



HAL
open science

Fast Large Deformation Matching with the Energy Distance Kernel

Siwan Boufadene, Jean Feydy, François-Xavier Vialard

► **To cite this version:**

Siwan Boufadene, Jean Feydy, François-Xavier Vialard. Fast Large Deformation Matching with the Energy Distance Kernel. 2025. <hal-05035450>

HAL Id: hal-05035450

<https://hal.science/hal-05035450v1>

Preprint submitted on 5 May 2025

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.



HAL Authorization

FAST LARGE DEFORMATION MATCHING WITH THE ENERGY DISTANCE KERNEL.

SIWAN BOUFADENE[†], JEAN FEYDY[◊], AND FRANÇOIS-XAVIER VIALARD[†]

ABSTRACT. We propose an efficient framework for point cloud and measure registration using bi-Lipschitz homeomorphisms, achieving $O(n \log n)$ complexity, where n is the number of points. By leveraging the Energy-Distance (ED) kernel, which can be approximated by its sliced one-dimensional projections, each computable in $O(n \log n)$, our method avoids hyperparameter tuning and enables efficient large-scale optimization. The main issue to be solved is the lack of regularity of the ED kernel. To this goal, we introduce two models that regularize the deformations and retain a low computational footprint. The first model relies on TV regularization, while the second model avoids the non-smooth TV regularization at the cost of restricting its use to the space of measures, or cloud of points. Last, we demonstrate the numerical robustness and scalability of our models on synthetic and real data.

Keywords. Calculus of variations · Energy-Distance kernel · Measure matching · Large Deformation Metric Mapping · Sliced approximations

Mathematics Subject Classification. 68U05, 65D18, 46E22, 58E50.

Email addresses: siwan.boufadene@live.fr, jean.feydy@inria.fr, francois-xavier.vialard@univ-eiffel.fr

CONTENTS

1	Introduction	2
1.1	Introduction to Large Deformation by Diffeomorphisms	4
1.1.1.	Deformation group	4
1.1.2.	Kernel methods	5
1.1.3.	General matching problem	5
1.1.4.	Measure matching using a positive definite kernel.	5
1.2	Reduction of the optimization set.	6
2	Interpolation using the Energy-Distance kernel, induced metrics and regularity.	6
2.1	Interpolation using a CPD kernel.	6
2.1.1.	Case of a positive definite kernel.	6
2.1.2.	Case of a \mathbb{P} conditionally positive definite kernel.	7
2.2	Case of the Energy-Distance kernel, induced norm on $\mathcal{H}_{K,\mathbb{P}}$.	9
2.3	Regularity and embeddings.	10
2.4	A few possible choices for the Energy-Distance kernel.	12
2.4.1.	The Energy-Distance as a pseudo-polynomial spline.	12
2.4.2.	Affine transformation invariance	12
2.4.3.	The Energy-Distance as a Thin-Plate Spline kernel (TPS)	12
3	Bi-Lipschitz homeomorphic matching using the Energy-Distance Kernel.	12
3.1	The variational problem on points	12
3.2	Total variation and regularity	15
3.3	Two different regularizations ensuring bi-Lipschitz homeomorphisms	17
3.3.1.	TV regularization on the dual space	18
3.3.2.	L^2 regularization on the momentum.	20
4	Sliced Energy Distance.	24
4.1	Fast estimators.	24

4.2 Use as a loss.	26
5 Experiments.	26
Acknowledgements	30
References	30
Appendix A Appendix	33

1. INTRODUCTION

The analysis of shape variations is a well-studied domain, with several applications in biomedical imaging [Rue+99; Dav97]. One of the standard tasks in this field is registering shapes, which can be of different natures, images, segmented surfaces, line bundles, etc... Several methods have been proposed over the last two decades, which can be divided into two main categories: the first category is optimization-based methods, in the LDM framework for example [Beg+05; Via+12], and the second category is learning-based methods [Dal+19; Zhu+21]. Typically, the second method requires a large database. Despite the name, recent foundation/universal models [Tia+24] may not generalize well for datasets that are extreme outliers to the set of training data classes. Fine-tuning of these models is recommended, particularly under supervision, which optimization-based methods can provide. However, these methods may suffer from long computational times, depending on the dimensionality of the input.

