscape of the energy functional. Specifically, it implies that there are no critical points other than global minima. However, in our case, since this inequality doesn't hold for every μ , a similar conclusion cannot be directly derived.

When the subdifferential of E_ν is non-empty, we can characterize it, using the same ideas as in [Bon13, Prop 4.3.1]. The proof is essentially the same for both Energy Distance in the Euclidean space and Coulomb kernels both in the Euclidean space and on Riemannian manifolds, but crucial parts about avoiding the singularity rely on different arguments (see Appendix C).

Lemma 2.4. *Let μ be a probability measure such that $\partial E_\nu(\mu)$ is non-empty. Then the vector field $\nabla \frac{\delta E_\nu}{\delta \mu}(\mu) = (\nabla G) \star (\mu - \nu)$ satisfies*

$$\left\| \nabla \frac{\delta E_\nu}{\delta \mu}(\mu) \right\|_{L_2(\mu)} \leq |\partial E_\nu|(\mu).$$

Then we need a regularity result about our functional, which is proven in Appendix C.

Lemma 2.5. *Let $\mu, \nu \in \mathcal{P}_2^r$ be density measures regarding the Lebesgue measure or the volume measure on a manifold. Then the functional E_ν has a non-empty subdifferential, in the sense of Definition A.6. Moreover, $\nabla \frac{\delta E_\nu}{\delta \mu}(\mu) \in \partial E_\nu(\mu)$.*

Now, combining these two lemmas, the element of minimal norm of $\partial \mathcal{F}(\mu_t)$ is precisely the time-dependent vector field $v_t = \nabla \frac{\delta E_\nu}{\delta \mu}(\mu_t)$, which drives mass transfer along time for the curve μ_t . In other words, the following property holds.

Proposition 2.6. *Let μ_t be a solution of equation (1.1) in a weak sense. We suppose $\mu_t \in \mathcal{P}_2^r$ at all times. Then it is a gradient flow of the functional E_ν starting from μ_0 , in the sense of Definition A.7.*

Then we can formulate a condition that implies a *local* Polyak-Lojasiewicz inequality in a compact manifold.

Proposition 2.7 (Local Polyak-Lojasiewicz inequality). *Let (M, g) be a closed Riemannian manifold and μ, ν be two measures with density w.r.t. the volume measure on M , such that $\log(\mu)$ is bounded below. Then, it holds:*

$$E_\nu(\mu) \leq \frac{1}{\underline{\mu}} \left\| \nabla \frac{\delta E_\nu}{\delta \mu}(\mu) \right\|_{L_2(\mu)}^2, \quad (2.3)$$

where $\underline{\mu}$ is a lower bound for μ on M .

Proof. The proof is straightforward since inequality 2.3 is exactly:

$$\int_M |\nabla \varphi_{\mu-\nu}(x)|^2 d\text{vol}(x) \leq \frac{1}{\underline{\mu}} \int_M |\nabla \varphi_{\mu-\nu}(x)|^2 d\mu(x),$$

where vol is the volume measure on M . The inequality follows from $\underline{\mu} d\text{vol} \leq \mu$. \square

However, in a non-compact setting, a probability measure such that $\log(\mu)$ is bounded from below does not exist. The question of formulating a condition implying a Polyak-Lojasiewicz inequality in the Euclidean space therefore remains an open question.

A crucial point in proving global convergence of a gradient flow that satisfies a local and partial PL inequality is to ensure that the constant $\lambda(\mu(t))$ is well-defined and does not deteriorate too rapidly along the curve $\mu(t)$. In the next section, we show that λ can be taken constant if the flow is sufficiently regular for all time.

2.2. Exponential Convergence for Globally Regular Data. First, we begin with a stability result, where the regularity of the density μ_t over time implies a global Polyak-Lojasiewicz inequality.

Proposition 2.8 (Stability of the Polyak-Lojasiewicz condition). *Let μ_0, ν be two \mathcal{C}^1 densities on M such that $\log(\mu_0)$ and $\log(\nu)$ are bounded. Then if the associated equation (1.1) admits a continuous density, it satisfies*

$$\min(\min \mu_0, \min \nu) \leq \mu_t(x) \leq \max(\max \mu_0, \max \nu).$$

Sketch of proof. We use the regularity and an optimality argument. Let us define $\underline{x}(t) \in \arg \min \mu_t(x)$ and $\bar{x}(t) \in \arg \max \mu_t(x)$. Since the manifold is closed and μ_t is \mathcal{C}^1 we have:

$$\nabla \mu_t(\underline{x}(t)) = \nabla \mu_t(\bar{x}(t)) = 0.$$

Moreover, using the regularity we obtain:

$$\begin{aligned} \partial_t \mu_t &= -\nabla \cdot (\mu_t v_t) \\ &= -v_t \cdot \nabla \mu_t - \mu_t \nabla \cdot v_t \\ \partial_t \mu_t &= -v_t \cdot \nabla \mu_t - \mu_t (\mu_t - \nu). \end{aligned}$$

Instantiating the previous inequality at $\underline{x}(t)$ for example we get:

$$\partial_t \mu_t(\underline{x}(t)) = (\min \mu_t) \nu(\underline{x}(t)) - (\min \mu_t)^2.$$

After a brief study of the phase diagram (see the proof of Lemma 2.12) we obtain our result. A similar argument applies to $\bar{x}(t)$. \square

This stability of the PL condition allows us to formulate a convergence result, if we make strong assumptions on the global existence and the regularity of $\mu(t, x)$.

Proposition 2.9 (Exponential convergence). *Let μ, ν be two measures with continuous densities and bounded logarithms on a closed Riemannian manifold M . If the density of the curve μ_t generated by the gradient flow of E_ν is $\mathcal{C}_{t,x}^1$ in both time and space for all $t \in \mathbb{R}_{\geq 0}$ and $x \in M$, then exponential convergence of μ_t to ν holds in the following two ways:*

$$\begin{cases} E_\nu(\mu_t) \leq E_\nu(\mu_0) e^{-\lambda t} \\ W_2^2(\mu_t, \nu) \leq \frac{4}{\lambda} E_\nu(\mu_0) e^{-\lambda t}, \end{cases}$$

where $\lambda = \min(\underline{\mu}_0, \nu)$.

Proof. The proof is a straightforward application of the Polyak-Lojasiewicz inequality. \square

In Theorem 2.23, we are able to relax the assumption on the regularity of the solution to Hölder continuous solutions. Therefore, we have just reduced the global convergence problem to a global existence and regularity problem. In the next section, we prove the global existence of Hölder continuous solutions if the initial and target densities are Hölder continuous. Our convergence result is then stated in Theorem 2.23.

2.3. Well-Posedness in Hölder Spaces. In this section, we consider solutions of the PDE:

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \\ v_t := -\nabla G \star (\mu_t - \nu). \end{cases} \quad (2.4)$$

The velocity vector field v_t satisfies $\nabla \cdot v_t = \mu_t - \nu$. Let us consider a solution μ_t of equation (2.4) on an open interval $]0, T[$. According to Definition A.2, it is an absolutely continuous curve associated with the time-dependent vector field $v_t := -\text{grad } \varphi$, where $\Delta \varphi_t = \mu_t - \nu$. If μ_0 has a density, μ_t also has a density at least for short times.

2.3.1. On the Euclidean Space. This subsection is devoted to the proof of the following theorem:

Theorem 2.10. *Let μ_0 and ν be Hölder continuous densities with compact support. Then, equation (1.1) admits a unique global solution μ_t that is Hölder continuous at all times.*

We use techniques from [BLL12] and adapt them to our case. The standard approach involves rewriting the problem in Lagrangian coordinates, following the particle flow, and leveraging ODE results to control the norm of the density.

Let us set the notations. We denote G as the Coulomb kernel, which is repulsive and a solution of $\Delta G = -\delta$. We define the particle flow ψ_t by $\psi_0 = \text{Id}$ and

$$\frac{d}{dt} \psi_t = v_t \circ \psi_t. \quad (2.5)$$

Let us consider the evolution of the time-dependent function $f_t := \mu_t \circ \psi_t$. As $\mu_t = f_t \circ \psi_t^{-1}$, the existence of μ_t is linked to the existence of f_t and ψ_t . This dependence will be made precise in Lemma 2.13. We can rewrite the particle flow equation (2.5), for $\alpha \in \mathbb{R}^d$:

$$\begin{aligned} \frac{d}{dt} \psi_t(\alpha) &= v_t \circ \psi_t(\alpha) = - \int \nabla G(\psi_t(\alpha) - y) (\mu_t - \nu)(y) dy \\ \frac{d}{dt} \psi_t(\alpha) &= - \int \nabla G(\psi_t(\alpha) - \psi_t(\alpha')) \det(d\psi_t(\alpha')) \mu_t(\psi_t(\alpha')) d\alpha' + \int \nabla G(\psi_t(\alpha) - y) \nu(y) dy. \end{aligned}$$

Let us note $J_t(\alpha) := \det(d\psi_t(\alpha))$ and remark:

$$\frac{d}{dt} J_t(\alpha) = J_t(\alpha) (\mu_t - \nu) \circ \psi_t(\alpha),$$

so that:

$$\frac{d}{dt} [t \mapsto J_t(\alpha) \mu_t(\psi_t(\alpha))] = 0, \quad (2.6)$$

and as $J_0(\alpha) = 1$ we get $J_t(\alpha) \mu_t(\psi_t(\alpha)) = \mu_0(\alpha)$. We end up with the Lagrangian formulation of the flow

$$\frac{d}{dt} \psi_t(\alpha) = - \int \nabla G(\psi_t(\alpha) - \psi_t(\alpha')) \mu_0(\alpha) d\alpha + \int \nabla G(\psi_t(\alpha) - y) \nu(y) dy = F(\psi_t(\alpha)). \quad (2.7)$$

We solve this equation in the Banach space \mathcal{B} defined by $\mathcal{B} := \{ \psi : \mathbb{R}^d \rightarrow \mathbb{R}^d \mid \|\psi\|_{1,\gamma} < \infty \}$, where $\|\psi\|_{1,\gamma} := |\psi(0)| + \|d\psi\|_\infty + |d\psi|_\gamma$ with $|\cdot|_\gamma$ the Hölder semi-norm. Now we can state our first local existence result.

Proposition 2.11. *Let μ_0 and ν be density measures in \mathcal{B} with compact support. Then equation (2.7) with initial condition $\psi_0 = \text{Id}$ admits a unique solution on a maximal time interval $[0, T[$. Either T is infinite or the Banach norm $\|\psi_t\|_{1,\gamma}$ blows up as $t \rightarrow T$.*

Proof. As F is Lipschitz [BLL12] and [MB01, Chap 4]) the Picard theorem in Banach spaces proves the result. \square

Now let us prove Theorem 2.10, that is $T = +\infty$ in the preceding proposition. To show this we suppose $T < \infty$ in the whole following discussion and show that $\|\psi_t\|_{1,\gamma}$ is bounded uniformly in time on the interval $[0, T[$. In [BLL12], an explicit formula is found for f_t , which cannot be done in our case. However, we can control the evolution of f_t .

Lemma 2.12. *The quantity $\|f_t\|_\infty$ is uniformly bounded on $[0, T[$.*

Proof. We can write:

$$\nabla \cdot (v_t \mu_t) = \mu_t \nabla \cdot v_t + v_t \cdot \nabla \mu_t = \mu_t^2 - \mu_t \nu + v_t \cdot \nabla \mu_t,$$

so that we get:

$$\begin{aligned} \frac{d}{dt} f_t(\alpha) &= \partial_t \mu_t(\psi_t(\alpha)) + \nabla \mu_t(\psi_t(\alpha)) \cdot v_t(\psi_t(\alpha)) \\ &= -\nabla \cdot (v_t \mu_t)(\psi_t(\alpha)) + \nabla \mu_t(\psi_t(\alpha)) \cdot v_t(\psi_t(\alpha)) \\ \frac{d}{dt} f_t(\alpha) &= f_t(\alpha) \nu(\psi_t(\alpha)) - f_t^2(\alpha). \end{aligned} \quad (2.8)$$

Although $\nu \circ \psi_t$ is of course not constant, this equation, which resembles a logistic equation, is well-behaved. We can study its phase diagram. First, if f_0 is a positive function then f_t will be positive too. Moreover, if $f_t(\alpha) > \|\nu\|_\infty$ then $\frac{d}{dt} f_t(\alpha)$ is negative, so $t \mapsto f_t(\alpha)$ is locally decreasing. This shows that for all $\alpha \in \mathbb{R}^d$:

$$\min(\min f_0, \min \nu) \leq f_t(\alpha) \leq \max(\|f_0\|_\infty, \|\nu\|_\infty).$$

□

This control allows us to bound the Holder norm by a quantity that depends on ψ_t .

