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. 2023. hal-04282762v1

HAL Id: hal-04282762

https://hal.science/hal-04282762v1

Preprint submitted on 20 Nov 2023 (v1), last revised 26 Jan 2024 (v2)

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 Boufadène[◇,△] and François-Xavier Vialard[◇*]

△ INRIA Paris,
◇ LIGM, Univ. Gustave Eiffel, CNRS, F-77454 Marne-la-Vallée, France.

Abstract

In this work, we study the Wasserstein gradient flow of the Riesz energy defined on the space of probability measures. The Riesz kernels define a quadratic functional on the space of measure which is not in general geodesically convex in the Wasserstein geometry, therefore one cannot conclude to global convergence of the Wasserstein gradient flow using standard arguments. Our main result is the exponential convergence of the flow to the minimizer on a closed Riemannian manifold under the condition that the logarithm of the source and target measures are Hölder continuous. To this goal, we first prove that the 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 space or on a closed Riemannian manifold. For general measures, we prove using flow interchange techniques that there is no local minima other than the global one for the Coulomb kernel. In fact, we prove that a Lagrangian critical point of the functional for the Coulomb (or Energy distance) kernel is equal to the target everywhere except on singular sets with empty interior. In addition, singular enough measures cannot be critical points.

Keywords. Non-convex optimization · Calculus of variation · Optimal transportation · Coulomb kernel · Flow interchange

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

*Email addresses: siwan.boufadene@ens-paris-saclay.fr, francois-xavier.vialard@univ-eiffel.fr.

Contents

1 Introduction

In this paper, we are interested in Wasserstein gradient flow of certain quadratic functionals on the set of probability measures on the Euclidean space or a Riemannian manifold. These quadratic functionals are convex for the usual convex structure of probability measures. However they are not geodesically convex for the Wassertein geometry. Therefore, the dynamic of their gradient flows with respect to the Wasserstein geometry may obstruct global convergence properties, which are granted in geodesically convex settings. One motivation for studying these gradient flows under an other geometry comes from machine learning, more precisely the mean-field limit of shallow neural networks studied in [12, 26]. This line of research explores 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 [5] consists in embedding the space of parameters into the space of probability measures. In this context, the corresponding objective functional becomes quadratic and the gradient flow corresponds to a Wasserstein gradient flow. In fact, the quadratic function on the space of probability measures is a particular case of quadratic functionals defined by reproducing kernels. These functionals have also raised interest in machine learning and statistics since it gives a discrepancy, in fact a distance squared, between probability measures. These discrepancies are called Maximum Mean Discrepancy (MMD) and they are defined below. Let us insist on two particularly nice properties of MMD: they have a quadratic computational cost which is better than, for instance, optimal transport and they benefit from a parametric rate of estimation from empirical measures, which is not the case for standard optimal transport.

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

$$E_\nu(\mu) = \iint (\mu(x) - \nu(x))G(x, y)(\mu(y) - \nu(y)) \, . \tag{1}$$

In fact, such a functional is a nonnegative strictly convex functional 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 μ will be optimized upon through the action of velocity fields. More precisely, we are interested in the Wasserstein gradient flow of the functional $E_\nu(\mu)$ with respect to μ. In particular, we want to treat the question of global convergence to the unique minimizer which is ν. For instance, if the kernel is smooth enough, empirical measures are 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 does not give convergence.

This fact motivates the exploration of non-smooth kernels such as the energy distance, namely $-\|x - y\|$ on the Euclidean space, which is not C^1. As a consequence, 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 one dimension, the corresponding functional is geodesically convex in Wasserstein and it implies global convergence of the solution μ_t towards ν. This particular kernel has been studied in the context of Wasserstein gradient flows in [17] and [16] 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 possible generalizations of this result. First, the kernel $-\|x - y\|$ is still conditionally positive definite on $\mathbb{R}^d$ for $d \geq 1$. Second, in one dimension, $-|x - y|$ is the Coulomb kernel, roughly speaking (proportional to) the inverse of the Laplacian, which is well-defined in higher-dimensions. For instance on $\mathbb{R}^3$, it is given by $\frac{1}{\|x-y\|}$. These two kernels belong to the family of Riesz kernels, which will be our main interest in this work. One motivation for using these kernels also comes from numerical experiments, in which the energy distance stands out. Indeed, the energy distance kernel is easy to implement and behaves particularly well, in comparison with other kernels such as the Gaussian kernel. More precisely, global convergence is observed. The Coulomb kernel is more intricate to implement due to its blow-up along the diagonal and drawing

conclusions from these numerical experiments is more delicate. Yet, from the theoretical point of view, the Coulomb kernel has been studied intensively in the mathematical literature, in particular due to its physical significance, see [30,31] for a review. Recent results exposed in [19] and [13] study large stochastic systems of interacting particles under Coulomb interaction, proving propagation of chaos and convergence to the limit PDE results using relative entropy methods. In [13] this work is done on the torus manifold.

For the Coulomb kernel, under 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,

$$\partial_t \mu = -\nabla \cdot \big(\mu \nabla \varphi\big), \qquad \Delta \varphi = \mu - \nu \,. \tag{2}$$

The potential φ is a solution of the previous Poisson equation with source term $\mu - \nu$. The potential φ can be written, up to a positive constant $\varphi = -G \star (\mu - \nu)$, 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 of

$$\frac{\partial \mu}{\partial t} = -\nabla \cdot \big(\mu(\nabla W \star \mu)\big) \,. \tag{3}$$

Confinement results for these dynamics have been proven, depending on W. In most cases, W is supposed to be λ-convex, preserving particles [9], so that mean-field techniques can be used. Attractive potentials, e.g. $w'(r) \geq 0$ everywhere are the simplest ones, and in some case total mass aggregate in its center. Some other potentials are said to be attractive-repulsive, e.g. there exists a radius $R_a > 0$ such as $w'(r) \leq 0$ if $r < R_a$ and $w'(r) \geq 0$ otherwise. That is the case for swarming systems models $w(r) = r^\gamma/\gamma - r^\alpha/\alpha$ where $\gamma > \alpha$ studied in [10], Morse potentials $w(r) = -C_a e^{-r/l_a} + C_r e^{-r/l_r}$ and characteristic function of sets for example [4,8,9] . Different hypothesis are used to prove confinement. In [9], W is continuous and $\lim_{r\to\infty} w'(r)\sqrt{r} = +\infty$. In [4] the weaker condition $\lim_{r\to\infty} w'(r) r = +\infty$ is enough. Some discontinuous potentials at 0 have been treated, for example in [4], potential $W(x) = G(x) + W_a(x)$ where G is the coulomb kernel $G(x) = \|x\|^{-d+2}$ and W_a is an attractive potential that verifies $\lim_{r\to\infty} w_a'(r) r^{1/d} = +\infty$. However these results do not apply to our subject, our functional is not λ-convex and has diffusive properties. Moreover, the confining part (depending on a target measure ν) is weaker than in previously cited papers and we were not able to prove mass confinement.

Main contributions. There are two main results in this paper, which are centered on the Coulomb kernel. The first main result is a proof of global convergence on a closed Riemannian manifold in a smooth setting. In the beginning of Section 2, a simple calculation shows that the Polyak-Lojasiewicz inequality is satisfied for all time on a closed Riemannian manifold under the condition that the solution is sufficiently smooth and that it exists for all time. Note that it implies that the question of (exponential) global convergence is reduced to a regularity question: Do the solutions exist for all time in sufficiently smooth functional spaces? Therefore, we first study the gradient flow of the Coulomb discrepancy in a smooth setting, that is, the corresponding PDE under some regularity assumptions. Under the assumption that initial and target measures are Hölder continuous, we prove that solutions exist for all time and they are Hölder continuous. It thus proves global convergence with an exponential rate of convergence in the case of closed Riemannian manifolds.

The second main result is found in Section 4 and it concerns the landscape of the Coulomb discrepancy on Euclidean spaces as well as on closed Riemannian manifolds: This functional has no local minima apart from the global one in the Wasserstein geometry. More precisely, we prove that if the current measure is different from the target, there always exists a path of measures (starting at the current one) which is 1/2-Hölder in Wassertein and such that the energy is strictly decreasing. This is done using flow interchange techniques, namely the use 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 interior. In a similar direction, we prove that if the difference between the current measure and

the target has Minkowski dimension less 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 in a closed Riemannian manifold for every source and target measures. Although the result seems highly plausible on a closed Riemannian manifold, a corresponding result on the Euclidean space would need to get around the confinement issue, which we have not proven in a Euclidean setting. Indeed, there is a competition between the repulsive behaviour of the Coulomb kernel and the attraction of the target measure. This repulsive behaviour can cause mass spread to infinity in Euclidean space, which makes the analysis more difficult in our opinion.

1.1 Notations

- If A is a subset of $\mathbb{R}^d$, $A^c \coloneqq \mathbb{R}^d \setminus A$ is its complementary
- If μ is a measure on $(\mathbb{R}^d, \mathcal{T}_1)$ and $T : (\mathbb{R}^d, \mathcal{T}_1) \to (\mathbb{R}^d, \mathcal{T}_2)$ is a measurable map then the pushforward measure of μ by T, denoted $T\#\mu$ verifies, for any $\mathcal{T}_2$ measurable function f

$$\int f dT\#\mu = \int f \circ T d\mu \,.$$

- $\mu^{\otimes 2}$ is the product measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that, for any Borel sets A, B, we have

$$\mu^{\otimes 2}(A \times B) = \mu(A)\mu(B).$$

- C_c^∞ is the space of test functions, i.e. of infinitely differentiable functions with compact support.
- The convolution is denoted by $\star$.
- If (M, g) is a Riemannian manifold, we denote $d_M(x, y)$ the geodesic distance between $x, y \in M$.