In this paper, we are interested in adapting regular and invertible shape registration methods in the so-called large deformation diffeomorphic metric mapping (LDDMM) framework [Tro98; Beg+05; Gla05] to the Energy-Distance kernel [Sej+13]. These deformations, which can be applied to images, clouds of points, or measures, are parameterized by time-dependent vector fields $v_t \in L^2([0, 1], V)$, where V is a (sufficiently smooth) reproducing kernel Hilbert space of vector fields defined on \mathbb{R}^d , such as Sobolev spaces of high-order. The corresponding flow ψ_t is obtained by solving $\partial_t \psi(t, x) = v(t, \psi(t, x))$ with the initial condition $\psi(0, x) = x \in \mathbb{R}^d$. Let A be a template object and B a target. The registration problem minimizes the energy functional:

$$\int_0^1 \frac{1}{2} \|v_t\|_V^2 dt + \text{Sim}(\psi(1) \cdot A, B), \quad (1.1)$$

where Sim is a similarity measure and $\psi \cdot A$ represents a natural transformation of the object A under the action of the diffeomorphism $\psi(1)$. For instance, this action is simply its image by $\psi(1)$ for point clouds. For measures, the action is the push-forward action of the map $\psi(1)$. By the Representer Theorem, the optimal v_t for a collection of points $(x_i)_{i=1, \dots, n}$ is written as $K \star \sum_{i=1}^n p_i(t) \delta_{x_i(t)}$ where K is the kernel of the RKHS and $p_i \in \mathbb{R}^d$. While this reduces the problem to finite dimensions, the regularization term $\sum_{i,j} \langle p_i, k(x_i, x_j) p_j \rangle$ retains a quadratic computational cost, typical of general kernel interaction functionals, for which trade-offs between memory and computations are sometimes necessary [Cha+21; Fey+20]. To overcome this computational bottleneck, various acceleration techniques have been proposed in these contexts, such as the fast multipole method [GR87; YBZ04] or other proximity-based algorithms such as k-d tree ([LS15] for an application in the Iterative Closest Point method). While these methods achieve efficiency by approximating the kernel, they often compromise the positivity of the kernel, a property essential for the well-posedness of the variational problem (1.1). As a result, such approximations limit the applicability of these methods within the LDDMM framework. Another approximation technique, called particle mesh method [Cot08], consists of using a mesh to encode the quantities of interest and exchange information between the particles $(x_i)_{i=1, \dots, n}$ and the grid. On the mesh, FFT can be used for a separable kernel to reduce the computational burden in $O(n \log(n))$, where n is the total grid size, by leveraging a regular grid structure. However, due to computational constraints, grid-based methods become very impractical for ambient dimensions greater than 4.

Main contribution: The first contribution of this article is to provide a simple computational framework for point clouds and measures registration in the LDMM framework with the Energy-Distance (ED) kernel. We prove that not only the gradient of the ED loss [Her+24], but also the convolution of any finite measure with the ED kernel can be approximated by a finite sum of its one-dimensional projections, with each projection being computable with an $O(n \log(n))$ time complexity. Using an Energy-Distance based loss function, this yields an efficient registration algorithm, which can be scaled to very large datasets. In addition, the ED kernel does not require any additional hyperparameter tuning, in sharp contrast with a mixture of Gaussian kernels. Using the ED kernel, we propose a registration algorithm that generates bi-Lipschitz homeomorphisms, which is a class of regular and invertible deformations strictly larger than diffeomorphisms. In that sense, our registration model is more expressive than standard LDDMM.

Main challenges: We propose to adapt the use of the energy distance kernel in the Euclidean space, defined by $K(x, y) = -|x - y|$. There are indeed several challenges that hinder the direct application of LDDMM theory with this kernel, which must be addressed to ensure theoretical well-posedness and practical feasibility. First, while the ED kernel is not positive definite but only conditionally positive, this can be resolved by constraining the so-called momentum measure $\sum_{i=1}^n p_i \delta_{x_i}$ to have zero-mean. Although this constraint reduces expressivity, the Energy-Distance registration framework naturally handles the addition of a translation term, maintaining universal approximation properties.