Lemma 2.13. *Let μ_t be a solution defined as in Proposition 2.11, where $\nu \in L_\infty$ satisfies $|\nu|_\gamma < \infty$. Then:*

$$|\mu_t|_\gamma \leq C \|d\psi_t\|_\infty^\gamma \left(\int_0^t (1 + \|d\psi_s^{-1}\|_\infty^\gamma) ds \right),$$

for some positive constant $C > 0$.

Proof. As $\mu_t = f_t \circ \psi_t^{-1}$, using that $|f \circ g|_\gamma \leq |f|_\gamma \|dg\|_\infty^\gamma$ for general functions f, g , we get the first estimate:

$$|\mu_t|_\gamma \leq |f_t|_\gamma \|d\psi_t^{-1}\|_\infty^\gamma.$$

Now, to control $|f_t|_\gamma$, we can take the Holder semi norm of equation (2.8), using the fact that ν is bounded and Holder continuous, that $\|f_t\|_\infty$ is bounded, and the property $|fg|_\gamma \leq |f|_\gamma \|g\|_\infty + \|f\|_\infty |g|_\gamma$ to get:

$$\begin{aligned} \frac{d}{dt} |f_t|_\gamma &\leq |f_t \cdot \nu \circ \psi_t| + |f_t^2|_\gamma \\ &\leq |f_t|_\gamma \|\nu \circ \psi_t\|_\infty + \|f_t\|_\infty |\nu \circ \psi_t|_\gamma + 2\|f_t\|_\infty |f_t|_\gamma \\ &\leq (\|\nu\|_\infty + 2\|f_t\|_\infty) |f_t|_\gamma + \|f_t\|_\infty |\nu|_\gamma \|d\psi_t\|_\infty^\gamma \\ \frac{d}{dt} |f_t|_\gamma &\leq C_1 |f_t|_\gamma + C_2 \|d\psi_t\|_\infty^\gamma, \end{aligned}$$

for some positive constants C_1, C_2 . Then we apply Gronwall's lemma with time-dependent terms, stating that if y is differentiable and a, b are continuous functions such that $\dot{y} \leq ay + b$, then $y(t) \leq y(0)e^{\int_0^t a(s)ds} + \int_0^t b(s)e^{\int_0^s a(u)du - \int_0^s a(u)du} ds$. We obtain, as $t < T < \infty$:

$$|f_t|_\gamma \leq |f_0|_\gamma \exp(C_1 t) + \int_0^t C_2 \|d\psi_s\|_\infty^\gamma \exp(C_1(t-s)) ds \leq C \left(1 + \int_0^t \|d\psi_s\|_\infty^\gamma ds \right),$$

ending the proof. □

Differentiating the particle equation (2.5) and taking the L_∞ norm we get:

$$\frac{d}{dt} \|d\psi_t\|_\infty \leq \|dv_t\|_\infty \|d\psi_t\|_\infty.$$

This gives by Gronwall's lemma:

$$\|d\psi_t\|_\infty \leq C \exp \int_0^t \|dv_s\|_\infty ds. \quad (2.9)$$

Let us observe that ψ_t^{-1} satisfies a similar bound.

Lemma 2.14. *Let ψ_t^{-1} be defined as the inverse flow of ψ_t . Then the majoration above is true for ψ_t^{-1} , i.e. for some constant $C > 0$:*

$$\|d\psi_t^{-1}\|_\infty \leq C \exp \int_0^t \|dv_s\|_\infty ds. \quad (2.10)$$

Proof. Let $\alpha := \psi_s(y)$ for some $y \in \mathbb{R}^d$ and $s > 0$. Then if $0 \leq t < s$, we have $\alpha \in \psi_t(\mathbb{R}^d)$, as $\alpha = \psi_t \circ \psi_{s-t}(y)$ by semi-group property. We get $\psi_t^{-1}(\alpha) = \psi_{s-t}(y) = \psi_{s-t} \circ \psi_s^{-1}(\alpha)$. Differentiating both sides of the equation in time we get:

$$\partial_t \psi_t^{-1}(\alpha) = \partial_t \psi_{s-t} \circ \psi_s^{-1}(\alpha) = -v_{s-t} \circ \psi_{s-t} \circ \psi_s^{-1}(\alpha) = -v_{s-t} \circ \psi_t^{-1}(\alpha).$$

This leads to the inequality:

$$\frac{d}{dt} \|d\psi_t^{-1}\|_\infty \leq \|dv_{s-t}\|_\infty \|d\psi_t^{-1}\|_\infty,$$

so that by Grönwall lemma:

$$\|d\psi_t^{-1}\|_\infty \leq C \exp \left(\int_0^t \|dv_{s-u}\|_\infty du \right).$$

We can choose $s = t$ in the previous inequality to get:

$$\|d\psi_t^{-1}\|_\infty \leq C \exp \left(\int_0^t \|dv_u\|_\infty du \right). \quad (2.11)$$

□

With this result, we managed to bound from above all of our quantities by functions of dv_t .

On the derivative of our velocity field. The following arguments are presented in [MB01, Sections 2.4.2, 3]. The kernel $G_2 := d_x \text{grad} G$ is homogeneous of degree -N. Due to this, the singularity at the diagonal cannot be integrated. However, it has mean-value zero and defines a singular integral operator through the convolution:

$$G_2 \star f(x) = PV \int G_2(x, y) f(y) dy := \lim_{\varepsilon \rightarrow 0} \int_{d(x, y) > \varepsilon} G_2(x, y) f(y) dy.$$

Using the same arguments than in [MB01, Prop 2.20], we know that for a velocity field defined by $\text{grad} G \star f$ with G the Coulomb kernel and $f \in \mathcal{C}^\gamma(\mathbb{R}^d; \mathbb{R}^d)$ we have:

$$dv_t(x) = PV \int G_2(x, y) f(y) dy.$$

The following result, found in [MB01, Lemma 4.5 and 4.6] and [BLL12, Lemma 2.2], is expressed in terms of principal value integral in [BLL12] but used in the following form.

Lemma 2.15. *Let $f \in \mathcal{C}^\gamma(\mathbb{R}^d; \mathbb{R}^d)$ be a compactly supported function in a ball of radius R . We define $v := \nabla G \star f$. Then, for some positive constant C independent of f and R we have:*

$$\begin{aligned} \|dv\|_\infty &\leq C \left[|f|_\gamma \varepsilon^\gamma + \max \left(1; \log \left(\frac{R}{\varepsilon} \right) \right) \|f\|_\infty \right]; \forall \varepsilon > 0, \\ |dv|_\gamma &\leq C |f|_\gamma. \end{aligned}$$

To use it we need to confine the support of μ_t . First, we can bound the velocity field v_t .

Lemma 2.16. *Let μ_t be a solution as defined in Proposition 2.11 and v_t the associated velocity field. Then for some constant $C > 0$ only depending on the dimension:*

$$\|v_t\|_\infty \leq C (\|\mu_t - \nu\|_\infty + \|\mu_t - \nu\|_1).$$

Proof. We write for $x \in \mathbb{R}^d$, knowing $|\nabla G(x)|$ is proportional to $|x|^{d-1}$:

$$\begin{aligned} |v_t(x)| &\leq \left[\int_{B(0,1)} + \int_{B(0,1)^c} \right] |\nabla G(x-y)(\mu_t - \nu)(y)| dy \\ &\leq \|\mu_t - \nu\|_\infty \int_{B(0,1)} |\nabla G(x-y)| dy + \int_{B(0,1)^c} |\mu_t - \nu|(y) dy \\ |v_t(x)| &= C(\|\mu_t - \nu\|_\infty + \|\mu_t - \nu\|_1). \end{aligned}$$

□

This allows us to bound the growth of the support.

Lemma 2.17. *Let μ_t be a solution defined as in Proposition 2.11. Suppose that the support of μ_0 is contained in the ball of center 0 and radius $R_0 > 0$. Then there exists a positive constant $C > 0$ such that the support of μ_t is contained in $R(t) := R_0 + Ct$.*

Proof. We use the inequality from Lemma 2.16. As $\|\mu_t\|_\infty = \|f_t\|_\infty \leq C$ by Lemma 2.12, the first term is uniformly bounded in $[0, T]$. As μ_t and ν are probability densities, so is the second term. This proves the result. □

Having proved that μ_t has bounded support, we can apply the first estimate from Lemma 2.15 with $\varepsilon = [\|\mu_t - \nu\|_\infty / |\mu_t - \nu|_\gamma]^{1/\gamma}$ to establish the existence of C independent of t such that:

$$\|dv_t\|_\infty \leq C \left[\|\mu_t - \nu\|_\infty + \max \left(1, \log \left(\frac{R(t) |\mu_t - \nu|_\gamma^{1/\gamma}}{\|\mu_t - \nu\|_\infty^{1/\gamma}} \right) \|\mu_t - \nu\|_\infty \right) \right].$$

Thanks to Lemma 2.12, $\|\mu_t - \nu\|_\infty$ is bounded and by Lemma 2.17, as $T < \infty$ we get the existence of $C_1, C_2 > 0$ such that:

$$\|dv_t\|_\infty \leq C_1 + C_2 \log(|\mu_t - \nu|_\gamma). \quad (2.12)$$

Now, we are ready to prove our first real boundedness result.

Proposition 2.18. *The quantity $\|dv_t\|_\infty$ is uniformly bounded on the time interval $[0, T]$.*

Proof. Injecting the inequality from Lemma 2.13 into equation (2.12), we get the existence of constants such that:

$$\|dv_t\|_\infty \leq A + B \log(\|d\psi_t\|_\infty) + C \log \left(\int_0^t (1 + \|d\psi_s^{-1}\|_\infty^\gamma) ds \right).$$

We write, using inequality (2.10), the fact that $s \leq t$ in the integrals and that $\|dv_t\|_\infty$ is a positive function:

$$\begin{aligned} \log \left(\int_0^t \|d\psi_s^{-1}\|_\infty^\gamma ds \right) &\leq C \log \left(\int_0^t \gamma \exp \left(\int_0^s \|dv_u\|_\infty du \right) ds \right) \\ &\leq C \log \left(\int_0^t \gamma \exp \left(\int_0^t \|dv_u\|_\infty du \right) ds \right) \\ &\leq C \log \left(\gamma t \exp \left(\int_0^t \|dv_u\|_\infty du \right) \right) \\ \log \left(\int_0^t \|d\psi_s^{-1}\|_\infty^\gamma ds \right) &\leq C \left(\log t + \int_0^t \|dv_u\|_\infty du \right), \end{aligned}$$

and we obtain the final differential inequality:

$$\|dv_t\|_\infty \leq C \left(1 + \int_0^t \|dv_s\|_\infty ds + \log(t) \right).$$

Once again, Gronwall's lemma applies to $t \mapsto \|dv_t\|_\infty$ and we get the existence of constants C_1, C_2 such that $\|dv_t\|_\infty \leq C_1 \exp(C_2 t)$, showing the result as $t < T < +\infty$. □

This immediately implies:

Proposition 2.19. *The quantities $\|d\psi_t\|_\infty$, $\|d\psi_t^{-1}\|_\infty$ and $|\mu_t|_\gamma$ are uniformly bounded in time on $[0, T[$.*

Proof. Using the fact that $\|dv_t\|_\infty$ is uniformly bounded in $[0, T[$, inequalities (2.9), (2.10) and Lemma 2.13 show the result. \square

Finally, we can control our last term:

Proposition 2.20. *The quantity $|d\psi_t|_\gamma$ is uniformly bounded in $[0, T[$.*

Proof. Differentiating the particle equation (2.5), taking the Hölder $|\cdot|_\gamma$ semi-norm and applying the preceding proposition along with the second potential theory estimate from Lemma 2.15 we get:

$$\begin{aligned} \frac{d}{dt}|d\psi_t|_\gamma &\leq |d(v_t \circ \psi_t)|_\gamma \|d\psi_t\|_\infty + \|d(v_t \circ \psi_t)\|_\infty |d\psi_t|_\gamma \\ &\leq |dv_t|_\gamma \|d\psi_t\|_\infty^{1+\gamma} + \|dv_t\|_\infty \|d\psi_t\|_\infty |d\psi_t|_\gamma \\ &\leq C_1 |\mu_t|_\gamma \exp\left(\left(1 + \gamma\right) \int_0^t \|dv_s\|_\infty ds\right) + C_2 |d\psi_t|_\gamma \\ \frac{d}{dt}|d\psi_t|_\gamma &\leq C_1 + C_2 |d\psi_t|_\gamma. \end{aligned} \tag{2.13}$$

One last application of Gronwall's lemma ends the proof. \square

We just proved that if $T < \infty$ and a solution of problem (2.7) in the Banach space \mathcal{B} exists in $[0, T[$, then the Banach norm $\|\psi_t\|_{1,\gamma}$ is uniformly bounded on $[0, T[$. If the maximal time of existence T satisfies $T < \infty$, we get a contradiction with the existence result of Proposition 2.11, which states that if $T < \infty$ a finite time blowup of $\|\psi_t\|_{1,\gamma}$ occurs. This proves Theorem 2.10.