1.2 Wasserstein gradient flows

Let us recall the definition of the Wasserstein distance W_2, see [1, 29].

Definition 1. *Let (X, d) be a metric space. We denote $\mathcal{P}_2(X)$ 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*

$$W_2^2(\mu, \nu) \coloneqq \min_{\gamma \in \Gamma(\mu,\nu)} \int d(x, y)^2 d\gamma(x, y) \,, \tag{4}$$

where, if denote π_i the projection on the i-th coordinate, $\Gamma(\mu, \nu) = \{\gamma \in \mathcal{P}_2(X \times X) \,|\, \pi_1\#\gamma = \mu, \pi_2\#\gamma = \nu\}$ 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 those 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 $\mathcal{P}_2(X)$; see [1, 29].

Definition 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 curves μ_t is said to satisfy the continuity equation associated with v_t if*

$$\partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \,, \tag{5}$$

in the sense of distributions.

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

Definition 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 $a \leq s \leq t \leq b$*

$$W_2(\mu(s), \mu(t)) \leq \int_s^t m(\tau) d\tau \,. \tag{6}$$

Proposition 1. 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 [27]. Indeed, in definition 2, the vector field v_t can be used to generate the tangent space at μ_t, by taking the minimal norm of v reproducing $\partial_t \mu$. In some cases however, the whole tangent space cannot be described through the action of a single vector field, but rather, informally, a multivalued vector field. This is why plans are used in some definitions below (see [1, 12.4] for more details on the geometry of Wasserstein spaces). Let us introduce subdifferentials in the Wasserstein space, see [1, Chap 10].

Definition 4 (Extended Fréchet subdifferential)**.** *Let us consider a probability measure μ and a functional $\mathcal{F} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$. A plan $\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ belongs to the Fréchet subdifferential $\boldsymbol{\partial}\mathcal{F}(\mu)$ of $\mathcal{F}$ at μ if $\pi_1 \# \gamma = \mu$ and if for every probability measure ρ*

$$\mathcal{F}(\rho) - \mathcal{F}(\mu) \geq \inf_{\alpha \in \Gamma(\gamma, \rho)} \int_{X^3} \langle x_2, x_3 - x_1 \rangle d\alpha + o(W_2^2(\mu, \rho)). \tag{7}$$

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

$$\gamma = (Id \times \xi) \# \mu \tag{8}$$

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

Definition 5. *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*

$$\mathcal{F}(\rho) - \mathcal{F}(\mu) \geq \inf_{\gamma_0 \in \Gamma_0(\mu, \rho)} \int \xi(x) \cdot (y - x) d\gamma_0(x, y) + o(W_2^2(\mu, \rho)) \,, \tag{9}$$

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 $\boldsymbol{\partial}\mathcal{F}(\mu)$ 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 manifold using the previous definitions.

Definition 6 (Pointwise differential formula)**.** *Let μ_t be an absolutely continuous curve in $\mathcal{P}_2(\mathbb{R}^d)$. By Proposition 1 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 $t > 0$*

$$v_t \in -\partial\mathcal{F}(\mu_t). \tag{10}$$

This definition is quite strong, and such curves may not exist at all. To build those, 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 ([1, 20]).

$$\mu_{k+1}^\tau \in \operatorname*{arg\,min}_{\rho \in \mathcal{P}_2} \mathcal{F}(\tau, \mu_k; \rho) \coloneqq \mathcal{F}(\rho) + \frac{1}{2\tau} W_2^2(\rho, \mu_k^\tau) \,. \tag{11}$$

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 built and we consider the piecewise constant curves defined, if $t \in [k\tau, (k+1)\tau[$ by $U_\tau(t) = \mu_k^\tau$.

Definition 7 (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 a priori no uniqueness guarantee 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 $\mathcal{F}(\tau, \mu; \cdot)$.

Then using [1, 11.1.6], one has:

Theorem 1. With the above assumptions, a Minimizing Movement curve always exists.

Why are we interested in such curves ? The iterations of the discrete scheme (11) satisfy an important regularity property, they are point of subdifferentiability of $\mathcal{F}$, see [1, Th 10.3.4; Remark 10.3.5].

Proposition 2. Let μ be a probability measure and $\mathcal{F}$ a functional such that we can define μ_τ as an iteration of (11) 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 $\partial\mathcal{F}(\mu)$. 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

$$\|\tilde{\gamma}_\tau^0\|_{L_2(\mu_\tau)} = \min_{\hat{\gamma}_\tau \in \Gamma_0(\mu_\tau, \mu)} \|\tilde{\gamma}_\tau - Id\|_{L_2(\mu_\tau)} \,. \tag{12}$$

Remark 1. The differentiation point is μ_τ, and 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

$$x_{k+1} \in \arg\min_y \; f(y) + \frac{1}{2\tau}\|y - x_k\|^2 \,. \tag{13}$$

Because of the result of Proposition 2, we would like to pass to the limit as $\tau \to 0$, and conclude that a Minimizing Movement curve is a gradient flow in the sense of definition 6. However, that 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 (see [1, Def 11.1.5]). In the case of a functional $\mathcal{F}$ which is λ-convex along generalized geodesics (see [1, Def 9.2.2, 9.2.4; Th 11.2.1]), more can be said.

Theorem 2 (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 (10).
- (EVI) The curve μ_t satisfies the EVI inequality, for all $\nu \in \mathcal{P}_2$

$$\frac{1}{2}\frac{d}{dt}W_2^2(\mu_t, \nu) \le \mathcal{F}(\nu) - \mathcal{F}(\mu_t) - \frac{\lambda}{2}W_2^2(\mu_t, \nu) \,. \tag{14}$$

- 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

$$\mathcal{F}(\mu_t) - \mathcal{F}_* \leq \frac{W_2(\mu_0, \mu_*)}{2t}. \tag{15}$$

Example 1.
- 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, verifies $F(0) = 0$, and that the map $s \mapsto s^d F(s^{-d})$ is convex and non increasing, then $\mathcal{F}$ is convex along generalized geodesics (see [29, 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 $V : \mathbb{R}^d \to \mathbb{R}$ is λ-convex, the associated potential functional $\mathcal{V} : \mu \mapsto \int V d\mu$ is λ-convex along geodesics.
- If $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ 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 1 (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 of the associated gradient flow equation (10) for the functional E_ν with the energy distance kernel, obtained as a Minimizing Movement curve from μ_0. Moreover

$$E_\nu(\mu_t) \leq \frac{W_2(\mu_0, \nu)}{2t}. \tag{16}$$

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 is a direct application of theorem 2. □

However, in higher-dimensions, the energy is not convex along generalized geodesics.

1.3 Kernels, MMD and Potential theory

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

Definition 8. *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*

$$E(\rho) := \frac{1}{2}\langle \rho, G \star \rho \rangle. \tag{17}$$

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

$$E_\nu(\mu) := E(\mu - \nu). \tag{18}$$

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

$$MMD(\mu, \nu) := \sqrt{E(\mu - \nu)}. \tag{19}$$

Maximum Mean Discrepancy have been studied in the context of Wasserstein gradient flows, but mainly in smooth cases, see [2]. A particular class of singular kernels are Riesz kernels.

Definition 9 (Riesz kernels)**.** *In the Euclidean space $\mathbb{R}^d$, $d \in \mathbb{N}$, the Riesz kernels are defined by*

$$k_s(x,y) = \frac{1}{s\|x-y\|^s}\,, \tag{20}$$

for $s \in [-1, d-2]/\{0\}$. If $s = 0$, we define the kernel $k_0(x,y) = -\log(\|x-y\|)$.

In this article, we consider mainly two kernels:

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

Proposition 3 (Corresponding differential operator)**.** Let G be the Coulomb Kernel in $\mathbb{R}^d$. There exists some 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

$$\frac{1}{c_d}(-\Delta_x)G(x,y) = \delta_y\,. \tag{21}$$

Definition 10. *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 4 (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

$$\frac{1}{c_d}(-\Delta)\varphi_\mu = \mu\,. \tag{22}$$

This equation proves that for the Coulomb kernel, the potential φ_μ is harmonic outside of the support of μ, and superharmonic in $\mathbb{R}^d$, see [28].

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

Proposition 5 (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) which 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

1.4 Extension to Riemannian manifolds

We are interested in compact manifolds without boundaries since this case avoids possible loss of mass at infinity in the gradient flow. Coulomb-like interactions between probability measures can be defined by fundamental solution of the Laplace equation on some Riemannian manifolds. We will use the formalism developed in [14] and [32] and results from [3].

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 $\pi(M) = 1$ and $\Delta_M : C^\infty(M) \to C^\infty(M)$ is the associated Laplace-Beltrami operator.

Definition 11 (Green's function on a manifold)**.** *A kernel $G : M \times M \to]-\infty, +\infty]$ 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*

$$-\Delta_M G_x = -\delta_x + 1\,, \tag{23}$$

in a distributional sense.

We have the following existence result, given in [3, Chapter 4].

Proposition 6. Let (M, g) as defined previously. Then, equation (23) 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 $\Delta_M \varphi_\mu = -\mu + \mu(M)\pi$. 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 as $\int G(x,y)d\pi(y) = 0$. The Green's function G is lower semicontinuous, so that the functional defined by

$$\mathcal{W}(\mu) = \iint_{M\times M} G(x,y)d\mu(x)d\mu(y) \tag{24}$$

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

$$E_\nu(\mu) := \frac{1}{2}\iint_{M\times M} G(x,y)d(\mu-\nu)^{\otimes 2}(x,y)\,. \tag{25}$$

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, see [11, 14].