The second issue is the smoothness of the corresponding Reproducing Kernel space, which is not contained in C^1 functions. Typically, the Energy-Distance kernel only generates Hölder-continuous velocity fields. Since such velocity fields fail to correspond to a well-defined unique flow, additional regularization is needed in theory. We propose two parameterizations, retaining the lightweight cost of the kernel, which guarantee that the resulting deformation is a bi-Lipschitz homeomorphism. While being theoretically more expressive than LDDMM deformations, which are smooth and invertible, our formulation maintains key theoretical properties such as completeness of the group and the existence of minimizers to the variational problem.

Structure of the paper: After a brief overview of the LDDMM framework, we present our main contributions. In section 2, we study in detail the finite interpolation problem using the ED kernel and general conditionally positive kernels. We identify the corresponding velocity fields space as a Beppo-Levi space, providing regularity results for the corresponding function space. In Section 3, we study the registration problem for point clouds using the ED kernel. While proving the existence of solutions, we demonstrate potential non-injectivity, motivating our two regularization approaches. In the first model, we use a Total Variation (TV) regularization of the momentum-measure. A Representer Theorem adapted to TV regularization maintains the $O(n \log n)$ time-complexity. Our second model only deals with measure matching, where we apply a L^2 regularization of the momentum w.r.t. the current measure. This formulation is easier to optimize numerically and applies to point clouds and discrete measures. In Section 4, we present how the ED kernel computations can be sliced with $O(n \log n)$ footprint, providing statistical error bounds for the approximations. In Section 5, we detail an $O(n \log n)$ algorithm for registration, and evaluate them on synthetic and real data. The ED kernel and its sliced version frequently outperform standard kernels w.r.t. the robustness of the optimization. Moreover, it requires **no hyperparameter** tuning, offering significant advantages in various scenarios.

Perspectives: Our light computational framework for LDDMM opens the way to its use in two important applications: processing small but high-precision datasets, and handling high-dimensional biomedical data [Bun+24], possibly across different modalities. The energy functional consists of two components: the energy, which can be computed in $O(n \log n)$, and the data discrepancy, which is particularly crucial in practice. Recent advancements have demonstrated the effectiveness of incorporating higher-order information, such as through currents or varifolds [CT13]. An open question remains to develop higher-order similarity measures between point clouds that can also be computed in $O(n \log n)$.

From a theoretical perspective, our experiments suggest the further study of a Hamiltonian formulation, which appears well-suited for this framework.

1.1. Introduction to Large Deformation by Diffeomorphisms. This section provides a brief overview of the mathematical framework for large deformations by diffeomorphisms (see for instance [Gla05]).

1.1.1. Deformation group. The idea is to model diffeomorphisms, i.e. smooth invertible maps, using displacement fields $v_t \in V$, where V is a space of sufficiently smooth vector fields. In such a case, it is called admissible.

Definition 1.1 (Admissible Hilbert spaces). A Hilbert space V is said to be *admissible* if

$$(V, \|\cdot\|_V) \hookrightarrow (\mathcal{C}_0^1(\mathbb{R}^d, \mathbb{R}^d), \|\cdot\|_{1,\infty}), \quad (1.2)$$

where $\mathcal{C}_0^1(\mathbb{R}^d, \mathbb{R}^d)$ is the space of \mathcal{C}^1 fields that tend to zero at infinity, along with their partial derivatives.

Example 1.2. The Sobolev space $H^s(\mathbb{R}^d, \mathbb{R}^d)$, where $s > 0$, is defined as

$$H^s(\mathbb{R}^d, \mathbb{R}^d) = \left\{ v \in L^2(\mathbb{R}^d, \mathbb{R}^d) \mid \|v\|_s^2 := \int (1 + |\xi|^2)^s |\hat{v}(\xi)|^2 d\xi < \infty \right\}, \quad (1.3)$$

A classical result is that if $s > \frac{d}{2} + 1$, by the Sobolev embedding theorem:

$$(H^s(\mathbb{R}^d, \mathbb{R}^d), \|\cdot\|_{\mathcal{H}^s}) \hookrightarrow (\mathcal{C}_0^1(\mathbb{R}^d, \mathbb{R}^d), \|\cdot\|_{1,\infty}). \quad (1.4)$$