2.3.2. On a Compact Riemannian Manifold. This Lagrangian formulation allows us to directly extend our result to a complete closed Riemannian manifold (M, g) . Indeed, taking G the Coulomb kernel on (M, g) , the Lagrangian formulation (2.7) still holds. As Hölder regularity is a local property, the functional F is still locally Lipschitz on the Banach space:

$$\mathcal{B}_M := \{\psi : M \rightarrow TM \mid \|\psi\|_{1,\gamma} < \infty\},$$

where $|\psi(0)| + \|d\psi\|_\infty + |d\psi|_\gamma$ with $\|\cdot\|_{\mathcal{C}_M^{0,\gamma}}$ the Hölder semi-norm on the manifold M , defined by:

$$\|f\|_{\mathcal{C}_M^{0,\gamma}} := \sup_{x,y \in M} \frac{|f(x) - f(y)|}{d_M(x,y)^\gamma}.$$

for scalar functions, and through parallel transport for tensors.

We get the existence of a flow ψ_t , at least locally. The rest of the proof is similar, the main difference being the equivalent of Lemma 2.15 on closed manifolds.

Lemma 2.21. *Let $u \in \mathcal{C}^\gamma(M)$ on a closed manifold M . Consider the equation:*

$$\Delta_M \varphi = u.$$

where u is Hölder continuous. Then, for some positive constant A, C (independent of u) and for all $\varepsilon > 0$ we have the Schauder estimates

$$\begin{aligned} \|\varphi\|_{2,\infty} &\leq C \left(\|u\|_{\mathcal{C}_M^{0,\gamma}} \varepsilon^\gamma + \log\left(\frac{A}{\varepsilon}\right) \|u\|_\infty \right), \\ \|\varphi\|_{\mathcal{C}_M^{2,\gamma}} &\leq C \left(\|u\|_{\mathcal{C}_M^{0,\gamma}} + \|\varphi\|_\infty \right). \end{aligned}$$

Proof of the first inequality. The proof of the first inequality is almost the same as in the Euclidean space. We denote G_2 as the differential (in coordinates) of the gradient of G , i.e. $G_2 := d_x \text{grad} G$. Its singularity on the diagonal behaves like the Coulomb kernel, implying [Aub98, Theorem 4.13.c]:

$$G_2(x, y) = O(1/d_M(x, y)^d).$$

Moreover, in Riemannian manifolds:

$$\int_{d_M(x, y) > \varepsilon} d_M(x, y)^{-d} dy = O(-\log \varepsilon),$$

and if $\alpha < d$:

$$\int_{d_M(x, y) < \varepsilon} d_M(x, y)^{-\alpha} dy = O(\varepsilon^{d-\alpha}).$$

We write:

$$\|\varphi\|_{2, \infty} = \left(PV \int_{d(x, y) \leq \varepsilon} + \int_{d(x, y) > \varepsilon} \right) G_2(x, y) f(y) dy = I_1(x) + I_2(x).$$

As $\int_B G_2(x, y) dy = 0$ on metric balls, we get:

$$\begin{aligned} |I_1(x)| &= \left| \int_{d(x, y) \leq \varepsilon} G_2(x, y) (f(y) - f(x)) dy \right| \\ &\leq \int_{d(x, y) \leq \varepsilon} \|G_2(x, y)\| \|f\|_{C_M^{0, \gamma}} d(x, y)^\gamma dy \\ &\leq C \|f\|_{C_M^{0, \gamma}} \int_{d(x, y) \leq \varepsilon} d_M(x, y)^{-d+\gamma} dy \\ |I_1(x)| &\leq C \|f\|_{C_M^{0, \gamma}} \varepsilon^\gamma. \end{aligned}$$

For the second term:

$$\begin{aligned} |I_2(x)| &= \int_{d(x, y) > \varepsilon} G_2(x, y) f(y) dy \\ &\leq C \|f\|_\infty \int_{d(x, y) > \varepsilon} d(x, y)^{-d} dy \\ |I_2(x)| &\leq C \|f\|_\infty \log \left(\frac{A}{\varepsilon} \right). \end{aligned}$$

This proves the first estimate. \square

Sketch of proof for the second inequality. The second estimate is proven, for a manifold embedded in the Euclidean space, with standard Schauder theory in \mathbb{R}^d [Aub98, p. 3.61]. We use local normal coordinates for the Laplace-Beltrami operator. The bound in the Euclidean space can be applied thanks to the following metric control of the Hölder norm. We refer to the lemma below. \square

Lemma 2.22. *There exists $r > 0$ and $C > 0$ such that, if $\Omega \subset M$ has diameter less than r , then if $u \in C^{2, \gamma}$:*

$$\frac{1}{C} \|u\|_{C_{Eucl}^{k, \gamma, \Omega}} \leq \|u\|_{C_M^{k, \gamma}} \leq C \|u\|_{C_{Eucl}^{k, \gamma, \Omega}}.$$

In this inequality, $\|u\|_{C_{Eucl}^{k, \gamma, \Omega}}$ denotes the Euclidean Hölder norm in Ω when considered as a subset of \mathbb{R}^d .

Now, we use these estimates to conclude the proof of global existence. In our case $\|\psi_t\|_{2, \infty} = \|dv_t\|_{0, \infty}$ and $\|\psi_t\|_{C_M^{2, \gamma}} = \|dv_t\|_{C_M^{0, \gamma}}$. Note that in the second inequality, since the manifold M is closed we have $\|\psi\|_\infty \leq C$ for some positive constant since $\nabla \psi$ is bounded in L^∞ . We can use the first estimate from Lemma 2.21 with, once again, $\varepsilon = [\|\mu_t - \nu\|_\infty / |\mu_t - \nu|_\gamma]^{1/\gamma}$ to obtain formula (2.12) on the manifold M . This allows to bound $\|dv_t\|_\infty$ in the same way as in the Euclidean space. To bound

$|dv_t|_\gamma$, we use the second inequality of the lemma. The principle of the proof is the same as in the Euclidean space and equation (2.13) becomes:

$$\begin{aligned} \frac{d}{dt}|d\psi_t|_\gamma &\leq |d(v_t \circ \psi_t)|_\gamma \|d\psi_t\|_\infty + \|d(v_t \circ \psi_t)\|_\infty |d\psi_t|_\gamma \\ &\leq |dv_t|_\gamma \|d\psi_t\|_\infty^{1+\gamma} + \|dv_t\|_\infty \|d\psi_t\|_\infty |d\psi_t|_\gamma \\ &\leq C_1 (|\mu_t|_\gamma + \|v_t\|_\infty) \exp\left((1+\gamma) \int_0^t \|dv_s\|_\infty ds\right) + C_2 |d\psi_t|_\gamma \\ \frac{d}{dt}|d\psi_t|_\gamma &\leq C_1 + C_2 |d\psi_t|_\gamma. \end{aligned} \tag{2.14}$$

The rest of the proof follows the same steps as in the Euclidean space. This proves:

Theorem 2.23 (Global convergence for Hölder initial and target data). *Let μ_0 and ν_0 be Hölder continuous probability densities on a closed manifold (M, g) . Consider the curve μ_t global solution of equation (2.4). We have just proved that it is defined and Hölder continuous at all times. Then*

$$\begin{cases} E_\nu(\mu_t) \leq E_\nu(\mu_0)e^{-\lambda t} \\ W_2^2(\mu_t, \nu) \leq \frac{4}{\lambda} E_\nu(\mu_0)e^{-\lambda t}. \end{cases}$$

Proof. Since $\log(\mu_t)$ is globally bounded from below thanks to Lemma 2.12, it satisfies a global Polyak-Lojasiewicz at all times. The rest of the proof is a straightforward application of this inequality. \square

3. CRITICAL POINTS FOR THE WASSERSTEIN FLOW

In the previous section, we studied regular solutions of Wasserstein gradient flows, with smooth initial and target data. Our proofs heavily relied on regularity results obtained through potential theory estimates. In this section, we consider some given probability measures μ and ν that are, depending on the context, of finite energy for the Coulomb kernel or the Energy Distance kernel. Our goal is to study critical points for the Wasserstein flow of the MMD energy E_ν .

3.1. Critical Points and Lagrangian Critical Points for MMD Wasserstein Gradient Flows.

For an arbitrary function \mathcal{F} , we define critical points of the associated Wasserstein gradient flows. Intuitively, they correspond to measures where the discrete JKO steps are blocked, in direct analogy with gradient flows in finite dimensions.

Definition 3.1 (Wasserstein critical point). Let μ be a probability measure such that $\mathcal{F}(\mu) < +\infty$. We say that μ is a Wasserstein critical point of \mathcal{F} if there exists $\tau_0 > 0$ such that for all $0 < \tau \leq \tau_0$ we have:

$$\mu \in \arg \min_{\rho \in \mathcal{P}(\mathbb{R}^d)} \mathcal{F}(\rho) + \frac{1}{2\tau} W_2^2(\mu, \rho).$$

We study a sub-class of critical points, which we call Lagrangian critical points or displacement critical points.

Definition 3.2. A probability measure μ is said to be a Lagrangian critical point for a functional \mathcal{F} if, on $\text{supp}(\mu)$ we have

$$\nabla \frac{\delta \mathcal{F}}{\delta \mu}(\mu) = 0.$$

For differentiable λ -convex functionals [San15], the two definitions are equivalent. In general settings, we cannot deduce one from the other, as the quantity above may not even belong to $\partial \mathcal{F}(\mu)$. However, in our cases, we can prove a partial result. Indeed, using the general differentiation result on JKO steps given by proposition A.10 we see that the following Proposition is true.

Proposition 3.3. *Let μ be a critical point as defined in 3.1. Then $0 \in \partial \mathcal{F}(\mu)$ (where 0 is seen as an element of $L_2(\mu)$).*

Combining this result and lemma 2.4, we see that if μ is a critical point of the G -energy E_ν , then the gradients of the potentials $G \star \mu$ and $G \star \nu$ are equal μ -almost everywhere. This observation is the key argument enabling us to prove the results of this section.

3.2. Characterization of Lagrangian Critical points for MMD Wasserstein Gradient Flows.

The main result of this section is presented in the following theorem.

Theorem 3.4. *Let μ be a Lagrangian critical point for the MMD functional E_ν^G . Then*

$$\mu|_{\text{Int}(\text{supp}(\mu))} = \nu|_{\text{Int}(\text{supp}(\mu))},$$

holds if G is the Coulomb kernel or if G is the Energy Distance kernel and d is odd.

This formulation may appear surprising. We obtain results only in the interior of the support of our measure, even though it may be empty, or μ may be composite, being the sum of a density measure and a singular measure for example. As the proof for the Energy Distance kernel is more involved, we focus on the first statement for now.

In both cases, the first variation of the E_ν is given by: $\frac{\delta E_\nu}{\delta \mu}(\mu) = \phi_\mu - \phi_\nu$. For measures μ et ν with finite G -energy, i.e. for measures such that $\int G d\mu^{\otimes 2}, \int G d\nu^{\otimes 2} < +\infty$, this function is locally integrable, and satisfies in a distributional sense:

$$\nabla \frac{\delta E_\nu}{\delta \mu}(\mu) = \nabla \phi_\mu - \nabla \phi_\nu. \quad (3.1)$$

If μ is a Lagrangian critical point for the Coulomb kernel, this quantity is constant equal to 0. Upon differentiating once again in a distributional sense we conclude that in the interior of the domain $\text{supp}(\mu)$, $d\mu = d\nu$ as per Proposition B.3, proving the first part of the theorem. The proof in the Energy Distance case requires iterating Laplacians.