2 Polyak-Lojasiewicz Inequality and exponential convergence

Our goal is to prove that a Wasserstein gradient flow curve μ_t converges to the target ν. In finite dimension, a standard condition is the Polyak-Lojasiewicz inequality.

Definition 12. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. It is said to satisfy the Polyak-Lojasiewicz with parameter $\lambda > 0$ if*

$$\frac{1}{2}\|\nabla f\|_2^2 \geq \lambda(f(x) - f^*)\,. \tag{26}$$

This condition is actually weaker than a lot of other classical conditions, such as strong convexity, weak strong convexity, or the restricted secant inequality (see [21] for a review). With this condition, an exponential convergence rate to the global minimum can be shown.

For our purpose, the key point of the Polyak-Lojasiewicz inequality is that it applies to more general settings such as Riemannian manifolds. Importantly, it can also be applied to the space of probability measures endowed with the Wasserstein metric. Let us give an example: As is well-known, the log-Sobolev inequality can be interpreted as a Polyak-Lojasiewicz inequality and exponential convergence of the solution of the Fokker-Planck equation to the equilibrium measure can be formulated in this language.

Proposition 7. Let $f : \mathbb{R}^d \to \mathbb{R}$ with L-Lipschitz continuous gradient. We suppose the optimisation problem

$$f^* = \min_{x\in\mathbb{R}^d} f(x) \tag{27}$$

has non-empty solution set Ω and that f satisfies the Polyak-Lojasiewicz inequality with parameter λ. Then both implicit and explicit gradient methods with step size $\tau \leq 1/L$ have global exponential convergence rate to the global minimum. Moreover, the corresponding equation $\dot{y} = -\nabla f(y)$ has exponential convergence rate to a global minimizer.

Note that the important point here is the assumption of L-smoothness of the function.

2.1 Polyak-Lojasiewicz inequality for Wasserstein gradient flows

To define a analog inequality in Wasserstein spaces we need the following chain rule, see [1, Prop 10.3.18].

Proposition 8. Let $\mathcal{F}$ be a proper lower semicontinuous 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 def 4) is non-empty. Then for any $\xi_t \in \partial\mathcal{F}(\mu_t)$

$$\frac{d}{dt}\mathcal{F}(\mu_t) = \int v_t \cdot \xi_t d\mu_t \,. \tag{28}$$

For regular enough functionals, for example if $\mathcal{F}$ is a λ-convex functionals and μ_t a Wasserstein gradient flow of $\mathcal{F}$, 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

$$\frac{d}{dt}\mathcal{F}(\mu_t) = -\left\|\nabla\frac{\delta\mathcal{F}}{\delta\mu}(\mu_t)\right\|^2_{L_2(\mu_t)} \,. \tag{29}$$

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

Definition 13. *Let $\mathcal{F}$ be a function with the same conditions than proposition 8. We suppose 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 with parameter $\lambda > 0$ along the curve μ_t if*

$$\left\|\nabla\frac{\delta\mathcal{F}}{\delta\mu}(\mu_t)\right\|^2_{L_2(\mu_t)} \geq \lambda\mathcal{F}(\mu_t) \tag{30}$$

If this inequality is verified, it leads thanks to equation (29) and Gronwall's lemma

$$\mathcal{F}(\mu_t) \leq \mathcal{F}(\mu_0)e^{-\lambda t} \,, \tag{31}$$

proving global exponential convergence.

2.2 Regularity and Polyak-Lojasiewicz inequality for the Energy functional

In this section, we will show how regularity of the solution of the Wasserstein gradient flow is linked to the Polyak-Lojasiewicz inequality in the setting of compact riemannian manifolds.

When the subdifferential of E_ν is non-empty, we can characterize it, using the same ideas than in [7, 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).

Lemma 1. Let μ be a probability measure such that $\partial E_\nu(\mu)$ is non-empty. Then the vector field $\nabla\frac{\delta E_\nu}{\delta\mu}(\mu) = (\nabla G) \star (\mu - \nu)$ verifies

$$\left\|\nabla\frac{\delta E_\nu}{\delta\mu}(\mu)\right\|_{L_2(\mu)} \leq |\partial E_\nu|(\mu) \,. \tag{32}$$

Then we need a regularity result about our functional, which is proved in the Appendix.

Lemma 2. 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 non-empty subdifferential, in the sense of definition 5. Moreover, $\nabla\frac{\delta E_\nu}{\delta\mu}(\mu) \in \partial E_\nu(\mu)$.

Now, combining these lemmas, the element of minimal norm of $\partial\mathcal{F}(\mu_t)$ is exactly the time dependent vector field $v_t = \nabla\varphi_t$, which drives mass transfer along time for the curve μ_t. In other words, the following property is true.

Proposition 9. Let μ_t be a solution of equation (2) 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, according to definition 6.

We are able to formulate a condition that implies a Polyak-Lojasiewicz inequality in a compact manifold.

Proposition 10 (Polyak-Lojasiewicz inequality)**.** Let (M, g) be a closed Riemannian manifold and μ, ν be two measures with density w.r.t. volume measure on M, such that $\log(\mu)$ is bounded below. Then, it holds

$$E_\nu(\mu) \leq \frac{1}{\underline{\mu}} \left\| \nabla \frac{\delta E_\nu}{\delta \mu}(\mu) \right\|^2_{L_2(\mu)} , \tag{33}$$

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

Proof. The proof is straightforward since the inequality is exactly the following

$$\int_M |\nabla \varphi_{\mu-\nu}(x)|^2 \, d\mathrm{vol}(x) \leq \frac{1}{\underline{\mu}} \int_M |\nabla \varphi_{\mu-\nu}(x)|^2 \, d\mu(x) , \tag{34}$$

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

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 remains an open question.

2.3 Exponential convergence for globally regular data

First we start with a stability result, where regularity of the density implies a global Polyak-Lojasiewicz inequallity.

Proposition 11 (Stability of Polyak-Lojasiewicz condition)**.** Let μ_0, ν be two C^1 densities on M such that $\log(\mu_0)$ and $\log(\nu)$ are bounded in L^∞. Then if the associated equation (2) admits a continuous density, it verifies

$$\min(\min \mu_0, \min \nu) \leq \mu_t(x) \leq \max(\max \mu_0, \max \nu) \tag{35}$$

Sketch of proof. We use the regularity and an optimality argument. Let us define $\underline{x}(t) \coloneqq \arg\min \mu_t(x)$ and $\overline{x}(t) \coloneqq \arg\max \mu_t(x)$. As the manifold is closed and μ_t is C^1 we get

$$\nabla \mu_t(\underline{x}(t)) = \nabla \mu_t(\overline{x}(t)) = 0 . \tag{36}$$

Moreover, using the regularity we have

$$\begin{aligned} \partial_t \mu_t &= -\nabla \cdot (\mu_t v_t) && (37)\\ &= -v_t \cdot \nabla \mu_t - \mu_t \nabla \cdot v_t && (38)\\ \partial_t \mu_t &= -v_t \cdot \nabla \mu_t - \mu_t(\mu_t - \nu) && (39)\\ & && (40) \end{aligned}$$

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

$$\partial_t \mu_t(\underline{x}(t)) = (\min \mu_t)\nu(\underline{x}(t)) - (\min \mu_t)^2 . \tag{41}$$

After a brief study of the phase diagram (see the proof of lemma 3) we get our result. A similar argument applies to $\overline{x}(t)$. □

We will later prove a stronger result independently: we show in the proof of lemma 3 that Hölder continuity is actually enough to guaranty the stability of the condition. Under this regularity assumption, it is direct to prove that both exponential convergences under the $\dot{H}^{-1}$-norm and Wasserstein L^2 distance hold.

Theorem 3 (Exponential convergence)**.** Let μ, ν be two measures with continuous densities with bounded logarithm on a closed Riemannian manifold. If the density of the curve μ_t generated by the gradient flow of E_ν is $C^1_{t,x}$ in both time and space, then exponential convergence of μ to ν holds in the two following ways

$$\begin{cases} E_\nu(\mu_t) \leq E_\nu(\mu_0)e^{-\lambda t} \\ W_2^2(\mu_t, \nu) \leq \frac{4}{\lambda} E_\nu(\mu_0)e^{-\lambda t}\,, \end{cases} \tag{42}$$

where $\lambda = \min(\underline{\mu_0}, \nu)$.

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

We just reduced this global convergence problem to a regularity problem. In the next section, we prove global existence of Hölder continuous solutions if the initial and target densities are Hölder continuous.

2.4 Well Posedness in Hölder spaces

In this section we consider solutions of the PDE

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \\ v_t := -\nabla G \star (\mu_t - \nu)\,. \end{cases} \tag{43}$$

The velocity vector field v_t verifies $\nabla \cdot v_t = \mu_t - \nu$. Let us consider a solution μ_t of equation (43) on an open interval $]0, T[$. According to definition 2, it is an absolutely continuous curve associated with the time dependent vector field $v_t := \nabla\varphi$, where $\Delta\varphi_t = \mu_t - \nu$. If μ_0 has a density, μ_t also ha a density at least for small times.

2.4.1 In the Euclidean space.

This subsection is devoted to the proof of the following theorem

Theorem 4. Let μ_0 and ν be Hölder continuous densities with a compact support. Then equation (2) admits a unique global solution μ_t that is Hölder continuous at all time.