We will now denote V an admissible space for the rest of the section, and define the time-dependent spaces

$$L^2([0, 1], V) = \left\{ v : (t, x) \in [0, 1] \times \mathbb{R}^d \mapsto v_t(x) \in V \mid \int_0^1 \|v_t\|_V^2 dt < \infty \right\}, \quad (1.5)$$

and

$$L^1([0, 1], V) = \left\{ v : (t, x) \in [0, 1] \times \mathbb{R}^d \mapsto v_t(x) \in V \mid \int_0^1 \|v_t\|_V dt < \infty \right\} \subset L^2([0, 1], V). \quad (1.6)$$

The admissibility condition is enough to prove the following theorem:

Theorem 1.3 (see [You10]). *Let $v \in L_V^1$. Then, for all $x \in \mathbb{R}^d$, there exists a unique continuous solution $t \mapsto \psi_t^v(x)$ that verifies the Lagrangian flow equation*

$$\psi_t^v(x) = x + \int_0^t v_s \circ \psi_s^v(x) ds. \quad (1.7)$$

Moreover, for all $t \geq 0$, the application ψ_t^v is a \mathcal{C}^1 diffeomorphism of \mathbb{R}^d . Finally, the applications $(t, x) \mapsto \psi_t^v(x)$ are uniformly continuous, in a uniform way regarding v on every bounded set in L_V^2 .

The deformation space of interest is then defined as

$$\mathcal{A}_V := \{\psi_1^v \mid v \in L_V^2\}, \quad (1.8)$$

which verifies the following property:

Proposition 1.4 ((see [Tro98])). *The function space \mathcal{A}_V is a diffeomorphism group, which is complete for the metric defined by*

$$d(\text{id}, \psi) := \inf \left\{ \|v\|_{L_V^2} \mid v \in L_V^2, \psi = \psi_1^v \right\}, \quad (1.9)$$

and d is extended to arbitrary diffeomorphisms $\phi, \psi \in \mathcal{A}_V$ by right-invariance $d(\phi, \psi) := d(\text{id}, \psi \circ \phi^{-1})$.

The deformation space \mathcal{A}_V is a geodesic space:

Proposition 1.5 (Existence of constant speed geodesics). *Let $\phi, \psi \in \mathcal{A}_V$. Then there exists $v \in L_V^2$ such that $\psi_1^v = \psi \circ \phi^{-1}$ and $d(\phi, \psi) = \|v\|_{L_V^2} = \|v\|_{L_V^1}$.*

1.1.2. *Kernel methods.* In applications, the admissible space V is a Reproducible Kernel Hilbert Space associated with a given positive definite kernel K , as provided by Aronza,jn's Theorem 2.1. Very often, kernels of the form $K(x, y) = h(|x - y|)\text{id}$ are used, where h is a scalar function.

Example 1.6. • **Gaussian kernel.** Choosing the Gaussian kernel defined, for $\sigma > 0$, through:

$$k_\sigma(x, y) = \exp\left(-\frac{|x - y|^2}{2\sigma^2}\right),$$

we get the Hilbert space \mathcal{G}_σ , which is embedded in $\mathcal{C}^k(\mathbb{R}^d, \mathbb{R}^d)$ for all $k \geq 0$. Remark that $k_{\sigma'} \in \mathcal{G}_\sigma$ if and only if $\sigma' > \sigma/\sqrt{2}$, so that the velocity fields spaces \mathcal{G}_σ are all distinct.

- **Sobolev kernel.** In general, describing the generated deformation group \mathcal{A}_V for a given admissible space V is a tricky operation. However, in the case of the Sobolev kernel, whose RKHS is the Sobolev space H^s , the following result was proven in [BV16]:

$$\mathcal{A}_{H^s} = (\mathcal{D}_s(\mathbb{R}^d))_0,$$

where $(\mathcal{D}_s(\mathbb{R}^d))_0$ is the connected component of the identity in the group of Sobolev diffeomorphisms.

1.1.3. *General matching problem.* In this paragraph, V is an admissible space.