Proposition 3.5. *Let μ be a probability measure on \mathbb{R}^d . Consider the associated potential $\phi_\mu(x) := \int -\|x - y\| d\mu(y)$. Then, its distributional Laplacian exists almost everywhere and is given by:*

$$\Delta \phi_\mu(x) := \int -\frac{d-1}{\|x-y\|} d\mu(y).$$

Proof. Let $g \in C_c^\infty$, and denote $P_\mu(x)$ as $\int -\frac{d-1}{\|x-y\|} d\mu(y)$. We wish to show $\langle \phi_\mu, \Delta g \rangle = \langle P_\mu, g \rangle$. The primary challenge arises from the singularity of the function $\|\cdot - y\|$ at y . We circumvent this by integrating on a small ball of radius $\varepsilon > 0$ near y , carefully controlling the error term. The inversion of integrals in the second equality is justified by the Fubini theorem.

$$\begin{aligned} \langle \phi_\mu, \nabla g \rangle &:= \int \phi_\mu(x) \Delta g(x) dx \\ &= \int \left(\int -\|x-y\| \Delta g(x) dx \right) d\mu(y) \\ &= \int \left[\left(\int_{B(y,\varepsilon)} -\|x-y\| \Delta g(x) dx \right) + \left(\int_{\mathbb{R}^d \setminus B(y,\varepsilon)} -\|x-y\| \Delta g(x) dx \right) \right] d\mu(y) \\ \langle \phi_\mu, \nabla g \rangle &= \int [I_\varepsilon(y) + J_\varepsilon(y)] d\mu(y), \end{aligned}$$

where the quantities I_ε and J_ε are defined by the formulas just above. First, as $g \in C_c^\infty$, we get:

$$|I_\varepsilon(y)| \leq 2\|\Delta g\|_\infty \varepsilon.$$

For $J_\varepsilon(y)$, we can use Green's Formula, since all our quantities are in C_c^∞ on the open set $\Omega := \mathbb{R}^d \setminus B(y, \varepsilon)$. If we take $\eta(x)$ to be the unit vector at $x \in \partial\Omega(y)$ pointing toward the exterior of $\Omega(y)$, then $\eta(x) = -\frac{x-y}{\|x-y\|}$. For $x \in \partial\Omega(y)$, the directional derivative of a function f at x is defined by

$\frac{\partial f}{\partial \eta}(x) := \langle \nabla f(x), \eta(x) \rangle$. We denote dS the usual area measure on $B(y, \varepsilon)$. Using Green's Formula we have:

$$J_\varepsilon(y) = \int_{\Omega(y)} \Delta - d_y(x)g(x)dx + \int_{\partial\Omega(y)} \frac{\partial d_y}{\partial \eta}(x)g(x)dS(x) - \int_{\partial\Omega(y)} d_y(x) \frac{\partial g}{\partial \eta}(x)dS(x).$$

First, we know that $\Delta d_y(x) = \frac{d-1}{\|x-y\|}$. Second, as $\nabla d_y(x) = \frac{x-y}{\|x-y\|^2}$, we have $\frac{\partial d_y}{\partial \eta}(x) = -1$. This leads to:

$$\left| \int_{\partial\Omega(y)} \frac{\partial d_y}{\partial \eta}(x)g(x)dS(x) \right| \leq \|g\|_\infty \mathcal{A}(S(y, \varepsilon)).$$

Finally, as g vanishes at infinity, there exists a constant C independent of y such that:

$$\left| \int_{\partial\Omega(y)} d_y(x) \frac{\partial g}{\partial \eta}(x)dS(x) \right| \leq C \mathcal{A}(S(y, \varepsilon)).$$

All these quantities vanish as $\varepsilon \rightarrow 0^+$, independently of y , which implies $\langle \phi_\mu, \Delta g \rangle = \langle P_\mu, g \rangle$, i.e. $\Delta \phi_\mu(x) = \int -\frac{d-1}{\|x-y\|} d\mu(y)$. \square

Now, we prove the following property, enabling us to prove next the second statement in Theorem 3.4.

Lemma 3.6. *Let μ be a Lagrangian critical point for the Energy Distance Wasserstein gradient flow towards ν . Then, on the open set $\text{Int}(\text{supp}(\mu))$ $\phi_\mu^G = \phi_\nu^G$ where G is the Coulomb kernel.*

Proof. We denote $P_k^\mu(x) := \int \frac{1}{\|x-y\|^k} d\mu(y)$ and prove the following result by finite induction:

$$\forall k < (d-1)/2, P_{2k+1}^\mu = P_{2k+1}^\nu.$$

We already established the result is true for $k = 1$. Now let us take $k < (d-3)/2$ and suppose $P_{2k+1}^\mu = P_{2k+1}^\nu$. We know that if $v : \mathbb{R}^+ \rightarrow \mathbb{R}$ and $u(x) := v(\|x\|)$, then its distributional Laplacian is given by:

$$\Delta u(x) = \frac{d-1}{\|x\|} v'(\|x\|) + v''(\|x\|).$$

If $v(r) := \frac{1}{r^{2k+1}}$, then $\Delta u(x) = \frac{(2k+1)(2k+3-d)}{\|x\|^{-(2k+3)}}$. Denoting $K_y(x) := \frac{1}{\|x-y\|^{2k+1}}$, we have $\Delta K_y(x) = \frac{(2k+1)(2k+3-d)}{\|x-y\|^{2k+3}}$. Again, let us take $g \in C_c^\infty$ and prove $\langle P_{2k+1}^\mu, \Delta g \rangle = (2k+1)(2k+3-d) \langle P_{2k+3}^\mu, g \rangle$. We use Fubini's theorem, and avoid the singularity around small balls of radius $\varepsilon > 0$:

$$\begin{aligned} \langle P_{2k+1}^\mu, g \rangle &:= \int P_{2k+1}^\mu(x) \Delta g(x) dx \\ &= \int \left(\int K_y(x) \Delta g(x) dx \right) d\mu(y) \\ &= \int \left[\left(\int_{B(y, \varepsilon)} K_y(x) \Delta g(x) dx \right) + \left(\int_{\mathbb{R}^d \setminus B(y, \varepsilon)} K_y(x) \Delta g(x) dx \right) \right] d\mu(y) \\ \langle P_{2k+1}^\mu, g \rangle &= \int [I_\varepsilon(y) + J_\varepsilon(y)] d\mu(y). \end{aligned}$$

Now, again, we need to control the growth of $I_\varepsilon(y)$ and $J_\varepsilon(y)$. Since $2k+1 < d$ we get:

$$\int_{B(y, \varepsilon)} K_y(x) \Delta g(x) dx \leq \|\Delta g(x)\|_\infty \int_{B(0, \varepsilon)} K_0(x) dx,$$

and this quantity tends to 0 uniformly as ε tends to 0. Now, as done previously, we use Green's Formula to express $J_\varepsilon(y)$:

$$J_\varepsilon(y) = \int_{\Omega(y)} \Delta K_y(x)g(x)dx - \int_{\partial\Omega(y)} \frac{\partial K_y}{\partial \eta}(x)g(x)dS(x) + \int_{\partial\Omega(y)} K_y(x) \frac{\partial g}{\partial \eta}(x)dS(x).$$

We have:

$$\frac{\partial K_y}{\partial \eta}(x) = \left\langle -(2k+1) \frac{x-y}{\|x-y\|^{2k+3}}, -\frac{x-y}{\|x-y\|} \right\rangle = \frac{1}{\|x-y\|^{2k+2}},$$

which gives:

$$\left| \int_{\partial\Omega(y)} \frac{\partial K_y}{\partial \eta}(x) g(x) dS(x) \right| \leq \|g\|_\infty \int \frac{1}{\|x-y\|^{2k+2}} dS(x) = \|g\|_\infty C_d \varepsilon^{d-1-(2k+2)},$$

and again the right term tends to 0 uniformly in y as ε tends to 0, as $2k+2 < d-1$. Finally:

$$\int_{\partial\Omega(y)} K_y(x) \frac{\partial g}{\partial \eta}(x) dS(x) \leq \left\| \frac{\partial g}{\partial \eta} \right\|_\infty \int_{\partial\Omega(y)} K_y(x) dS(x) = \left\| \frac{\partial g}{\partial \eta} \right\|_\infty C_d \varepsilon^{d-1-(2k+1)},$$

which tends to 0 uniformly in y as ε tends to 0. This proves:

$$\Delta P_{2k+1}^\mu = (2k+1)(2k+3-d)P_{2k+3}^\mu.$$

Now, back to our hypothesis $P_{2k+1}^\mu = P_{2k+1}^\nu$. By taking the distributional Laplacian from both sides, we get $(2k+1)(2k+3-d)P_{2k+3}^\mu = (2k+1)(2k+3-d)P_{2k+3}^\nu$, and as $2k < d-3$, we get: $P_{2k+3}^\mu = P_{2k+3}^\nu$. This proves our result. Now, to obtain the conclusion, if d is odd, by taking $k = (d-3)/2 < (d-1)/2$, we finally get in the interior of $\text{supp}(\mu)$ the equality $\phi_\mu^G = \phi_\nu^G$. \square

The potentials ϕ_μ^G and ϕ_ν^G are superharmonic on the interior of $\text{supp}(\mu)$, and taking their distributional Laplacian yields $\mu|_{\text{Int}(\text{supp}(\mu))} = \nu|_{\text{Int}(\text{supp}(\mu))}$, which is exactly the second statement of Theorem 3.4.

In this section, we were able to characterize a particular class of critical points. However, our result does not capture the singular parts of our measures. In the next section, we study how singular measures behave under the Wasserstein gradient flow of E_ν , proving in Theorem 4.11 that singular energy measures cannot be critical points.

4. NO LOCAL MINIMA IN THE WASSERSTEIN GEOMETRY

Recall that our PL inequality does not hold at a measure μ if there is no strictly positive lower bound on the density of μ . In particular, it is unclear if local non-global minima of the MMD norm E_ν can exist. In this section, we establish that there are no local minima other than the global one. The previous section was dedicated to Lagrangian critical points. We described measures μ where the velocity field induced by $\nabla \frac{\delta E_\nu}{\delta \mu}(\mu)$ is equal to 0 everywhere. In these cases, a JKO step cannot be described through a push-forward map. However, this pushforward action by maps does not describe all possible dynamics, and diffusion of the mass can occur. In terms of Wasserstein geometry, it means, heuristically, that a descent direction has to be found in the space of velocity plans [AGS05, Section 12.4], instead of L_2 velocity maps. To do this, we use concepts inspired by flow interchange techniques developed in [MMS09; KMX15]. Instead of studying the gradient flow of E , we study how E behaves along a certain auxiliary flow, here associated with the Boltzmann entropy functional.

We show, as the main result of this section, the following theorem.

Theorem 4.1 (No local non global minima). *Let μ be a probability measure. Then, if $\mu \neq \nu$, there exists a curve μ_t which is 1/2-Hölder for the Wasserstein distance, such that $t \mapsto E_\nu(\mu_t)$ is strictly decreasing for t small enough.*

We begin by noting that the MMD energy only depends on the difference between the measures. Indeed, since $(\mu, \nu) \mapsto E_\nu(\mu)$ is a function of $\mu - \nu$, we can use the Hahn-Jordan decomposition of the signed measure $\mu - \nu$.

Lemma 4.2 (Hahn-Jordan decomposition). *Let μ and ν be two probability measures. Then there exists a unique decomposition $\mu - \nu = \mu_+ - \nu_-$, where μ_+ and ν_- are mutually singular measures. Additionally, $\mu_+ \ll \mu$ and $\nu_- \ll \nu$.*

A heat diffusion process can be applied to μ_+ , enabling us to prove the main theorem of this section, stating that even though our functional is non-convex in the Wasserstein geometry, it does not admit any local minima for the Wasserstein geometry that is not global (meaning $\mu = \nu$ or in an equivalently $\mu_+ = \nu_- = 0$).

4.1. Heat Diffusion Perturbation. Let ρ_0 be a measure dominated by μ . We write $\mu = \mu - \rho + \rho$ and consider the curve ρ_t defined as the solution of the heat equation $\partial_t \rho_t = \Delta \rho_t$ with initial condition $\rho_0 = \rho$. As is well-known, the heat equation is the Wasserstein gradient flow of the Boltzmann entropy functional defined by: $\mathcal{H}(\rho) := \int \rho \log(\rho) dx$ if the measure has a density ρ (with abuse of notation) with respect to the Lebesgue measure and $+\infty$ otherwise. Furthermore, the solution ρ_t is explicit: $\rho_t = K_t \star \rho$, where K_t is the heat kernel in \mathbb{R}^d defined by

$$K_t(x) := \frac{1}{(4\pi t)^{d/2}} \exp(-\|x\|^2/4t).$$

Remark 4.3. Here, ρ is not necessarily a probability measure. However, we can write the whole Wasserstein formalism for any measure space $\mathcal{M}_m := \{\mu \in \mathcal{M}_+(\mathbb{R}^d), \mu(\mathbb{R}^d) = m\}$. We denote $W_2(\alpha, \beta)$ the Wasserstein distance on \mathcal{M}_m if $\alpha, \beta \in \mathcal{M}_m$.