We use techniques from [6] and adapt them to our case. The standard technique is to rewrite the problem in Lagrangian coordinates, following the particle flow and using ODE results. However, long time existence follows from the fact that one can write an explicit evolution equation for a quantity that involves the jacobian of the map, and this equation resembles to

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

$$\frac{d}{dt}\psi_t = v_t \circ \psi_t\,. \tag{44}$$

We 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 Hölder solutions to the system. This dependance will be made precise in lemma 4.

We can rewrite the particle flow equation (44), 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

$$\frac{d}{dt}J_t(\alpha) = J_t(\alpha)(\mu_t - \nu)\circ\psi_t(\alpha)\,, \tag{45}$$

so that

$$\frac{d}{dt}\left[t \mapsto J_t(\alpha)\mu_t(\psi_t(\alpha))\right] = 0\,, \tag{46}$$

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

$$\frac{d}{dt}\psi_t(\alpha) = -\int \nabla G(\psi_t(\alpha) - \psi_t(\alpha'))\mu_0(\alpha)d\alpha + \int \nabla G(\psi_t(\alpha) - y)\nu(y)dy = F(\psi_t(\alpha))\,. \tag{47}$$

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

Proposition 12. Let μ_0 and ν be density measures in $\mathcal{B}$ with compact support. Then equation (47) with initial condition $\psi_0 = Id$ admits a unique solution on a maximal time interval $[0, T[$. Either T is infinite either the Banach norm $\|\psi_t\|_{1,\gamma}$ blows up as $t \to T$.

Proof. As F is Lipschitz (see [6] and [24, Chap 4]) we can use the Picard theorem in Banach space to prove the result. □

Now let us prove theorem 4, that is $T = +\infty$ in the preceding proposition. To show that $T = \infty$ we will 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 [6], an explicit formula is found for f_t, which can not be done in our case. However we are able to control the evolution of f_t.

Lemma 3. The quantity $\|f_t\|_\infty$ is uniformly bounded on $[0, T[$.

Proof. We can write

$$\nabla\cdot(v_t\mu_t) = \mu_t\nabla\cdot v_t + v_t\cdot\nabla\mu_t = \mu_t^2 - \mu_t\nu + v_t\cdot\nabla\mu_t\,, \tag{48}$$

so that we get

$$\frac{d}{dt}f_t(\alpha) = \partial_t\mu_t(\psi_t(\alpha)) + \nabla\mu_t(\psi_t(\alpha))\cdot v_t(\psi_t(\alpha)) \tag{49}$$

$$= -\nabla\cdot(v_t\mu_t)(\psi_t(\alpha)) + \nabla\mu_t(\psi_t(\alpha))\cdot v_t(\psi_t(\alpha)) \tag{50}$$

$$\frac{d}{dt}f_t(\alpha) = f_t(\alpha)\nu(\psi_t(\alpha)) - f_t^2(\alpha)\,. \tag{51}$$

Even if $\nu\circ\psi_t$ is of course not constant, this equation, that looks like 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 for all $\alpha \in \mathbb{R}^d$

$$\min(\min f_0, \min\nu) \le f_t(\alpha) \le \max(\|f_0\|_\infty, \|\nu\|_\infty)\,. \tag{52}$$

□

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

Lemma 4. Let μ_t be a solution defined as in proposition 12, where $\nu \in L_\infty$ verifies $|\nu|_\gamma < \infty$. Then

$$|\mu_t|_\gamma \le C\|d\psi_t\|_\infty^\gamma\left(\int_0^t (1 + \|d\psi_s^{-1}\|_\infty^\gamma)ds\right) \tag{53}$$

for some positive constants $C > 0$.

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

$$|\mu_t|_\gamma \leq |f_t|_\gamma \|d\psi_t^{-1}\|_\infty^\gamma . \tag{54}$$

Now, to control $|f_t|_\gamma$, we can take the Holder semi norm of equation (51), using the fact that v 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

$$\begin{aligned}
\frac{d}{dt}|f_t|_\gamma &\leq |f_t \cdot v \circ \psi_t| + |f_t^2|_\gamma && (55)\\
&\leq |f_t|_\gamma \|v \circ \psi_t\|_\infty + \|f_t\|_\infty |v \circ \psi_t|_\gamma + 2\|f_t\|_\infty |f_t|_\gamma && (56)\\
&\leq (\|v\|_\infty + 2\|f_t\|_\infty)|f_t|_\gamma + \|f_t\|_\infty |v|_\gamma \|d\psi_t\|_\infty^\gamma && (57)\\
\frac{d}{dt}|f_t|_\gamma &\leq C_1 |f_t|_\gamma + C_2 \|d\psi_t\|_\infty^\gamma , && (58)
\end{aligned}$$

for some positive constants C_1, C_2. Next we are able to 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 $y(t) \leq y(0) e^{\int_0^t a(s)ds} + \int_0^t b(s) e^{\int_0^t a(u)du - \int_0^s a(u)du} ds$. This provides, as $t < T < \infty$

$$|f_t|_\gamma \leq |f_0|_\gamma \exp(C_1 t) + \int_0^t C_2 \|d\psi_s\|_\infty^\gamma \exp(C_1(t-s))ds \leq C\left(1 + \int_0^t \|d\psi_s\|_\infty^\gamma ds\right) , \tag{59}$$

ending the proof. □

Differentiating the particle equation (44) and taking the L_∞ norm we get

$$\frac{d}{dt}\|d\psi_t\|_\infty \leq \|dv_t\|_\infty \|d\psi_t\|_\infty , \tag{60}$$

which gives by Gronwall's lemma

$$\|d\psi_t\|_\infty \leq C \exp \int_0^t \|dv_s\|_\infty ds . \tag{61}$$

Let us remark that ψ_t^{-1} admits a similar bound.

Lemma 5. Let ψ_t^{-1} be defined as the inverse flow of φ_t. Then the majoration above is true for ψ_t^{-1}, i.e. for some constant $C > 0$

$$\|d\psi_t^{-1}\|_\infty \leq C \exp \int_0^t \|dv_s\|_\infty ds . \tag{62}$$

Proof. Let $\alpha := \psi_s(y)$ for some $y \in \mathbb{R}^d$ and $s > 0$. Then if $0 \leq t < s$, we get that $\alpha \in \psi_t(\mathbb{R}^d)$, 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

$$\partial_t \psi_t^{-1}(\alpha) = \partial_t \psi_{s-t} \circ \psi_s^{-1}(\alpha) = -v_{s-t} \circ \psi_{s-t} \circ \psi_s^{-1}(\alpha) = -v_{s-t} \circ \psi_t^{-1}(\alpha) . \tag{63}$$

We get the inequality

$$\frac{d}{dt}\|d\psi_t^{-1}\|_\infty \leq \|dv_{s-t}\|_\infty \|d\psi_t^{-1}\|_\infty , \tag{64}$$

so that by Grönwall lemma

$$\|d\psi_t^{-1}\|_\infty \leq C \exp\left(\int_0^t \|dv_{s-u}\|_\infty du\right) . \tag{65}$$

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

$$\|d\psi_t^{-1}\|_\infty \leq C \exp\left(\int_0^t \|dv_u\|_\infty du\right). \tag{66}$$

□

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 found in [24, Sections 2.4.2,3]. The kernel $G_2 := d_x \nabla G$ is homogeneous of degree -N. Because of this, the singularity at the diagonal cannot be integrated. However it has mean-value zero and defines a singular integral operator through the convolution

$$G_2 \star f(x) = PV \int G_2(x,y) f(y) dy := \lim_{\varepsilon \to 0} \int_{d(x,y)>\varepsilon} G_2(x,y) f(y) dy \,. \tag{67}$$

Using the same arguments than in [24, Prop 2.20] we prove that we actually have for a velocity field defined by $\nabla G \star f$ with G the Coulomb kernel and $f \in C^\gamma(\mathbb{R}^d;\mathbb{R}^d)$

$$dv_t(x) = PV \int G_2(x,y) f(y) dy \,. \tag{68}$$

The following lemma, found in [24, Lemma 4.5 and 4.6] and [6, Lemma 2.2], is written in terms of principal value integral in [6] but used in the following form.