Definition 1.7 (General matching problem). Let $A : \mathcal{A}_V \rightarrow \mathbb{R}^+$ be a discrepancy function and $\lambda > 0$. The corresponding minimal matching problem is defined as

$$\min_{\psi \in \mathcal{A}_V} A(\psi) + \lambda d_V(\text{id}, \psi), \quad (1.10)$$

or in an equivalent way

$$\min_{v \in L_V^2} A(\psi_1^v) + \lambda \int_0^1 \|v_t\|_V^2 dt. \quad (1.11)$$

The following result can be written:

Theorem 1.8 (Existence of a minimizer). *Let A be defined as in 1.7. Let us suppose that $v \mapsto A(\psi_1^v)$ is weakly continuous from L_2^V to \mathbb{R} . Then there always exists a solution to the matching problem from Definition 1.7.*

1.1.4. *Measure matching using a positive definite kernel.* The previous general matching problem can be instantiated in the context of measure matching. Let μ_0 be a probability measure. If ψ_t^v is the \mathcal{C}^∞ diffeomorphism from Theorem 1.3, then the time-dependent measure $\mu_t := \psi_t^v \# \mu_0$ is a solution in the sense of distributions to the continuity equation:

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \quad (1.12)$$

for $t \in (0, 1)$ [AGS05, p. 8.1]. By a direct application of Theorem 1.8, there exist solutions to the following matching problem:

Proposition 1.9. *Let $\mu_0, \nu \in \mathcal{P}(\Omega)$ and $\mathcal{L} : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \mapsto \mathbb{R}^+$ be a lower semi-continuous loss, i.e. a function such that $\mathcal{L}(\alpha, \beta) = 0$ iff $\alpha = \beta$ in $\mathcal{P}(\Omega)$. Then the application $v \in L_2^V \mapsto \mathcal{L}(\psi_1^v \# \mu_0, \nu)$ is weakly continuous, and for $\lambda > 0$ the matching problem*

$$\min_{v \in L_2^V} E(v) := \mathcal{L}(\psi_1^v \# \mu_0, \mu_1) + \int_0^1 \|v_t\|_V^2 dt \quad (1.13)$$

has a solution in L_V^2 .

Example 1.10. There are multiple possible choices for \mathcal{L} . Some of them, that are specific to this measure context, are:

- Maximum Mean Discrepancy losses: If K is a kernel, define $\mathcal{L}_K(\mu, \nu) := \langle \mu - \nu, K \star (\mu - \nu) \rangle$. One can choose K to be a Gaussian kernel, Laplacian kernel, or the Energy-Distance kernel. For more information see [Sri+10]. In general, these discrepancies have a quadratic computational cost in terms of the number of points used to represent the measure μ . However, for the ED kernel, an approximation of it can be reduced to a $O(N \log(N))$ computational cost, which makes it attractive.
- A natural choice is to choose the optimal transport loss $\mathcal{L}(\mu, \nu) = W_2^2(\mu, \nu)$, where W_2 is L^2 Wasserstein distance, or its entropic regularization [Fey+17; LGSL22]. These two losses are at the very least quadratic if not cubic. An advantageous loss function derived from OT ideas is the sliced Wasserstein distance, whose cost can be approximated in $O(n \log(n))$.

1.2. Reduction of the optimization set. The optimization set $L^2([0, 1], V)$ can be reduced to $L^2([0, 1], (\mathbb{R}^d)^n)$ in the case of empirical measure matching with n Dirac masses (or cloud of n points). To fix the setting, we consider $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ as the source measure. Given the trajectory of the points $x_i(t) = \varphi_t(x_i)$, the minimizing vector field of the energy functional $\int_0^1 \|v_t\|_V^2 dt$ admits a sparse representation due to the Representer Theorem for RKHS. Specifically, v_t is written as $v_t(x) = \sum_{i=1}^n p_i(t) \delta_{x_i(t)}$, where $p_i(t) \in L^2([0, 1], (\mathbb{R}^d)^n)$. In terms of these new variables, the energy functional reads:

$$\frac{1}{2} \int_0^1 \langle p_i(t), K(x_i(t), x_j(t)) p_j(t) \rangle dt + \mathcal{L}_K \left(\frac{1}{n} \sum_{i=1}^n \delta_{x_i(1)}, \nu \right), \quad (1.14)$$

under the constraints $\dot{x}_i(t) = \sum_{j=1}^n K(x_i(t), x_j(t)) p_j(t)$. Without any further additional structure, the computational cost of this functional is quadratic in n .