Lemma 4.4. *Consider the curve defined by $\mu_t := \mu + \rho_t - \rho$, where ρ_t is the heat flow at time t of ρ_+ . Then, the curve μ_t is absolutely continuous. More precisely, there exists a constant $C > 0$ such as, for any $s, t > 0$:*

$$W_2(\mu_t, \mu_s) \leq \sqrt{|t-s|}C.$$

Proof. Standard results exposed in [AGS05, Theorem 11.2.8] show that $K_t \star \rho$ is absolutely continuous for the Wasserstein metric. We fix some $s, t > 0$, and write the EVI equation associated with the heat flow (see (A.6)), that states that for all α density measure with the same mass as ρ and for almost every $t > 0$:

$$\frac{1}{2} \frac{d}{dt} W_2^2(\rho_t, \alpha) \leq \mathcal{H}(\alpha) - \mathcal{H}(\rho_t).$$

This implies that there exists a constant $C > 0$ such that, for any $s, t > 0$ $W_2(\mu_t, \mu_s) \leq \sqrt{|t-s|}C$. Let γ be an optimal transport plan γ between $K_t \star \rho$ and $K_s \star \rho$, for $s, t > 0$. The plan $\tilde{\gamma} := \gamma + (\mu - \rho)^{\otimes 2}$ is a transport plan between μ_t and μ_s . Thus:

$$W_2(\mu_t, \mu_s) \leq W_2(K_t \star \rho, K_s \star \rho),$$

which proves the lemma. \square

Now, we need some estimates on E_ν along the flow defined above. It is possible thanks to the next lemma.

Lemma 4.5. *The function $t \mapsto E_\nu(\mu_t)$ is differentiable for every $t > 0$, and its derivative is given by:*

$$\frac{1}{c_d} \frac{d}{dt} E_\nu(\mu_t) = -\langle \rho, K_{2t} \star \rho + K_t \star (\mu - \rho) - K_t \star \nu \rangle. \quad (4.1)$$

4.2. Estimates of Mass Transfers in the Diffusion Process. In this paragraph, we show that for a well-chosen positive measure ρ and for t small enough, the quantity in Formula (4.1) is strictly negative. An essential property is the following lemma [Wat94].

Lemma 4.6 (Heat kernel estimates, [Wat94]). *Let α and β be two positive mutually singular measures. Then, as $t \rightarrow 0$, for μ almost every x :*

$$K_t \star \beta(x) = o(K_t \star \alpha(x)),$$

We use this lemma to derive the following estimate.

Lemma 4.7. *Let α and β be two positive mutually singular measures. Then, as $t \rightarrow 0$, for μ almost every x :*

$$K_t \star \beta(x) = o(K_{2t} \star \alpha(x)),$$

Proof. From the expression of K_t , we have $\frac{K_t(x)}{K_{2t}(x)} = 2^{d/2} \exp(-\|x\|^2/8t) \leq 2^{d/2}$, from which we deduce $K_t \star \alpha(x) \leq 2^{d/2} K_{2t} \star \alpha(x)$ and:

$$\frac{K_t \star \beta(x)}{K_{2t} \star \alpha(x)} = \frac{K_t \star \alpha(x)}{K_{2t} \star \alpha(x)} \frac{K_t \star \beta(x)}{K_t \star \alpha(x)} \leq 2^{d/2} \frac{K_t \star \beta(x)}{K_t \star \alpha(x)}.$$

This quantity tends to 0 from the preceding lemma. \square

4.3. Proof of Theorem 4.1. We now prove the main theorem of this section, which is implied by the following proposition.

Proposition 4.8. *Let μ be a probability measure. We suppose $\mu \neq \nu$, and write $\mu - \nu = \mu_+ - \nu_-$ the unique associated Hahn-Jordan decomposition. Then there exists a μ_+ measurable set A such as $\mu_+(A) \geq \mu_+(\mathbb{R}^d)/2$ and $t_0 > 0$ such as, for the curve $\mu_t = \mu + K_t \star \mu_{+|A} - \mu_{+|A}$, then for all $t < t_0$ we get $\frac{d}{dt} E_\nu(\mu_t) < 0$.*

Proof. The preceding lemma gives the following result: for μ_+ almost every x , there exists $t_x > 0$ such as, for all $t < t_x : K_t \star \nu_-(x) < \frac{1}{2} K_{2t} \star \mu_+(x)$. Let us consider the set sequence defined for $N > 0$ by:

$$X_N := \{x \in \text{supp}(\mu_+) | t_x \geq 1/N\}.$$

This is a growing sequence for inclusion, and it verifies $\text{supp}(\mu_+) = \bigcup_{N \in \mathbb{N}^*} X_N$. Now if we write $\widetilde{X}_1 := X_1$ $\widetilde{X}_{N+1} := X_{N+1} \setminus X_N$, we get a countable sequence of disjoint sets whose union is of total mass for μ_+ . This implies, by σ -additivity, that there exists an integer N_0 such that:

$$\mu_+ \left(\bigcup_{N=1}^{N_0} X_N \right) \geq \mu_+(\mathbb{R}^d)/2.$$

Now, we take $A := \bigcup_{N=1}^{N_0} X_N$ and $t_0 = 1/N_0$, which gives, for $t < t_0$:

$$\frac{d}{dt} E_\nu(\mu_t) = -c_d \langle \mu_{+|A}, K_{2t} \star \mu_{+|A} - K_t \star \nu_- \rangle \quad (4.2)$$

$$\leq -\frac{1}{2} c_d \langle \mu_{+|A}, K_{2t} \star \mu_{+|A} \rangle. \quad (4.3)$$

In particular, we get the conclusion $\frac{d}{dt} E_\nu(\mu_t) < 0$. \square

4.4. Dimension of Measure and Critical Points. In Theorem 4.1, we saw that for any measure μ distinct from the target measure ν we could find an absolutely continuous curve for the Wasserstein distance μ_t such as $t \mapsto E_\nu(\mu_t)$ was strictly decreasing near 0. However, it is not sufficient to guarantee that the JKO steps (A.5) do not remain stationary at μ . Our Theorem 3.4 states that if μ is a critical point, then on the interior of its support μ is equal to ν . We will prove that if the part of μ singular to ν is supported on sets singular enough, then μ cannot be a critical point as in Definition 3.1. In analogy to the finite dimensions and regular case, it corresponds to the fact that it cannot be a second-order critical point.

To quantify the singularity of the support of a measure, we use geometric measure theory properties, mainly the growth of $\mu(B(x, r))$. If it grows faster than r^d , we prove that the heat diffusion makes the energy E_ν decrease fast enough to compensate for the growth of $W_2^2(\mu, K_t \star \mu)$, so that μ is not a point where JKO steps get stuck. To do this, we use a more precise version of Lemma 4.6, see [Wat94]:

Lemma 4.9. *Let μ be a positive measure in \mathbb{R}^d and $q \in [0, n]$. Then, there exists a universal constant $c_{d,q}$ only dependent on d and q such as, for all $x \in \mathbb{R}^d$:*

$$\liminf_{r \rightarrow 0} r^{-q} \mu(B(x, r)) \leq c_{d,q} \liminf_{t \rightarrow 0} t^{(d-q)/2} K_t \star \mu(x) \quad (4.4)$$

$$\leq c_{d,q} \limsup_{t \rightarrow 0} t^{(d-q)/2} K_t \star \mu(x) \leq \limsup_{r \rightarrow 0} r^{-q} \mu(B(x, r)). \quad (4.5)$$

How to interpret this result? This result allows, in some cases, to get precise estimates on the decay of $K_t \star \mu$. To illustrate, if μ has a continuous bounded density function f with respect to the Lebesgue measure, then, with $q = d$, we get that $\lim_{r \rightarrow 0} r^{-d} \mu(B(x, r)) = f(x)$, which gives:

$$\lim_{t \rightarrow 0} K_t \star \mu(x) = f(x)/c_{d,q}.$$

If μ charges more singular sets than open sets, for example if there exists some $\delta > 0$ such as, for all $x \in A$, where $\mu(A) > 0$ $\lim_{r \rightarrow 0} r^{-(d-\delta)} \mu(B(x, r)) = f(x)$, then, for all $x \in A$ we have $\lim_{t \rightarrow 0} t^{\delta/2} K_t \star \mu(x) = f(x)/c_{d,q}$. That is, when $t \rightarrow 0$:

$$K_t \star \mu(x) \sim \frac{f(x)}{c_{d,q}} \frac{1}{t^{\delta/2}}.$$

Let μ and ν be two probability measures with finite Coulomb energy. Once again we write the Hahn-Jordan decomposition $\mu - \nu = \mu_+ - \nu_-$. From the result of Section 3, we know that at a critical point that is not the global minimum, μ_+ has an empty interior support.

We study what happens if μ_+ is singular in the sense that the growth of $\mu_+(B(x, r))$ on the support of μ_+ as r tends to 0. We prove the following lemma:

Lemma 4.10. *If there exists a set $A \subset \mathbb{R}^d$ such that $\mu_+(A) > 0$, some positive number $0 < \delta < 2$, a constant $C > 0$ and $r_0 > 0$ such that, for all $0 \leq r \leq r_0$ and μ_+ almost every $x \in A$:*

$$\mu_+(B(x, r)) \geq Cr^{d-\delta},$$

then μ is not a critical point as in Definition 3.1.

Proof. In this proof, we denote by C any strictly positive constant that does not depend on t or τ . Once again, we consider the Wasserstein curve defined by $\mu_0 = \mu$ and $\mu_t = \mu + K_t \star \mu_{+|A} - \mu_{+|A}$. Using Lemma 4.9, we get the existence of $t_0 > 0$ such that, for all $0 < t < t_0$ and μ_+ almost every $x \in A$

$$K_{2t} \star \mu_+(x) \geq \frac{C}{t^{\delta/2}}. \quad (4.6)$$

Moreover, by the same reasoning as in the proof of Theorem 4.8, we can suppose, if we take a subset of A , that there exists $t_1 > 0$ such as, for any $0 < t < t_1$:

$$\frac{d}{dt} E_\nu(\mu_t) \leq -C \langle \mu_{+|A}, K_{2t} \star \mu_{+|A} \rangle.$$

Now, using Formula (4.6), this gives along μ_t , $\frac{d}{dt} E_\nu(\mu_t) \leq -Ct^{-\delta/2}$. From the EVI inequality associated with the entropy gradient flow, we get the existence of a constant C such that $\frac{d}{dt} W_2^2(\mu_0, \mu_t) \leq C$. Combining these inequalities, we get an estimate on the derivative of our proximal functional $E_\nu^{\tau, \mu} := E_\nu + \frac{1}{2\tau} W_2^2(\mu_0, \cdot)$:

$$\frac{d}{dt} E_\nu^{\tau, \mu}(\mu_t) \leq C \left(-t^{-\delta/2} + \frac{1}{2\tau} \right).$$

This provides, integrating from 0 to t , as $\delta/2 < 1$:

$$E_\nu^{\tau, \mu}(\mu_t) - E_\nu^{\tau, \mu}(\mu_0) \leq C \left(-\frac{2t^{-\delta/2+1}}{2-\delta} + \frac{t}{2\tau} \right). \quad (4.7)$$

Optimizing this quantity over $t > 0$, we get an optimal $t_\tau := (2\tau)^{2/\delta}$, smaller than t_0, t_1 if τ is small enough so that all our inequalities are true for t_τ . Injecting t_τ in equation (4.7) we get:

$$E_\nu^{\tau, \mu}(\mu_{t_\tau}) - E_\nu^{\tau, \mu}(\mu_0) \leq -C\tau^{\frac{2-\delta}{\delta}},$$

concluding the proof since:

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} E_\nu^{\tau, \mu}(\rho) - E_\nu^{\tau, \mu}(\mu_0) \leq \min_{t > 0} E_\nu^{\tau, \mu}(\mu_t) - E_\nu^{\tau, \mu}(\mu_0) \leq -C\tau^{\frac{2-\delta}{\delta}}. \quad \square$$

This lemma is used to prove the main result of this subsection.

Theorem 4.11. *Let us assume there exists a set A such that $\mu_+(A) > 0$ and such that $\mu_{+|A}$ has Minkowski dimension less than $d - \delta$ (see [Heu07]) for some $0 < \delta \leq 1$. Then μ is not a critical point as in Definition 3.1.*

Proof. The local Minkowski dimension of a measure at x is defined, if the limit exists, as

$$\dim_{\mathcal{M}}^x(\mu) := \lim_{r \rightarrow 0} \frac{\log(\mu(B(x, r)))}{\log(r)}.$$

If μ_+ charges some set A such as, for all $x \in A$, $\dim_{\mathcal{M}}^x(\mu)$ exists and is less than $d - \delta$, then for every $\varepsilon > 0$ there exists a radius r_x such that for any $r < r_x$, $\mu(B(x, r)) \geq r^{d-\delta+\varepsilon}$. Using the same σ -additivity arguments as in the proof of Proposition 4.8, we prove we can choose a uniform r_0 by considering a subset of A of non-zero measure for μ_+ . This is exactly the hypothesis of Lemma 4.10, concluding the proof. \square

Remark 4.12. The condition on the Minkowski dimension is satisfied if, for example, μ_+ is absolutely continuous with respect to the volume measure on some manifold of dimension $d - 1$.