Lemma 6. Let $f \in C^\gamma(\mathbb{R}^d;\mathbb{R}^d)$ 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

$$\|dv\|_\infty \leq C\left[|f|_\gamma \varepsilon^\gamma + \max\left(1; \log\left(\frac{R}{\varepsilon}\right)\right)\|f\|_\infty\right]; \forall \varepsilon > 0\,, \tag{69}$$

$$|dv|_\gamma \leq C|f|_\gamma \,. \tag{70}$$

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

Lemma 7. Let μ_t be a solution defined as in proposition 12 and v_t the associated velocity field. Then for some constant $C > 0$ only depending on the dimension

$$\|v_t\|_\infty \leq C(\|\mu_t - \nu\|_\infty + \|\mu_t - \nu\|_1)\,. \tag{71}$$

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

$$|v_t(x)| \leq \left[\int_{B(0,1)} + \int_{B(0,1)^c}\right] |\nabla G(x-y)(\mu_t - \nu)(y)| dy \tag{72}$$

$$\leq \|\mu_t - \nu\|_\infty \int_{B(0,1)} |\nabla G(x-y)| dy + \int_{B(0,1)^c} |\mu_t - \nu|(y) dy \tag{73}$$

$$|v_t(x)| = C(\|\mu_t - \nu\|_\infty + \|\mu_t - \nu\|_1)\,. \tag{74}$$

□

Lemma 8. Let μ_t be a solution defined as in proposition 12. Let us 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 7. As $\|\mu_t\|_\infty = \|f_t\|_\infty \leq C$ by lemma 3, the first term is uniformly bounded in $[0, T[$. As μ_t and ν are probability densities, so is the second term. This shows the result. □

As we proved that μ_t has bounded support, we can use the first estimate from lemma 6 with $\varepsilon = \left[\|\mu_t - \nu\|_\infty / |\mu_t - \nu|_\gamma\right]^{1/\gamma}$ to get the existence of C independent of t such as

$$\|dv_t\|_\infty \leq C\left[\|\mu_t - \nu\|_\infty + \max\left(1, \log\left(\frac{R(t)|\mu_t - \nu|_\gamma^{1/\gamma}}{\|\mu_t - \nu\|_\infty^{1/\gamma}}\right)\|\mu_t - \nu\|_\infty\right)\right]. \tag{75}$$

We know thanks to lemma 3 that $\|\mu_t - \nu\|_\infty$ is bounded, and by lemma 8, as $T < \infty$ we get the existence of $C_1, C_2 > 0$ such that

$$\|dv_t\|_\infty \leq C_1 + C_2 \log(|\mu_t - \nu|_\gamma). \tag{76}$$

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

Proposition 13. The quantity $\|dv_t\|_\infty$ is uniformly bounded on the time interval $[0, T[$.

Proof. Injecting the inequality from lemma 4 into (76), we get the existence of constants such that

$$\|dv_t\|_\infty \leq A + B\log(\|d\psi_t\|_\infty) + C\log\left(\int_0^t (1 + \|d\psi_s^{-1}\|_\infty^\gamma) ds\right). \tag{77}$$

We write, using inequality (62), the fact that $s \leq t$ in the integrals and that $\|dv_t\|_\infty$ is a positive function

$$\log\left(\int_0^t \|d\psi_s^{-1}\|_\infty^\gamma ds\right) \leq C\log\left(\int_0^t \gamma \exp\left(\int_0^s \|dv_u\|_\infty du\right) ds\right) \tag{78}$$

$$\leq C\log\left(\int_0^t \gamma \exp\left(\int_0^t \|dv_u\|_\infty du\right) ds\right) \tag{79}$$

$$\leq C\log\left(\gamma t \exp\left(\int_0^t \|dv_u\|_\infty du\right)\right) \tag{80}$$

$$\log\left(\int_0^t \|d\psi_s^{-1}\|_\infty^\gamma ds\right) \leq C\left(\log t + \int_0^t \|dv_u\|_\infty du\right), \tag{81}$$

and we get the final differential inequality

$$\|dv_t\|_\infty \leq C\left(1 + \int_0^t \|dv_s\|_\infty ds + \log(t)\right). \tag{82}$$

Once again, Gronwall's lemma applies to $t \mapsto \|dv_t\|_\infty$ and we obtain the existence of some 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 14. 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 (61), (62) and lemma 4 show the result. □

Finally, we can control our last term

Proposition 15. The quantity $|d\psi_t|_\gamma$ is uniformly bounded in $[0, T[$.

Proof. Differentiating the particle equation (44), taking the Hölder $|\cdot|_\gamma$ semi-norm and using the preceding proposition with the second potential theory estimate from lemma 6 we get

$$\frac{d}{dt}|d\psi_t|_\gamma \leq |d(v_t \circ \psi_t)|_\gamma \|d\psi_t\|_\infty + \|d(v_t \circ \psi_t)\|_\infty |d\psi_t|_\gamma \tag{83}$$

$$\leq |dv_t|_\gamma \|d\psi_t\|_\infty^{1+\gamma} + \|dv_t\|_\infty \|d\psi_t\|_\infty |d\psi_t|_\gamma \tag{84}$$

$$\leq C_1 |\mu_t|_\gamma \exp\left((1+\gamma)\int_0^t \|dv_s\|_\infty ds\right) + C_2 |d\psi_t|_\gamma \tag{85}$$

$$\frac{d}{dt}|d\psi_t|_\gamma \leq C_1 + C_2 |d\psi_t|_\gamma \,. \tag{86}$$

One last application of Gronwall's lemma end the proof.

□

We just proved that if $T < \infty$ and that a solution of problem (47) 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 verifies $T < \infty$, we get a contradiction with the existence result 12, which states that if $T < \infty$ a finite time blowup of $\|\psi_t\|_{1,\gamma}$ occurs. This proves theorem 4.

2.4.2 In a compact Riemannian manifold. This Lagrangian formulation allows to directly extend our result on a complete closed Riemannian manifold (M, g). Indeed, if we take G to be the Coulomb kernel on M, then the Lagrangian formulation (47) still holds. As Hölder regularity is a local property, the functionnal F is still locally Lipschitz on the Banach space

$$\mathcal{B}_M := \left\{\psi : M \to TM \,|\, \|\psi\|_{1,\gamma} < \infty\right\} \,, \tag{87}$$

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

$$\|f\|_{C_M^{0,\gamma}} := \sup_{x,y\in M} \frac{|f(x) - f(y)|}{d_M(x,y)^\gamma} \,. \tag{88}$$

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 6 on closed manifolds.

Lemma 9. Let $u \in C^\gamma(M)$ on a closed manifold M. We consider the equation

$$\Delta_M \varphi = u \,. \tag{89}$$

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

$$\|\varphi\|_{2,\infty} \leq C\left(\|u\|_{C_M^{0,\gamma}} \varepsilon^\gamma + \log\left(\frac{A}{\varepsilon}\right)\|u\|_\infty\right) \,, \tag{90}$$

$$\|\varphi\|_{C_M^{2,\gamma}} \leq C\left(\|u\|_{C_M^{0,\gamma}} + \|\varphi\|_\infty\right) \,. \tag{91}$$

Proof of the first inequality. The proof of the first inequality is almost the same as in the Euclidean space. We denote G_2 the differential (in coordinates) of the gradient of G, i.e. $G_2 := d_x \nabla G$. Its singularity on the diagonal behaves like the Coulomb kernel, meaning we have (see [3, Theorem 4.13.c])

$$G_2(x,y) = O(1/d_M(x,y)^d) \,. \tag{92}$$

Moreover, in Riemannian manifolds

$$\int_{d_M(x,y)>\varepsilon} d_M(x,y)^{-d} dy = O(-\log \varepsilon) \,, \tag{93}$$

and if $\alpha < d$

$$\int_{d_M(x,y)<\varepsilon} d_M(x,y)^{-\alpha} dy = O(\varepsilon^{d-\alpha}) \,. \tag{94}$$

We write

$$\|\varphi\|_{2,\infty} = \left(PV \int_{d(x,y)\leq\varepsilon} + \int_{d(x,y)>\varepsilon}\right) G_2(x,y)f(y)dy = I_1(x) + I_2(x)\,. \tag{95}$$

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

$$|I_1(x)| = \left|\int_{d(x,y)\leq\varepsilon} G_2(x,y)(f(y)-f(x))dy\right| \tag{96}$$

$$\leq \int_{d(x,y)\leq\varepsilon} \|G_2(x,y)\|\|f\|_{C_M^{0,\gamma}} d(x,y)^\gamma dy \tag{97}$$

$$\leq C|f\|_{C_M^{0,\gamma}} \int_{d(x,y)\leq\varepsilon} d_M(x,y)^{-d+\gamma} dy \tag{98}$$

$$|I_1(x)| \leq C|f\|_{C_M^{0,\gamma}} \varepsilon^\gamma\,. \tag{99}$$

For the second term

$$|I_2(x)| = \int_{d(x,y)>\varepsilon} G_2(x,y)f(y)dy \tag{100}$$

$$\leq C\|f\|_\infty \int_{d(x,y)>\varepsilon} d(x,y)^{-d} dy \tag{101}$$

$$|I_2(x)| \leq C\|f\|_\infty \log\left(\frac{A}{\varepsilon}\right)\,. \tag{102}$$

This proves the first estimate. □

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 $\mathbb{R}^d$ (see [3, 3.61]). We use local normal coordinates for the Laplace-Beltrami operator. The bound can be used thanks to the following metric control of the Hölder norm. We refer to the lemma below. □

Lemma 10. There exists $r > 0$ and $C > 0$ such that, if $\Omega \subset M$ has diameter less than r, then if $u \in C^{2,\gamma}$

$$\frac{1}{C}\|u\|_{C_{Eucl}^{k,\gamma,\Omega}} \leq \|u\|_{C_M^{k,\gamma}} \leq C\|u\|_{C_{Eucl}^{k,\gamma,\Omega}}\,. \tag{103}$$

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

We now 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, as 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 9 with, once again, $\varepsilon = \left[\|\mu_t - \nu\|_\infty/|\mu_t - \nu|_\gamma\right]^{1/\gamma}$ to get formula (76) 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, equation (83) becomes

$$\frac{d}{dt}|d\psi_t|_\gamma \leq |d(v_t\circ\psi_t)|_\gamma \|d\psi_t\|_\infty + \|d(v_t\circ\psi_t)\|_\infty |d\psi_t|_\gamma \tag{104}$$

$$\leq |dv_t|_\gamma \|d\psi_t\|_\infty^{1+\gamma} + \|dv_t\|_\infty \|d\psi_t\|_\infty |d\psi_t|_\gamma \tag{105}$$

$$\leq C_1\left(|\mu_t|_\gamma + \|v_t\|_\infty\right)\exp\left((1+\gamma)\int_0^t \|dv_s\|_\infty ds\right) + C_2|d\psi_t|_\gamma \tag{106}$$

$$\frac{d}{dt}|d\psi_t|_\gamma \leq C_1 + C_2|d\psi_t|_\gamma \,. \tag{107}$$

The rest of the proof is the same as in the Euclidean space. This proves the following theorem.