The goal of the following sections is to adapt the ED kernel to the LDMM framework, so that efficient $O(n \log(n))$ algorithm can be performed.

2. INTERPOLATION USING THE ENERGY-DISTANCE KERNEL, INDUCED METRICS AND REGULARITY.

In this section, we present results on interpolation using a conditionally positive definite kernel. The motivation is as follows: Let μ be a probability measure on \mathbb{R}^d . To transport the measure, we consider the action of a vector field v whose restriction to $\text{supp}(\mu)$ is in $L^2(\mu)$, defined via

$$\mu \cdot v = (\text{Id} + v) \# \mu, \quad (2.1)$$

where $\#$ is the push-forward operation. Here, v is chosen to belong to a Hilbert space (resp. semi-Hilbert space), associated with a PD (resp. CPD) kernel. This space is constructed as the tensorization of the RKHS (resp. \mathbb{P} -RKSHS) with values in \mathbb{R} associated with the kernel. In the two following subsections, we present the theory of interpolation theory for the case of scalar values.

2.1. Interpolation using a CPD kernel. In this paragraph, we present the theory of interpolation using a positive definite kernel, and how it extends to conditionally positive definite (CPD) kernels [AB09].

2.1.1. Case of a positive definite kernel. Let $K : \Omega \times \Omega \rightarrow \mathbb{R}$ be a symmetric positive definite kernel, meaning that for any $(x_1, \dots, x_n) \in \Omega^n$ and $(c_1, \dots, c_n) \in \mathbb{R}^n$, we have

$$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0, \quad (2.2)$$

with equality if and only if $c_1 = \dots = c_n = 0$. Define the partial function $K_x : y \in \Omega \mapsto K(x, y)$, and let \mathcal{F}_K be the space of linear combinations of functions in the set $\{K_x; x \in \Omega\}$. The inner product

$$\langle \lambda_i K_{x_i}, \beta_j K_{y_j} \rangle_{\mathcal{F}_K} := \sum_{i,j} \lambda_i \beta_j K(x_i, y_j) \quad (2.3)$$

defines a symmetric positive bilinear form on \mathcal{F}_K . Now, an implication of Aronszajn's famous theorem [Aro50] states:

Theorem 2.1 (Aronszajn). (1) The bilinear form $\langle \cdot, \cdot \rangle_{\mathcal{F}_K}$ is positive definite.

(2) There exists a unique Hilbert space of functions \mathcal{H}_K such that

(a) $(\mathcal{F}_K, \langle \cdot, \cdot \rangle_{\mathcal{F}_K})$ is a pre-Hilbert subspace of \mathcal{H}_K .

(b) (Reproducing property) $\forall f \in \mathcal{H}_K, x \in \Omega, f(x) = \langle f, K_x \rangle_{\mathcal{H}_K}$.

The space \mathcal{H}_K is called the Reproducible Kernel Hilbert Space associated with K .

In spline interpolation, the goal is to approximate a function $f \in \mathcal{H}_K$ that is known on a finite design set $X = \{x_1, \dots, x_N\} \subset \Omega$ using a finite-dimensional approximation. More precisely, let $\mathcal{F}_K(X)$ the space of finite linear combinations of the functions $\{K_x | x \in X\}$. The previous theorem leads to the following result:

Proposition 2.2 (Interpolators in \mathcal{H}_K). Let X be a finite subset of Ω and $f \in \mathcal{H}_K$. Then, the problem

$$\min_{g \in \mathcal{F}_K(X)} \|f - g\|_{\mathcal{H}_K} \quad (2.4)$$

has a unique solution, given by $g = S_{K,X}(f)$, where $S_{K,X}$ is the orthogonal projection on $\mathcal{F}_K(X)$. Moreover, $S_{K,X}(f)$ is the interpolator of f on X with minimal \mathcal{H}_K norm.