Note that the result is stated only for $\delta \leq 1$ instead of $\delta < 2$ (as assumed in the lemma). This is explained by the fact that for $\delta > 1$, the energy functional is infinite. In such a case, it is straightforward to prove from the very definition that such a measure (of infinite energy) cannot be a critical point.

Comments on possible critical points. Our main goal in this section was to prove that there is no local minima apart from the global one. However, it does not prevent the existence of critical points. Passing by, we were able to rule out some singular measures from being critical points. Indeed, we proved that if the current measure has a Minkowski dimension strictly less than the ambient one, then it cannot be a critical point as defined in Section 3. However, there are still many candidates for critical points such as measures proportional to 1_A where A is a positive Lebesgue measure closed set of empty interior such as a fat Cantor set. A complete characterization of critical points, and the existence of a nontrivial example, is left open for future works.

4.5. Extension to Riemannian Manifolds. Recall that the heat kernel is defined on a general Riemannian manifold.

Proposition 4.13 (Heat kernel on a manifold). *Let (M, g) be a Riemannian manifold (it doesn't need to be compact). Then we can define its heat kernel $K : (0, \infty) \times M \times M \rightarrow \mathbb{R}$ as the smallest positive fundamental solution of the heat equation, meaning that for any $y \in M$:*

$$\begin{cases} \partial_t K = \Delta_x K \\ K(t, \cdot, y) \xrightarrow{t \rightarrow 0} \delta_y. \end{cases}$$

To generalize the conclusions of Section 4, we need estimates resembling the one in the proof of Lemma 4.7, mainly that there is some constant $C > 0$ such that:

$$K_{2t} \leq CK_t. \quad (4.8)$$

In the Euclidean case $M = \mathbb{R}^d$, we got $C = 2^{d/2}$. We do not detail the conditions for this to hold, but it is true if M is a nilpotent Lie group equipped with a left-invariant metric ([Var90]), if (M, g) is a geodesically complete non-compact Riemannian manifold of nonnegative Ricci curvature ([Gri99]), or if (M, g) is compact ([LY86]). For example, on the flat 1-dimensional torus $\mathbb{T} := \mathcal{S}^1$ which we represent as $[0, 1]/\{0 \sim 1\} = \mathbb{R}/\mathbb{Z}$, the heat kernel is given by the periodization of the Euclidean heat kernel:

$$K_{\mathbb{T}}(t, x, y) := \sum_{n \in \mathbb{Z}} K_{\mathbb{R}}(t, x + n, y).$$

As the flat d -dimensional torus $\mathbb{T}^d := (\mathcal{S}^1)^d$ is a product space, its heat kernel is the product, defined for $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$ by:

$$K_{\mathbb{T}^d}(t, x, y) = \prod_{1 \leq k \leq d} K_{\mathbb{T}}(t, x_k, y_k),$$

so that the desired estimate in Formula (4.8) holds for some $C > 0$. An example of a manifold where it does not hold is the hyperbolic space.

ACKNOWLEDGEMENTS

The authors thank very warmly Stephen C. Preston for his help on radial solutions. They also thank Adrien Vacher for his help at the beginning of this work and Théo Dumont for his careful proofreading. This work was supported by the Bézout Labex (New Monge Problems), funded by ANR, reference ANR-10-LABX-58.

REFERENCES

- [AGS05] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005 (cit. on pp. 5, 17, 18, 26–28).
- [Arb+19] Michael Arbel, Anna Korba, Adil Salim, and Arthur Gretton. “Maximum mean discrepancy gradient flow”. In: *Advances in Neural Information Processing Systems 32* (2019) (cit. on p. 29).
- [Aub98] T. Aubin. *Some Nonlinear Problems in Riemannian Geometry*. Springer, 1998 (cit. on pp. 13, 30).
- [Bar93] A.R. Barron. “Universal approximation bounds for superpositions of a sigmoidal function”. In: *IEEE Transactions on Information Theory* 39.3 (1993), pp. 930–945. DOI: [10.1109/18.256500](https://doi.org/10.1109/18.256500) (cit. on p. 3).
- [BCY14] Daniel Balagué, José A. Carrillo, and Yao Yao. “Confinement for repulsive-attractive kernels”. In: *Discrete and Continuous Dynamical Systems - B* 19.5 (2014), pp. 1227–1248. ISSN: 1531-3492. DOI: [10.3934/dcdsb.2014.19.1227](https://doi.org/10.3934/dcdsb.2014.19.1227). URL: <https://www.aims sciences.org/article/id/60> (cit. on p. 4).
- [BLL12] Andrea L. Bertozzi, Thomas Laurent, and Flavien Leger. “Aggregation and spreading via the Newtonian potential : The dynamics of patch solutions”. In: *Mathematical Models and Methods in Applied Sciences* 22.supp01 (2012). <http://www.math.ucla.edu/~bertozzi/papers/Patch-Galley.pdf>, p. 1140005. DOI: [10.1142/s0218202511400057](https://doi.org/10.1142/s0218202511400057). URL: <https://app.dimensions.ai/details/publication/pub/10.1142/s0218202511400057> (cit. on pp. 8, 10).
- [Bon13] Bonaschi. “Gradient flows driven by a non-smooth repulsive interaction potential”. In: (2013) (cit. on p. 6).
- [Car+11] J. Carrillo, Massimo Di Francesco, Alessio Figalli, T. Laurent, and D. Slepčev. “Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations”. In: *Duke Mathematical Journal* 156 (Feb. 2011). DOI: [10.1215/00127094-2010-211](https://doi.org/10.1215/00127094-2010-211) (cit. on p. 4).
- [Car+12] J.A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev. “Confinement in nonlocal interaction equations”. In: *Nonlinear Analysis: Theory, Methods & Applications* 75.2 (2012), pp. 550–558. ISSN: 0362-546X. DOI: <https://doi.org/10.1016/j.na.2011.08.057>. URL: <https://www.sciencedirect.com/science/article/pii/S0362546X11006195> (cit. on p. 4).
- [Car45] Cartan. “Théorie du potentiel newtonien : énergie, capacité, suites de potentiels”. In: *Bulletin de la Société Mathématique de France* 73 (1945), pp. 74–106. DOI: [10.24033/bsmf.1368](https://doi.org/10.24033/bsmf.1368) (cit. on p. 30).
- [CB18] Lenaïc Chizat and Francis Bach. *On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport*. 2018. arXiv: [1805.09545](https://arxiv.org/abs/1805.09545) [math.OA] (cit. on p. 3).
- [CCH14] José Antonio Carrillo, Michel Chipot, and Yanghong Huang. “On global minimizers of repulsive–attractive power-law interaction energies”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2028 (Nov. 2014), p. 20130399. DOI: [10.1098/rsta.2013.0399](https://doi.org/10.1098/rsta.2013.0399). URL: <https://doi.org/10.1098/rsta.2013.0399> (cit. on p. 4).

- [CRS23] Antonin Chodron de Courcel, Matthew Rosenzweig, and Sylvia Serfaty. *Sharp uniform-in-time mean-field convergence for singular periodic Riesz flows*. 2023. arXiv: [2304.05315](https://arxiv.org/abs/2304.05315) [[math.AP](#)] (cit. on p. 4).
- [Gar19] David García-Zelada. “Concentration for Coulomb gases on compact manifolds”. In: *Electronic Communications in Probability* 24.none (2019), pp. 1–18. DOI: [10.1214/19-ECP211](https://doi.org/10.1214/19-ECP211) (cit. on p. 30).
- [Gri99] Grigor’yan. “Estimates of heat kernels on Riemannian manifolds”. In: (1999) (cit. on p. 21).
- [Hag+23] Hagemann, Hertrich, Altekrüger, Beinert, Chemseddine, and Steidl. “Posterior Sampling Based on Gradient Flows of the MMD with Negative Distance Kernel”. In: *arXiv preprint arXiv:2310.03054* (2023) (cit. on p. 3).
- [Her+23] Johannes Hertrich, Christian Wald, Fabian Altekrüger, and Paul Hagemann. *Generative Sliced MMD Flows with Riesz Kernels*. 2023. arXiv: [2305.11463](https://arxiv.org/abs/2305.11463) [[cs.LG](#)] (cit. on p. 3).
- [Her+24] Johannes Hertrich, Manuel Gräf, Robert Beinert, and Gabriele Steidl. “Wasserstein steepest descent flows of discrepancies with Riesz kernels”. In: *Journal of Mathematical Analysis and Applications* 531.1 (Mar. 2024), p. 127829. DOI: [10.1016/j.jmaa.2023.127829](https://doi.org/10.1016/j.jmaa.2023.127829). URL: <https://doi.org/10.1016/j.jmaa.2023.127829> (cit. on p. 3).
- [Heu07] Yanick Heurteaux. “Dimension of measures: the probabilistic approach.” eng. In: *Publicacions Matemàtiques* 51.2 (2007), pp. 243–290. URL: <http://eudml.org/doc/41895> (cit. on p. 21).
- [JKO98] Richard Jordan, David Kinderlehrer, and Felix Otto. “The variational formulation of the Fokker–Planck equation”. In: *SIAM Journal on Mathematical Analysis* 29.1 (1998), pp. 1–17 (cit. on p. 27).
- [JW18] Pierre-Emmanuel Jabin and Zhenfu Wang. “Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels”. In: *Inventiones mathematicae* 214.1 (July 2018), pp. 523–591. DOI: [10.1007/s00222-018-0808-y](https://doi.org/10.1007/s00222-018-0808-y). URL: <https://doi.org/10.1007/s00222-018-0808-y> (cit. on p. 4).
- [KMV+16] Stanislav Kondratyev, Léonard Monsaingeon, Dmitry Vorotnikov, et al. “A new optimal transport distance on the space of finite Radon measures”. In: *Advances in Differential Equations* 21.11/12 (2016), pp. 1117–1164 (cit. on p. 5).
- [KMX15] Kinderlehrer, Monsaingeon, and Xu. “A Wasserstein gradient flow approach to Poisson–Nernst–Planck equations”. In: *arXiv preprint arXiv:1501.04437v3* (2015) (cit. on p. 17).
- [KNS16] Hamed Karimi, Julie Nutini, and Mark Schmidt. “Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak–Lojasiewicz Condition”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Paolo Frasconi, Niels Landwehr, Giuseppe Manco, and Jilles Vreeken. Cham: Springer International Publishing, 2016, pp. 795–811. ISBN: 978-3-319-46128-1 (cit. on p. 5).
- [LY86] Peter Li and Shing Tung Yau. “On the parabolic kernel of the Schrödinger operator”. In: (1986) (cit. on p. 21).
- [MB01] Andrew J. Majda and Andrea L. Bertozzi. *Vorticity and Incompressible Flow*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2001. DOI: [10.1017/CB09780511613203](https://doi.org/10.1017/CB09780511613203) (cit. on pp. 8, 10).
- [MMN18] Song Mei, Andrea Montanari, and Phan-Minh Nguyen. “A mean field view of the landscape of two-layer neural networks”. In: *Proceedings of the National Academy of Sciences* 115.33 (July 2018). DOI: [10.1073/pnas.1806579115](https://doi.org/10.1073/pnas.1806579115). URL: <https://doi.org/10.1073/pnas.1806579115> (cit. on p. 3).
- [MMS09] D. Matthes, R. McCann, and G. Savaré. “A Family of Nonlinear Fourth Order Equations of Gradient Flow Type”. In: *arXiv preprint arXiv:0901.0540v1* (2009) (cit. on p. 17).
- [Ott01] Felix Otto. “The geometry of dissipative evolution equations : The porous medium equation”. In: *Communications in Partial Differential Equations* 26.1-2 (2001), pp. 101–174. DOI: [10.1081/PDE-100002243](https://doi.org/10.1081/PDE-100002243). eprint: <http://dx.doi.org/10.1081/PDE-100002243>. URL: <http://dx.doi.org/10.1081/PDE-100002243> (cit. on p. 26).