Theorem 5 (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). We consider the curve μ_t global solution of equation (43). We just proved it is defined and Hölder continuous at all time. Then

$$\begin{cases} E_\nu(\mu_t) \leq E_\nu(\mu_0)e^{-\lambda t} \\ W_2^2(\mu_t, \nu) \leq \frac{4}{\lambda} E_\nu(\mu_0)e^{-\lambda t} \,. \end{cases} \tag{108}$$

Proof. As $\log(\mu_t)$ is globally bounded from below thanks to lemma 3, it verifies a global Polyak-Lojasiewisz at all time. The rest of the proof is a simple application of this inequality. □

3 Critical points for the Wasserstein flow.

In the previous section, we studied some solutions of Wasserstein gradient flows, with regular initial data. Our proofs heavily relied on regularity results that we were able to get thanks to regular initial and target data. In this section we consider a given probability measure ν which is, depending on the context, of finite energy for the Coulomb kernel or the Energy Distance kernel. We do not make any other assumptions on ν and μ_0. 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 get stuck, in direct analogy with gradient flows in finite dimensions.

Definition 14 (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 $\tau \leq \tau_0$ we have*

$$\mu \in \operatorname*{arg\,min}_{\rho \in \mathcal{P}(\mathbb{R}^d)} \mathcal{F}(\rho) + \frac{1}{2\tau} W_2^2(\mu, \rho) \,. \tag{109}$$

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

Definition 15 (Lagrangian Wasserstein critical point)**.** *A probability measure μ is said to be a Lagrangian critical point for a functional $\mathcal{F}$ if, on $supp(\mu)$ we have*

$$\nabla \frac{\delta \mathcal{F}}{\delta \mu}(\mu) = 0 \,. \tag{110}$$

For λ-convex functionnals, see [29], 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 2 we see that

Proposition 16. 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 1, we see that if μ is a critical point of the G-energy from ν E_ν, then the gradients of the potentials $G \star \mu$ and $G \star \nu$ are equal μ almost everywhere. This observation is the principal argument allowing 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 the following theorem.

Theorem 6. Let μ be a Lagrangian critical point for the MMD functional E_ν^G. Then

$$\mu_{|\mathrm{Int}(\mathrm{supp}(\mu))} = \nu_{|\mathrm{Int}(\mathrm{supp}(\mu))}\,, \tag{111}$$

in the two cases

1. if G is the Coulomb kernel,
2. if G is the Energy Distance kernel and d is odd.

This formulation may seem surprising. We only get results in the interior of the support of our measure, however 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 will focus on the first statement for now.

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

$$\nabla \frac{\delta E_\nu}{\delta \mu}(\mu) = \nabla \varphi_\mu - \nabla \varphi_\nu\,. \tag{112}$$

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

Proposition 17. Let μ be a probability measure on $\mathbb{R}^d$. We consider the associated potential $\varphi_\mu(x) \coloneqq \int -\|x-y\| d\mu(y)$. Then, its distributional Laplacian exists almost everywhere and is given by

$$\Delta \varphi_\mu(x) \coloneqq \int -\frac{d-1}{\|x-y\|} d\mu(y)\,. \tag{113}$$

Proof. Let $g \in C_c^\infty$, and let us write $P_\mu(x) \coloneqq \int -\frac{d-1}{\|x-y\|} d\mu(y)$. We wish to prove $\langle \varphi_\mu, \Delta g \rangle = \langle P_\mu, g \rangle$. The main issue is that the function $\|\cdot - y\|$ admits a singularity in y. We circumvent this by integrating on a small ball of radius $\varepsilon > 0$ near y, and by controlling the error term. The inversion of integrals in the second equality comes from the Fubini theorem

$$\begin{aligned}
\langle \varphi_\mu, \nabla g \rangle &\coloneqq \int \varphi_\mu(x) \Delta g(x) dx && (114)\\
&= \int \left(\int -\|x-y\| \Delta g(x) dx \right) d\mu(y) && (115)\\
&= \int \left[\left(\int_{B(y,\varepsilon)} -\|x-y\| \Delta g(x) dx \right) + \left(\int_{\mathbb{R}^d \setminus B(y,\varepsilon)} -\|x-y\| \Delta g(x) dx \right) \right] d\mu(y) && (116)\\
\langle \varphi_\mu, \nabla g \rangle &= \int \left[I_\varepsilon(y) + J_\varepsilon(y) \right] d\mu(y)\,, && (117)
\end{aligned}$$

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

$$|I_\varepsilon(y)| \leq 2\|\Delta g\|_\infty \varepsilon\,. \tag{118}$$

For $J_\varepsilon(y)$, we are able to use Green's Formula, as 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 $\Omega(y)$, we get $\eta(x) = -\frac{x-y}{\|x-y\|}$. 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 $B(y,\varepsilon)$. By Green's Formula

$$J_\varepsilon(y) = \int_{\Omega(y)} \Delta - d_y(x)g(x)dx + \int_{\partial\Omega(y)} \frac{\partial d_y}{\partial \eta}(x)g(x)dS(x) - \int_{\partial\Omega(y)} d_y(x)\frac{\partial g}{\partial \eta}(x)dS(x). \tag{119}$$

First, we know : $\Delta d_y(x) = \frac{d-1}{\|x-y\|}$. Second, as $\nabla d_y(x) = \frac{x-y}{\|x-y\|}$, we get : $\frac{\partial d_y}{\partial \eta}(x) = -1$. This gives

$$\left| \int_{\partial\Omega(y)} \frac{\partial d_y}{\partial \eta}(x)g(x)dS(x) \right| \leq \|g\|_\infty \mathcal{A}(S(y,\varepsilon))\,. \tag{120}$$

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

$$\left| \int_{\partial\Omega(y)} d_y(x)\frac{\partial g}{\partial \eta}(x)dS(x) \right| \leq C\mathcal{A}(S(y,\varepsilon))\,. \tag{121}$$

All these quantities vanish when $\varepsilon \to 0^+$, independently from y, this gives $\langle \varphi_\mu, \Delta g\rangle = \langle P_\mu, g\rangle$, that is $\Delta\varphi_\mu(x) = \int -\frac{d-1}{\|x-y\|} d\mu(y)$. □

Now, we will prove the following property, which will allow us to prove the second statement in theorem 6.

Lemma 11. Let μ be a lagrangian critical point for the Energy Distance Wasserstein gradient flow towards ν. Then, on the open set $\mathrm{Int}(\mathrm{supp}(\mu))$ $\varphi_\mu^G = \varphi_\nu^G$ where G is the Coulomb kernel.

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

$$\forall k < (d-1)/2, P_{2k+1}^\mu = P_{2k+1}^\nu\,. \tag{122}$$

We already saw 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 now that if $v : \mathbb{R}^+ \to \mathbb{R}$ and $u(x) := v(\|x\|)$, then its distributional Laplacian is given by

$$\Delta u(x) = \frac{d-1}{\|x\|} v'(\|x\|) + v''(\|x\|)\,. \tag{123}$$

We get, if $v(r) := \frac{1}{r^{2k+1}}$, that $\Delta u(x) = \frac{(2k+1)(2k+3-d)}{\|x\|^{-(2k+3)}}$. If we denote $K_y(x) := \frac{1}{\|x-y\|^{2k+1}}$, 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 $\varepsilon > 0$

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

$$\langle P^{\mu}_{2k+1}, g\rangle = \int [I_\varepsilon(y) + J_\varepsilon(y)]\, d\mu(y)\,.$$

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

$$\int_{B(y,\varepsilon)} K_y(x)\Delta g(x)dx \leq \|\Delta g(x)\|_\infty \int_{B(0,\varepsilon)} K_0(x)dx\,, \tag{124}$$

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

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

which gives

$$\left|\int_{\partial\Omega(y)} \frac{\partial K_y}{\partial \eta}(x)g(x)dS(x)\right| \leq \|g\|_\infty \int \frac{1}{\|x-y\|^{2k+2}} dS(x) = \|g\|_\infty C_d \varepsilon^{d-1-(2k+2)}\,, \tag{125}$$

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

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

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

$$\Delta P^{\mu}_{2k+1} = (2k+1)(2k+3-d)P^{\mu}_{2k+3} \tag{126}$$

Now, back to our hypothesis $P^{\mu}_{2k+1} = P^{\nu}_{2k+1}$. By taking the distributional Laplacian from both sides, we get : $(2k+1)(2k+3-d)P^{\mu}_{2k+3} = (2k+1)(2k+3-d)P^{\nu}_{2k+3}$, and as $2k < d-3$, we get $P^{\mu}_{2k+3} = P^{\nu}_{2k+3}$. 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 $\mathrm{supp}(\mu)$ the equality $\varphi^G_\mu = \varphi^G_\nu$. □

The potentials φ^G_μ and φ^G_ν are superharmonic on the interior of $\mathrm{supp}(\mu)$, and taking their distributional Laplacian yields $\mu_{|\mathrm{Int}(\mathrm{supp}(\mu))} = \nu_{|\mathrm{Int}(\mathrm{supp}(\mu))}$, which is exactly the second statement of Theorem 6.