- [Pap04] Papadimitrakis. “Notes on classical potential theory”. In: (2004) (cit. on p. 29).
- [San15] Filippo Santambrogio. “Optimal transport for applied mathematicians”. In: *Birkhäuser, NY* 55.58-63 (2015), p. 94 (cit. on pp. 14, 26, 28).
- [Ser15] Serfaty. *Coulomb Gases and Ginzburg-Landau Vortices*. 2015. arXiv: [1403.6860](https://arxiv.org/abs/1403.6860) [math-ph] (cit. on p. 4).
- [Ser16] Serfaty. “Large systems with Coulomb interactions: Variational study and statistical mechanics”. In: *Portugaliae Mathematica* 73 (2016), pp. 247–278. URL: <https://api.semanticscholar.org/Co> (cit. on p. 4).
- [Ste21] Stefan Steinerberger. “A Wasserstein inequality and minimal Green energy on compact manifolds”. In: *Journal of Functional Analysis* 281.5 (2021), p. 109076. ISSN: 0022-1236. DOI: <https://doi.org/10.1016/j.jfa.2021.109076>. URL: <https://www.sciencedirect.com/science/ai> (cit. on p. 30).
- [Var90] N.Th Varopoulos. “Small time Gaussian estimates of heat diffusion kernels. II. The theory of large deviations”. In: *Journal of Functional Analysis* 93.1 (1990), pp. 1–33 (cit. on p. 21).
- [Wat94] Watson. “Applications of geometric measure theory to the study of Gauss-Weierstrass and Poisson integrals”. In: *Annales Academiæ Scientiarum Fennicæ* 19.3 (194), f115–132 (cit. on pp. 18, 19).

APPENDIX A. WASSERSTEIN GRADIENT FLOWS

Let us recall the definition of the Wasserstein distance W_2 [San15; AGS05].

Definition A.1. Let (X, d) be a metric space. We denote $\mathcal{P}_2(X)$ as the set of probability measures on X with bounded second moment, i.e. such that $\int d(x, x_0)^2 d\mu(x) < \infty$ for some $x_0 \in X$. The 2-Wasserstein distance W_2 between two measures $\mu, \nu \in \mathcal{P}_2(X)$ is defined by:

$$W_2^2(\mu, \nu) := \min_{\gamma \in \Gamma(\mu, \nu)} \int d(x, y)^2 d\gamma(x, y), \quad (\text{A.1})$$

where, if we denote π_i as the projection on the i -th coordinate, $\Gamma(\mu, \nu) = \{\gamma \in \mathcal{P}_2(X \times X) \mid \pi_1 \# \gamma = \mu, \pi_2 \# \gamma = \nu\}$ is the space of transport plans between μ and ν .

In this paper, we only consider X to be the Euclidean space or a Riemannian manifold. In these cases, the metric space $(\mathcal{P}_2(X), W_2)$ is a complete geodesic space. The following results are about some properties of continuous curves in $\mathcal{P}_2(X)$ [San15; AGS05].

Definition A.2 (Continuity equation). Let μ_t be a time-indexed family of measures on \mathbb{R}^d and v_t a time-dependent μ -integrable vector field. The curve μ_t is said to satisfy the continuity equation associated with v_t if:

$$\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0,$$

in the sense of distributions.

The continuity equation above means that along the curve μ_t , mass moves along the vector field v_t .

Definition A.3 (Absolutely continuous curves). A curve of measures μ_t defined for $t \in [a, b]$ is said to be absolutely continuous if there exists some function $m \in L_2(\mathbb{R})$ such that for all $a \leq s \leq t \leq b$:

$$W_2(\mu(s), \mu(t)) \leq \int_s^t m(\tau) d\tau.$$

Proposition A.4. A curve μ_t satisfies a continuity equation if and only if it is absolutely continuous.

This structure allows some Riemannian-like calculus on the geodesic space $(\mathcal{P}_2(X), W_2)$ as introduced by Otto [Ott01]. Indeed, in Definition A.2, the vector field v_t is to be seen as the time derivative of the curve μ_t , taking value in the tangent space at μ_t for the Wasserstein metric. In some diffusive cases, the time derivative cannot be described through the action of a single vector field at all times but rather, informally, a multivalued vector field taking value in a product space. This is why plans are used in some definitions below (see [AGS05, p. 12.4] for more details on the geometry of Wasserstein spaces). Let us introduce subdifferentials in the Wasserstein space [AGS05, Chap 10].

Definition A.5 (Extended Fréchet subdifferential). Consider a probability measure μ and a functional $\mathcal{F} : \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$. A plan $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ belongs to the Fréchet subdifferential $\succ \partial \mathcal{F}(\mu)$ of \mathcal{F} at μ if $\pi_1 \# \gamma = \mu$ and if for every probability measure ρ :

$$\mathcal{F}(\rho) - \mathcal{F}(\mu) \geq \inf_{\alpha \in \Gamma(\gamma, \rho)} \int_{X^3} \langle x_2, x_3 - x_1 \rangle d\alpha + o(W_2^2(\mu, \rho)). \quad (\text{A.2})$$

In some cases, a transport plan $\gamma \in \succ \partial \mathcal{F}(\mu)$ may be concentrated on the graph of a vector field, being of the form:

$$\gamma = (Id \times \xi) \# \mu,$$

for a vector field $\xi \in L_2(\mu)$. Consequently, the subdifferential $\partial \mathcal{F}(\mu)$ is defined as follows:

Definition A.6. Let \mathcal{F} be a functional on \mathcal{P}_2 and $\mu \in \mathcal{P}_2$. A vector field $\xi \in L_2(\mu)$ belongs to the subdifferential $\partial \mathcal{F}(\mu)$ of \mathcal{F} at μ if for every probability measure ρ

$$\mathcal{F}(\rho) - \mathcal{F}(\mu) \geq \inf_{\gamma_0 \in \Gamma_0(\mu, \rho)} \int \xi(x) \cdot (y - x) d\gamma_0(x, y) + o(W_2^2(\mu, \rho)),$$

where $\Gamma_0(\mu, \rho)$ is the set of optimal transport plans between μ and ρ .

For a general functional \mathcal{F} and a general probability measure μ , both $\partial\mathcal{F}(\mu)$ and $\succ\partial\mathcal{F}(\mu)$ may be empty.

There are several ways to approach gradient flows in the Wasserstein space. The first one is a direct analogy of gradient flows on a manifold using the previous definitions.

Definition A.7 (Pointwise differential formula). Let μ_t be an absolutely continuous curve in $\mathcal{P}_2(\mathbb{R}^d)$. By Proposition A.4 it is a weak solution of a continuity equation with a time-dependent vector field v_t . The curve μ_t is said to be a gradient flow a functional \mathcal{F} if for almost any $t > 0$:

$$v_t \in -\partial\mathcal{F}(\mu_t). \quad (\text{A.3})$$

This definition is quite strong, and such curves may not exist at all. To construct them, the idea is to use a discrete algorithm approximation. Let \mathcal{F} be a functional defined on \mathcal{P} . Let us fix an initial measure μ_0 , $\tau > 0$ and consider the following discrete recursive scheme, called Minimizing Movement or JKO steps ([AGS05; JKO98]):

$$\mu_{k+1}^\tau \in \arg \min_{\rho \in \mathcal{P}_2} \mathcal{F}(\tau, \mu_k; \rho) := \mathcal{F}(\rho) + \frac{1}{2\tau} W_2^2(\rho, \mu_k^\tau). \quad (\text{A.4})$$

It consists in updating μ_{k+1} using the proximal function $\mathcal{F}(\tau, \mu_k; \cdot)$. We assume that the associated sequence $(\mu_k^\tau)_{k \in \mathbb{N}}$ can be constructed and we consider the piecewise constant curves defined, if $t \in [k\tau, (k+1)\tau[$ by $U_\tau(t) = \mu_k^\tau$.

Definition A.8 (Minimizing movement curve). Let μ_0 an initial probability measure and \mathcal{F} a functional such that the associated sequence $(\mu_k^\tau)_{k \in \mathbb{N}}$ can be built. A curve μ_t is said to be a minimizing movement curve if there exists some sequence $(\tau_k) \searrow 0$ such that (U_{τ_k}) converges narrowly to μ_t .

There is no a priori guarantee of uniqueness in the definition above: to a functional \mathcal{F} and an initial probability measure μ_0 can correspond an infinity of minimizing movements.

Let us make the following assumptions on \mathcal{F} :

- (1) \mathcal{F} is proper (not everywhere $+\infty$) and lower semi continuous for the weak topology.
- (2) Coercivity: there exists some $\tau_0 > 0$ such that for all $\tau_0 > \tau > 0$ and $\mu \in \mathcal{P}_2$, there exists some probability measure μ_τ minimizing the proximal function $\mathcal{F}(\tau, \mu; \cdot)$.

Using [AGS05, Proposition 11.1.6], we get:

Theorem A.9. *With the above assumptions, a Minimizing Movement curve always exists.*

Why are we interested in such curves? The iterations of the discrete scheme (A.4) satisfy an important regularity property, they are point of subdifferentiability of \mathcal{F} [AGS05, Th 10.3.4, Remark 10.3.5].

Proposition A.10. *Let μ be a probability measure and \mathcal{F} a functional such that we can define μ_τ as an iteration of (A.4) from μ . Then $\partial\mathcal{F}(\mu_\tau)$ is not empty.*

Let $\hat{\gamma}_\tau \in \Gamma_0(\mu_\tau, \mu)$. The rescaled velocity plan $\gamma_\tau := f_\tau \# \hat{\gamma}_\tau$, where $f_\tau(x_1, x_2) = (x_1, \frac{x_2 - x_1}{\tau})$, is in the extended Fréchet subdifferential $\succ\partial\mathcal{F}(\mu)$. Moreover, there exists a unique optimal plan such that its barycenter projection $\tilde{\gamma}_\tau^0$ is in the subdifferential $\partial\mathcal{F}(\mu)$. It is characterized by the strictly convex minimum condition:

$$\|\tilde{\gamma}_\tau^0\|_{L_2(\mu_\tau)} = \min_{\tilde{\gamma}_\tau \in \Gamma_0(\mu_\tau, \mu)} \|\tilde{\gamma}_\tau - Id\|_{L_2(\mu_\tau)}.$$

Remark A.11. The differentiation point is μ_τ , not directly μ . This result is to be compared with the fact that in the Euclidean space, the iteration of the implicit gradient descent scheme $x_{k+1} = x_k - \tau \nabla f(x_{k+1})$ can be obtained as:

$$x_{k+1} \in \arg \min_y f(y) + \frac{1}{2\tau} \|y - x_k\|^2. \quad (\text{A.5})$$

Due to Proposition A.10, we would like to pass to the limit as $\tau \rightarrow 0$ and conclude that a Minimizing Movement curve is a gradient flow in the sense of Definition A.7. However, this is not always the case and such a curve only satisfies a relaxed gradient equation, with the time-dependent vector field v_t only belonging to the limiting subdifferential of \mathcal{F} at μ_t [AGS05, Def 11.1.5]. In the case of a functional \mathcal{F} which is λ -convex along generalized geodesics [AGS05, Def 9.2.2, 9.2.4, Th 11.2.1], more can be said.

Theorem A.12 (Gradient flow for λ -convex functionals). *Let \mathcal{F} be a λ -convex functional along generalized geodesics and $\mu_0 \in \mathcal{P}_2$. Then:*

- There exists a unique Minimizing Movement curve starting from μ_0 .
- This limiting curve μ_t is a solution of the gradient flow equation (A.3).
- (EVI) The curve μ_t satisfies the EVI inequality, for all $\nu \in \mathcal{P}_2$

$$\frac{1}{2} \frac{d}{dt} W_2^2(\mu_t, \nu) \leq \mathcal{F}(\nu) - \mathcal{F}(\mu_t) - \frac{\lambda}{2} W_2^2(\mu_t, \nu). \quad (\text{A.6})$$

- If $\lambda > 0$, \mathcal{F} admits a unique minimum μ_* and both μ_t and $\mathcal{F}(\mu_t)$ converge exponentially respectively to μ_* and $\mathcal{F}(\mu_*)$.
- If $\lambda = 0$ and \mathcal{F} admits a minimum \mathcal{F}_* then

$$\mathcal{F}(\mu_t) - \mathcal{F}_* \leq \frac{W_2(\mu_0, \mu_*)}{2t}.$$

Example A.13. • If \mathcal{F} is defined on measures with density with respect to the Lebesgue measure ($\mu = \rho(x)dx$) by the formula $\mathcal{F}(\mu = \rho(x)dx) := \int F(\rho)dx$ where F is convex, superlinear, satisfies $F(0) = 0$, and that the map $s \mapsto s^d F(s^{-d})$ is convex and non increasing, then \mathcal{F} is convex along generalized geodesics [San15, Th 7.28]. For example, this condition is verified for $H : x \mapsto x \log(x)$ which gives the Boltzmann entropy \mathcal{H} , and $F_p : x \mapsto x^p$ where $p > 1$, defining p-energies.

- If $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is λ -convex, the associated potential functional $\mathcal{V} : \mu \mapsto \int V d\mu$ is λ -convex along geodesics.
- If $W : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is λ -convex, the associated auto-interaction functional $\mathcal{W} : \mu \mapsto \int W d\mu^{\otimes 2}$ is λ -convex along geodesics.
- The previous condition is not necessary. In dimension 1, let us define $\Delta_+ := \{(x, y), x \leq y\}$ and $\Delta_- := \{(x, y), x \geq y\}$. Then if W is convex when restricted to Δ_+ and Δ_- , the associated energy \mathcal{W} is convex along generalized geodesics. That is the case for the Energy Distance kernel $(x, y) \mapsto -\|x - y\|$.

Corollary A.14 (Convergence of the Energy Distance gradient flow in 1D). *Let $\mu_0, \nu \in \mathcal{P}_2(\mathbb{R})$ with finite first moment. Using the previous results, there exists a unique solution μ_t to the associated gradient flow equation (A.3) for the functional E_ν with the energy distance kernel, obtained as a Minimizing Movement curve from μ_0 . Furthermore:*

$$E_\nu(\mu_t) \leq \frac{W_2(\mu_0, \nu)}{2t}.$$

Proof. Let $x, y \in \mathbb{R}$. If $(x, y) \in \Delta_+$, then $-\|x - y\| = x - y$ and the Energy distance kernel is convex on Δ_+ . If $(x, y) \in \Delta_-$, then $-\|x - y\| = y - x$ and the Energy distance kernel is convex on Δ_- . This proves that $E_\nu(\cdot)$ is convex along Wasserstein geodesics. The result directly follows from Theorem A.12. \square

However, it is important to note that in higher dimensions, the energy is not convex along generalized geodesics.

APPENDIX B. KERNELS, MMD AND POTENTIAL THEORY

In this paper, we are interested in particular functionals on probability measure spaces: MMD-energies.

Definition B.1. Let G be a conditionally positive kernel (possibly taking infinite values).

- The associated internal G -energy functional of a signed measure ρ is defined by:

$$E(\rho) := \frac{1}{2} \langle \rho, G \star \rho \rangle.$$

- If μ and ν are probability measures of finite internal energy, we define the functional:

$$E_\nu(\mu) := E(\mu - \nu).$$

The Maximum Mean Discrepancy of kernel G between μ and ν is defined as:

$$MMD(\mu, \nu) := \sqrt{E(\mu - \nu)}.$$

Maximum Mean Discrepancies have been studied in the context of Wasserstein gradient flows, but mainly in smooth cases [Arb+19].

B.1. In the Euclidean Space. A particular class of singular kernels o Riesz kernels.

Definition B.2 (Riesz kernels). In the Euclidean space \mathbb{R}^d , $d \in \mathbb{N}$, Riesz kernels are defined by:

$$k_s(x, y) = \frac{1}{s \|x - y\|^s},$$

for $s \in [-1, d - 2] \setminus \{0\}$. If $s = 0$, the kernel is defined as $k_0(x, y) = -\log(\|x - y\|)$.

In this article, we focus on two kernels:

The Coulomb kernel. In \mathbb{R}^d , the Coulomb Kernel is given by $k_{d-2}(x, y) = \frac{1}{(d-2)\|x-y\|^{d-2}}$. It is remarkable as a fundamental solution of the Laplace equation in \mathbb{R}^d in the following sense.

Proposition B.3 (Corresponding differential operator). *Let G be the Coulomb Kernel in \mathbb{R}^d . There exists a positive constant c_d such that G is the fundamental solution of $\frac{1}{c_d}(-\Delta)^{\frac{d-s}{2}}$ in \mathbb{R}^d , which is, if $\Delta_x G(x, y)$ is equal to the Laplacian of $x \mapsto G(x, y)$:*

$$\frac{1}{c_d}(-\Delta_x)G(x, y) = \delta_y.$$

Definition B.4. For a kernel G and a measure ρ , we define the associated potential by $\varphi_\rho^G := G \star \mu$ when it is well-defined.

When there is no ambiguity on the kernel G or the measure ρ we will only write φ instead of φ_ρ^G . In the Coulomb case, this potential corresponds to the electric field generated by a distribution μ of electric charges.

Proposition B.5 (Coulomb kernel and Laplace equation). *Let G be the Coulomb kernel in \mathbb{R}^d and μ be a positive measure. Then, the associated potential φ_μ satisfies the equation:*

$$\frac{1}{c_d}(-\Delta)\varphi_\mu = \mu.$$

This equation establishes that, for the Coulomb kernel, the potential φ_μ is harmonic outside of the support of μ , and superharmonic in \mathbb{R}^d [Pap04].

The Energy Distance kernel. The Energy Kernel is independent of the dimension and defined by $k_{-1}(x, y) = -\|x - y\|$. It has the good property that any measure with a finite first moment has finite internal energy. This is to compare with the Coulomb kernel, where additional regularity is required, e.g. finite H_{-1} norm.

Proposition B.6 (Convexity on probability measures). *Let G be the Coulomb or Energy Distance kernel and $\nu \in \mathcal{P}(\mathbb{R}^d)$ with finite G -energy. Then the associated functional $E_\nu(\cdot) = E(\cdot - \nu)$ is a quadratic functional (in μ) that is strictly convex (for the convex structure on $\mathcal{P}(\mathbb{R}^d)$) on its domain. It is positive and equal to 0 if and only if $\mu = \nu$. Moreover, the application $(\mu, \nu) \mapsto MMD(\mu, \nu) := \sqrt{E(\mu - \nu)}$ defines a distance on probability measures with finite G -energy*

B.2. Extension to Riemannian Manifolds. We focus on compact manifolds without boundaries to avoid possible loss of mass at infinity in the dynamics. Coulomb-like interactions among probability measures can be defined through the fundamental solution of the Laplace equation on some Riemannian manifolds. We use the formalism developed in [Gar19] and [Ste21], along with results from [Aub98].

Let (M, g) be a compact oriented n -dimensional Riemannian manifold without boundary where g is the Riemannian metric. We denote π its volume form, assuming $\pi(M) = 1$. Additionally, $\Delta_M : C^\infty(M) \rightarrow C^\infty(M)$ is the associated Laplace-Beltrami operator.

Definition B.7 (Green's function on a manifold). A kernel $G : M \times M \rightarrow]-\infty, +\infty]$ is said to be a Green function if it is symmetric, if $G_x : y \in M \mapsto G(x, y)$ is integrable for all $x \in M$, and if it satisfies the Laplacian equation:

$$-\Delta_M G_x = -\delta_x + 1, \quad (\text{B.1})$$

in a distributional sense.

The following existence and quasi-uniqueness result is given in [Aub98, Chapter 4].

Proposition B.8. *Let (M, g) as defined previously. Then, equation (B.1) admits a unique solution, up to an additive constant.*

In particular, if μ is a measure on M , we can define the potential $\varphi_\mu(\mu)(x) := \int G(x, y) d\mu(y)$. Then, in the sense of distributions $\Delta_M \varphi_\mu = -\mu + \pi$. In particular, G is bounded from below and $\int G(x, y) d\pi(y)$ does not depend on x . We denote by G the unique Green function such that $\int G(x, y) d\pi(y) = 0$. The Green's function G is lower semi-continuous, so that the functional defined by:

$$\mathcal{W}(\mu) = \iint_{M \times M} G(x, y) d\mu(x) d\mu(y) \quad (\text{B.2})$$

is lower semi-continuous for the weak topology. Furthermore, the kernel G is C^∞ outside of the diagonal $\mathcal{D} := \{(x, x) \mid x \in M\}$. Now, let ν be a density probability measure on M . The energy functional is defined similarly:

$$E_\nu(\mu) := \frac{1}{2} \iint_{M \times M} G(x, y) d(\mu - \nu)^{\otimes 2}(x, y).$$

This energy is lower semi-continuous, positive, and equal to 0 if and only if $\mu = \nu$. As in the Euclidean case, the square root is a distance between μ and ν [Car45; Gar19].

APPENDIX C. SOME PROOFS

Proof of Lemma 2.4. Let us fix a vector field $v \in \mathcal{C}_c^\infty$, and prove the following equality in both cases:

$$\lim_{t \rightarrow 0} \frac{E_\nu((id + tv)\#\mu) - E_\nu(\mu)}{t} = \int (\nabla G) \star (\mu - \nu) \cdot v d\mu. \quad (\text{C.1})$$

First case: If $G = -\|\cdot\|$. First let us remark that if $x = y$, $G(x - y + t(v(x) - v(y))) = G(x - y) = 0$. Then:

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{1}{2} \int \frac{G(x - y + t(v(x) - v(y))) - G(x - y)}{t} d\mu(x) d\mu(y) \\ &= \lim_{t \rightarrow 0} \frac{1}{2} \int_{x \neq y} \frac{G(x - y + t(v(x) - v(y))) - G(x - y)}{t} d\mu(x) d\mu(y) \\ &= \frac{1}{2} \int_{x \neq y} \nabla G(x - y) \cdot (v(x) - v(y)) d\mu(x) d\mu(y) \\ &= \int_{x \neq y} \nabla G(x - y) \cdot v(x) d\mu(x) d\mu(y). \end{aligned}$$

Same for the other term of the MMD energy:

$$\begin{aligned}
& \lim_{t \rightarrow 0} \frac{1}{2} \int \frac{G(x-y+tv(x)) - G(x-y)}{t} d\mu(x) d\nu(y) \\
&= \lim_{t \rightarrow 0} \int_{x \neq y} \frac{G(x-y+t(v(x)-v(y))) - G(x-y)}{t} d\mu(x) d\mu(y) \\
&= \int_{x \neq y} \nabla G(x-y) \cdot v(x) d\mu(x) d\nu(y) \\
&= \int_{x \neq y} \nabla G(x-y) \cdot v(x) d\mu(x) d\nu(y).
\end{aligned}$$

We conclude that Formula (C.1) holds for the Energy Distance kernel.

Second case: If $G = \frac{1}{\|\cdot\|^{d-2}}$. Here, we need to use the result that if a positive measure has finite Coulomb energy, then it cannot be too singular, i.e. if we define the diagonal $\mathcal{D} := \{x = y\} \subset X \times X$ then $\mu^{\otimes 2}(\mathcal{D}) = 0$. Indeed, $\int_{\mathcal{D}} G d\mu^{\otimes 2} \leq \int G d\mu^{\otimes 2} < \infty$, and $G = +\infty$ on \mathcal{D} , which proves $\mu^{\otimes 2}(\mathcal{D}) = 0$. This means that for any measure μ with finite energy:

$$\int G d\mu^{\otimes 2} = \int_{\mathcal{D}^c} G d\mu^{\otimes 2}.$$

Thus we can conclude that Formula (C.1) holds for the Coulomb kernel with similar computations, as G is \mathcal{C}^∞ on \mathcal{D}^c . We write:

$$\frac{E_\nu((id+tv)\#\mu) - E_\nu(\mu)}{t} = \frac{E_\nu((id+tv)\#\mu) - E_\nu(\mu)}{W_2((id+tv)\#\mu, \mu)} \frac{W_2((id+tv)\#\mu, \mu)}{t}.$$

By definition of the slope, we get:

$$\limsup_{t \rightarrow 0} \frac{E_\nu((id+tv)\#\mu) - E_\nu(\mu)}{W_2((id+tv)\#\mu, \mu)} \leq |\partial E_\nu|(\mu).$$

Additionally,

$$\limsup_{t \rightarrow 0} \frac{W_2((id+tv)\#\mu, \mu)}{t} \leq \|v\|_{L_2(\mu)},$$

so that, taking the limit for $t \rightarrow 0$ we get:

$$\int_{x \neq y} (\nabla G) \star (\mu - \nu) \cdot v d\mu \leq |\partial E_\nu|(\mu) \|v\|_{L_2(\mu)}.$$

Similarly, using $-v$ instead of v , we have:

$$\left| \int_{x \neq y} (\nabla G) \star (\mu - \nu) \cdot v d\mu \right| \leq |\partial E_\nu|(\mu) \|v\|_{L_2(\mu)}.$$

As the vector field v chosen is arbitrary, we get:

$$\|(\nabla G) \star (\mu - \nu)\|_{L_2(\mu)} \leq |\partial E_\nu|(\mu),$$

which proves the result. \square

Proof of Lemma 2.5. Let ρ be a measure, and $id + v$ an optimal transport between μ and ν (so that $\rho = (id + v)\#\mu$) which exists as $\mu \in \mathcal{P}_2^r$. We note $\xi := \nabla G \star (\mu - \nu)$, which is a well-defined vector field everywhere. Using the same computations as in the proof of equality (C.1), we get:

$$E_\nu(\rho) - E_\nu(\mu) = \int \xi \cdot v d\mu + o(\|v\|_{L_2(\mu)}^2).$$

However, as v is an arbitrary optimal transport plan between μ and ρ and μ is regular, we get:

$$E_\nu(\rho) - E_\nu(\mu) \geq \inf_{\gamma_0 \in \Gamma_0(\mu, \rho)} \int \xi \cdot (y - x) d\gamma_0 + o(W_2^2(\mu, \rho)),$$

which concludes the proof by Definition A.6. \square