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 18 that singular enough measures cannot be critical points.

4 No local minima in the Wasserstein geometry

In this section, we denote by G the Coulomb kernel in $\mathbb{R}^d$ and we only study this kernel. The last section was dedicated to Lagrangian critical points, who describe one kind of evolution for our curve: cases where the current measure is push forwarded through a map. For functionals defined only on absolutely continuous functions (entropy, p-energy for Porous Medium Equations...) or for

functionals convex along Wasserstein geodesics (potentials $\mathcal{V} : \mu \mapsto \int V(x)d\mu(x)$ where V is convex, internal non-local energies $\mathcal{W} : \mu \mapsto \iint W(x,y)d\mu(x)d\mu(y)$ where $W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is convex [29]), both notions are the same.

However, push forwarded measures do not describe all possible dynamics in our case. For non-convex functionals, diffusion phenomena can happen. Using Wasserstein geometry terms, it means, heuristically, that a descent direction has to be found in the space of velocity plans (see section 12.4 in [1]), instead of L_2 velocity maps. To do this, we use concepts inspired from flow interchange techniques developed in [22,25]. Instead of studying the gradient flow of E, we study how E behaves along a certain auxiliary flow. Here, this flow is associated the Boltzmann entropy functional.

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

Theorem 7 (No local minima and only global)**.** 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 start with the remark that the MMD energy only depends on the difference of 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 12 (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 singulars measures. In addition, $\mu_+ \ll \mu$ and $\nu_- \ll \nu$.

A heat diffusion process can be applied to μ_+, which allows 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 equivalent way $\mu_+ = \nu_- = 0$).

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 an 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

$$K_t(x) := \frac{1}{(4\pi t)^{d/2}} \exp(-\|x\|^2/4t)\,. \tag{127}$$

Remark 2. Here, ρ is not necessarily a probability measure. However, we can write the whole Wasserstein formalism for any space of measures $\mathcal{M}_m := \{\mu \in \mathcal{M}_+(\mathbb{R}^d), \mu(\mathbb{R}^d) = m\}$. We will abuse notations, and denote $W_2(\alpha, \beta)$ the Wasserstein distance on $\mathcal{M}_m$ if $\alpha, \beta \in \mathcal{M}_m$.

Lemma 13. 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 $s, t > 0$

$$W_2(\mu_t, \mu_s) \leq \sqrt{|t-s|}C\,. \tag{128}$$

Proof. Standard results exposed in [1, 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 to the heat flow, that states that for all α density measure with the same mass as ρ and for almost every $t > 0$

$$\frac{1}{2}\frac{d}{dt} W_2^2(\rho_t, \alpha) \leq \mathcal{H}(\alpha) - \mathcal{H}(\rho_t)\,. \tag{129}$$

This implies that there exists a constant $C > 0$ such as,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 $\tilde{\gamma} \coloneqq \gamma + (\mu - \rho)^{\otimes 2}$ is a transport plan between μ_t and μ_s. Thus

$$W_2(\mu_t, \mu_s) \leq W_2(K_t \star \rho, K_s \star \rho)\,, \tag{130}$$

which proves the lemma. □

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

Lemma 14. The function $t \mapsto E_\nu(\mu_t)$ is differentiable for every $t > 0$, and its derivative is given by

$$\frac{1}{c_d}\frac{d}{dt}E_\nu(\mu_t) = -\left\langle \rho, K_{2t} \star \rho + K_t \star (\mu - \rho) - K_t \star \nu \right\rangle\,. \tag{131}$$

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 (131) is strictly negative. An essential property is the following lemma (see [34]).

Lemma 15 (Heat kernel estimates, [34])**.** Let α and β be two positive mutually singular measures. Then, as $t \to 0$

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

for μ almost every x.

We use this lemma to derive the following estimate.

Lemma 16. Let α and β be two positive mutually singular measures. Then, as $t \to 0$

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

for μ almost every 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. □

4.3 Proof of theorem 7

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

Proposition 18. 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

$$X_N \coloneqq \left\{x \in \mathrm{supp}(\mu_+) | t_x \geq 1/N\right\}\,.$$

This is a growing sequence for inclusion, and it verifies $\mathrm{supp}(\mu_+) = \bigcup_{N\in\mathbb{N}^*} X_N$. 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 $t < t_0$

$$\frac{d}{dt}E_v(\mu_t) = -c_d\left\langle \mu_{+|A}, K_{2t}\star\mu_{+|A} - K_t\star\nu_-\right\rangle \tag{132}$$

$$\leq -\frac{1}{2}c_d\left\langle \mu_{+|A}, K_{2t}\star\mu_{+|A}\right\rangle\,. \tag{133}$$

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

4.4 Dimension of measure and critical points

In theorem 7, 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 $t \mapsto E_v(\mu_t)$ was strictly decreasing near 0. However, it is not enough to guarantee that the JKO steps (13) do not stay stationnary at μ. Our theorem 6 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 it can not be a critical point as in definition 3.1. In analogy to the finite dimension 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 then we prove that the heat diffusion makes the energy E_v decrease fast enough to compensate the growth of $W_2^2(\mu, K_t\star\mu)$, so that μ is not a point where JKO steps stay stationnary. To do this, we can use a more precise version of 15, see [34]:

Lemma 17. 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 $x \in \mathbb{R}^d$

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

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

$$\lim_{t\to 0} K_t\star\mu(x) = f(x)/c_{d,q}\,. \tag{134}$$

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 $\mu(A) > 0$ $\lim_{r\to 0} r^{-(d-\delta)}\mu(B(x,r)) = f(x)$, then, for all $x \in A$ one has $\lim_{t\to 0} t^{\delta/2}K_t\star\mu(x) = f(x)/c_{d,q}$. That is, when $t \to 0$

$$K_t\star\mu(x) \sim \frac{f(x)}{c_{d,q}}\frac{1}{t^{\delta/2}}\,. \tag{135}$$

Let μ and ν be two probability measures with finite Coulomb energy. We write, once again, the Hahn-Jordan decomposition $\mu - \nu = \mu_+ - \nu_-$. From the result of Section 3, we know that at a critical

point which is not the global minimum, μ_+ cannot have a density w.r.t. the Lebesgue measure. We study what happens if μ_- is singular, in a precise sense. The point of interest, to be able to use Lemma 17, will be the growth of $\mu_+(B(x,r))$ on the support of μ_+ as r tends to 0. We prove the following lemma:

Lemma 18. If there exists a set $A \subset \mathbb{R}^d$ such that $\mu_+(B) > 0$, 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 $x \in A$

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

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

Proof. In this proof, we denote by C any strictly positive constant which 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 17, we get the existence $t_0 > 0$ such that, for all $0 < t < t_0$ and μ_+ almost every $x \in A$

$$K_{2t} \star \mu_+(x) \geq \frac{C}{t^{\delta/2}} . \tag{136}$$

Moreover, by the same reasoning as in the proof of Theorem 18, we can suppose, if we take a subset of A, that there exists $t_1 > 0$ such as, for any $0 < t < t_1$

$$\frac{d}{dt} E_v(\mu_t) \leq -C \left\langle \mu_{+|A}, K_{2t} \star \mu_{+|A} \right\rangle . \tag{137}$$

Now, using Formula (136), this gives along μ_t, $\frac{d}{dt} E_v(\mu_t) \leq -C t^{-\delta/2}$. 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 $E_v^{\tau,\mu} := E_v + \frac{1}{2\tau} W_2^2(\mu_0, \cdot)$

$$\frac{d}{dt} E_v^{\tau,\mu}(\mu_t) \leq C \left(-t^{-\delta/2} + \frac{1}{2\tau} \right) , \tag{138}$$

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

$$E_v^{\tau,\mu}(\mu_t) - E_v^{\tau,\mu}(\mu_0) \leq C \left(-\frac{2t^{-\delta/2+1}}{2-\delta} + \frac{t}{2\tau} \right) . \tag{139}$$

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 (139) we get

$$E_v^{\tau,\mu}(\mu_{t_\tau}) - E_v^{\tau,\mu}(\mu_0) \leq -C \tau^{\frac{2-\delta}{\delta}} , \tag{140}$$

concluding the proof since

$$\min_{\rho \in \mathcal{P}_2(\mathbb{R}^d)} E_v^{\tau,\mu}(\rho) - E_v^{\tau,\mu}(\mu_0) \leq \min_{t>0} E_v^{\tau,\mu}(\mu_t) - E_v^{\tau,\mu}(\mu_0) \leq -C \tau^{\frac{2-\delta}{\delta}} . \tag{141}$$

□

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

Theorem 8. Let us suppose there exists a set A such that $\mu_+(A) > 0$ and such that $\mu_{+|A}$ has Minkowski dimension less than $d - \delta$ (see [18]) 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

$$\dim^x_{\mathcal{M}}(\mu) := \lim_{r\to 0} \frac{\log(\mu(B(x,r)))}{\log(r)}\,. \tag{142}$$

If μ_+ charges some set A such as, for all $x \in A$, $\dim^x_{\mathcal{M}}(\mu)$ exists and is inferior to $d-\delta$, we get that 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 than in the proof of proposition 18, 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 18, ending the proof. □

Remark 3. 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) can be explained by the fact that for $\delta > 1$, the energy functional is infinite. In such a case, it is direct to prove from the very definition that such a measure (of infinite energy) cannot be a critical point.

4.5 Extension to Riemannian manifolds

Recall that the heat kernel is defined on a general Riemannian manifold.

Proposition 19 (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 $K : (0,\infty)\times M\times M \to \mathbb{R}$ as the smallest positive fundamental solution of the heat equation, meaning that for any $y \in M$

$$\begin{cases} \partial_t K = \Delta_x K \\ K(t,\cdot,y) \underset{t\to 0}{\to} \delta_y\,. \end{cases} \tag{143}$$

In order to generalize the conclusions of section 4, we need estimates resembling to the one in the proof of Lemma 16, mainly that there is some constant $C > 0$ such that

$$K_t \leq C K_{2t}\,. \tag{144}$$

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 ([33]), if (M,g) is a geodesically complete non-compact Riemannian manifold of nonnegative Ricci curvature ([15]), or if (M,g) is compact ([23]). 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 periodisation of the euclidean heat kernel, that is

$$K_{\mathbb{T}}(t,x,y) := \sum_{n\in\mathbb{Z}} K_{\mathbb{R}}(t,x+k,y)\,. \tag{145}$$

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

$$K_{\mathbb{T}^d}(t,x,y) = \prod_{1\leq k\leq d} K_{\mathbb{T}}(t,x_k,y_k)\,, \tag{146}$$

so that the desired estimate in Formula (144) holds for some $C > 0$. An example of manifold where it does not hold is the hyperbolic space.

Acknowledgements

The authors would like to thank Stephen C. Preston and Adrien Vacher for their help in this work. This work was supported by the Bézout Labex (New Monge Problems), funded by ANR, reference ANR-10-LABX-58.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2005. 5, 6, 7, 11, 24

[2] M. Arbel, A. Korba, A. Salim, and A. Gretton, *Maximum mean discrepancy gradient flow*, Advances in Neural Information Processing Systems, 32 (2019). 8

[3] T. Aubin, *Some Nonlinear Problems in Riemannian Geometry*, Springer, 1998. 9, 10, 18, 19

[4] D. Balagué, J. A. Carrillo, and Y. Yao, *Confinement for repulsive-attractive kernels*, Discrete and Continuous Dynamical Systems - B, 19 (2014), pp. 1227–1248. 4

[5] A. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Transactions on Information Theory, 39 (1993), pp. 930–945. 3

[6] A. L. Bertozzi, T. Laurent, and F. Leger, *Aggregation and spreading via the newtonian potential: The dynamics of patch solutions*, Mathematical Models and Methods in Applied Sciences, 22 (2012), p. 1140005. http://www.math.ucla.edu/ bertozzi/papers/Patch-Galley.pdf. 13, 14, 16

[7] Bonaschi, *Gradient flows driven by a non-smooth repulsive interaction potential*, (2013). 11

[8] J. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepčev, *Confinement in nonlocal interaction equations*, Nonlinear Analysis: Theory, Methods & Applications, 75 (2012), pp. 550–558. 4

[9] J. Carrillo, M. Difrancesco, A. Figalli, T. Laurent, and D. Slepčev, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Mathematical Journal, 156 (2011). 4

[10] J. A. Carrillo, M. Chipot, and Y. Huang, *On global minimizers of repulsive–attractive power-law interaction energies*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372 (2014), p. 20130399. 4

[11] Cartan, *Théorie du potentiel newtonien : énergie, capacité, suites de potentiels*, Bulletin de la Société Mathématique de France, 73 (1945), pp. 74–106. 10

[12] L. Chizat and F. Bach, *On the global convergence of gradient descent for over-parameterized models using optimal transport*, 2018. 3

[13] A. C. de Courcel, M. Rosenzweig, and S. Serfaty, *Sharp uniform-in-time mean-field convergence for singular periodic riesz flows*, 2023. 4

[14] D. García-Zelada, *Concentration for Coulomb gases on compact manifolds*, Electronic Communications in Probability, 24 (2019), pp. 1 – 18. 9, 10

[15] Grigor'yan, *Estimates of heat kernels on riemannian manifolds*, (1999). 28

[16] Hagemann, Hertrich, Altekrüger, Beinert, Chemseddine, and Steidl, *Posterior sampling based on gradient flows of the mmd with negative distance kerne*, arXiv preprint arXiv:2310.03054, (2023). 3

[17] J. Hertrich, M. Gräf, R. Beinert, and G. Steidl, *Wasserstein steepest descent flows of discrepancies with riesz kernels*, Journal of Mathematical Analysis and Applications, 531 (2024), p. 127829. 3

[18] Y. Heurteaux, *Dimension of measures: the probabilistic approach.*, Publicacions Matemàtiques, 51 (2007), pp. 243–290. 27

[19] P.-E. Jabin and Z. Wang, *Quantitative estimates of propagation of chaos for stochastic systems with $$w^{-1,\infty }$$ w - 1 , ∞ kernels*, Inventiones mathematicae, 214 (2018), pp. 523–591. 4

[20] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the fokker–planck equation*, SIAM journal on mathematical analysis, 29 (1998), pp. 1–17. 6

[21] H. Karimi, J. Nutini, and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition*, in Machine Learning and Knowledge Discovery in Databases, P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, eds., Cham, 2016, Springer International Publishing, pp. 795–811. 10

[22] Kinderlehrer, Monsaingeon, and Xu, *A wasserstein gradient flow approach to poisson-nernst-planck equations*, arXiv preprint arXiv:1501.04437v3, (2015). 24

[23] P. Li and S. T. Yau, *On the parabolic kernel of the schrödinger operator*, (1986). 28

[24] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow*, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2001. 14, 16

[25] D. Matthes, R. McCann, and G. Savaré, *A family of nonlinear fourth order equations of gradient flow type*, arXiv preprint arXiv:0901.0540v1, (2009). 24

[26] S. Mei, A. Montanari, and P.-M. Nguyen, *A mean field view of the landscape of two-layer neural networks*, Proceedings of the National Academy of Sciences, 115 (2018). 3

[27] F. Otto, *The geometry of dissipative evolution equations: The porous medium equation*, Communications in Partial Differential Equations, 26 (2001), pp. 101–174. 6

[28] Papadimitrakis, *Notes on classical potential theory*, (2004). 9

[29] F. Santambrogio, *Optimal transport for applied mathematicians*, Birkäuser, NY, 55 (2015), p. 94. 5, 8, 20, 24

[30] Serfaty, *Coulomb gases and ginzburg-landau vortices*, 2015. 4

[31] ———, *Large systems with coulomb interactions: Variational study and statistical mechanics*, Portugaliae Mathematica, 73 (2016), pp. 247–278. 4

[32] S. Steinerberger, *A wasserstein inequality and minimal green energy on compact manifolds*, Journal of Functional Analysis, 281 (2021), p. 109076. 9

[33] N. Varopoulos, *Small time gaussian estimates of heat diffusion kernels. ii. the theory of large deviations*, Journal of Functional Analysis, 93 (1990), pp. 1–33. 28

[34] Watson, *Applications of geometric measure theory to the study of gauss-weierstrass and poisson integrals*, Annales Academiæ Scientiarum Fennicæ, 19 (194), pp. f115–132. 25, 26

A Appendix

Proof of lemma 1. Let us fix a vector field $v \in C_c^\infty$, and prove the following equality in both cases

$$\lim_{t\to 0} \frac{E_\nu((id+tv)\#\mu) - E_\nu(\mu)}{t} = \int (\nabla G) \star (\mu - \nu) \cdot v d\mu \,. \tag{147}$$

First case: If $G = -\|.\|$.

First let us remark that if $x = y$, $G(x - y + t(v(x) - v(y))) = G(x - y) = 0$. Then

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

Same for the other term of the MMD energy

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

We conclude that Formula (147) holds for the Energy Distance kernel.

Second case: If $G = \frac{1}{\|.\|^{d-2}}$. 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

$$\int G d\mu^{\otimes 2} = \int_{\mathcal{D}^c} G d\mu^{\otimes 2} \,. \tag{148}$$

Thus we can conclude that Formula (147) holds for the Coulomb kernel with similar computations, as G is 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

$$\limsup_{t\to 0} \frac{E_\nu((id+tv)\#\mu) - E_\nu(\mu)}{W_2((id+tv)\#\mu, \mu)} \leq |\partial E_\nu|(\mu) \,. \tag{149}$$

In addition,

$$\limsup_{t\to 0} \frac{W_2((id+tv)\#\mu,\mu)}{t} \leq \|v\|_{L_2(\mu)}\,, \tag{150}$$

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

$$\int_{x\neq y} (\nabla G) \star (\mu-\nu)\cdot v d\mu \leq |\partial E_\nu|(\mu)\|v\|_{L_2(\mu)}\,. \tag{151}$$

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

$$\left|\int_{x\neq y} (\nabla G) \star (\mu-\nu)\cdot v d\mu\right| \leq |\partial E_\nu|(\mu)\|v\|_{L_2(\mu)}\,. \tag{152}$$

As the vector field v chosen is arbitrary, we get

$$\|(\nabla G) \star (\mu-\nu)\|_{L_2(\mu)} \leq |\partial E_\nu|(\mu)\,, \tag{153}$$

which proves the result. □

Proof of Lemma 2. 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 than in the proof of (147), we get :

$$E_\nu(\rho) - E_\nu(\mu) = \int \xi\cdot v d\mu + o(\|v\|^2_{L_2(\mu)}). \tag{154}$$

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

$$E_\nu(\rho) - E_\nu(\mu) \geq \inf_{\gamma_0\in\Gamma_0(\mu,\rho)} \int \xi\cdot(y-x) d\gamma_0 + o(W_2^2(\mu,\rho))\,, \tag{155}$$

which concludes the proof by definition 5. □