An direct consequence of this result is the important Representer Theorem, which applies to regularized regression:

Proposition 2.3 (Representer Theorem). Let $X = (x_1, \dots, x_n) \in \Omega^n$ and $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$. Then, the solution to the regularized regression problem

$$\min_{f \in \mathcal{H}_K} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \|f\|_{\mathcal{H}_K}^2 \quad (2.5)$$

lies in $\mathcal{F}_K(X)$. Moreover, it is given by

$$f = \sum_{j=1}^n \gamma_j K_{x_j}, \quad (2.6)$$

where $\gamma = (\gamma_1, \dots, \gamma_n)$ is obtained as

$$\gamma = (K_X + \lambda \text{Id}_n)^{-1} Y. \quad (2.7)$$

2.1.2. *Case of a \mathbb{P} conditionally positive definite kernel.* Now, we extend the previous results to the case of conditionally positive definite kernels. This exposition is mainly based on [AB09; Wen04].

Definition 2.4 (\mathbb{P} -CPD kernels, point-wise version). Let $K : \Omega \times \Omega \rightarrow \mathbb{R}$ be a symmetric kernel and $\mathbb{P} \subset \mathbb{R}^\Omega$ be a finite-dimensional function space. The kernel K is said to be \mathbb{P} -conditionally positive definite if, for any distinct points $x_1, \dots, x_N \in \Omega$, $c_1, \dots, c_N \in \mathbb{R}$ satisfying

$$\forall p \in \mathbb{P}, \sum_{i=1}^n c_i p(x_i) = 0, \quad (2.8)$$

the following inequality holds

$$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0 \quad (2.9)$$

with an equality if and only if $c_1 = \dots = c_n = 0$.

Example 2.5. • The TPS kernels

Let us define the measure space

$$\mathcal{M}_{\mathbb{P}} := \{\gamma \in \mathcal{M}(\Omega) | \forall p \in \mathbb{P}, \gamma(p) = 0\}, \quad (2.10)$$

as well as the function space, for a finite set X :

$$\mathcal{F}_{K,\mathbb{P}}(X) := \{K \star \gamma | \gamma \in \mathcal{M}_{\mathbb{P}} \cap \{\text{supp}(\gamma) = X\}\}, \quad (2.11)$$

and

$$\mathcal{F}_{K,\mathbb{P}} := \bigcap_{\substack{X \subset \Omega \\ X \text{ is finite}}} \mathcal{F}_{K,\mathbb{P}}(X). \quad (2.12)$$

By Definition 2.4, the bilinear form

$$\langle \cdot, \cdot \rangle_{\mathcal{F}_{K,\mathbb{P}}} : \left(K \star \alpha = \sum_i \alpha_i K_{x_i}, K \star \beta = \sum_j \beta_j K_{y_j} \right) \in \mathcal{F}_{K,\mathbb{P}}^2 \mapsto \sum_{i,j} \alpha_i \beta_j K(x_i, y_j) \in \mathbb{R} \quad (2.13)$$

is positive definite on $\mathcal{F}_{K,\mathbb{P}}$. To generalize Aronszajn's result in this setting [AB09], we need the following definition:

Definition 2.6. A vector space \mathcal{N} equipped with a symmetric positive bilinear form $\langle \cdot, \cdot \rangle_{\mathcal{N}}$ is said to be a semi-Hilbert if \mathcal{N}/\mathcal{K} is a Hilbert space, where \mathcal{K} is the null-space of $\langle \cdot, \cdot \rangle_{\mathcal{N}}$.

Now, the generalized Aronszajn's Theorem from [AB09, Theorem 4.1] states:

Theorem 2.7. Let K be a \mathbb{P} -CPD kernel. Then, there exists a unique semi-Hilbert space $(\mathcal{H}_{K,\mathbb{P}}, \langle \cdot, \cdot \rangle_{\mathcal{H}_{K,\mathbb{P}}})$ such that

- (1) The space $(\mathcal{F}_{K,\mathbb{P}}, \langle \cdot, \cdot \rangle_{\mathcal{F}_{K,\mathbb{P}}})$ is a pre-Hilbert subspace of $(\mathcal{H}_{K,\mathbb{P}}, \langle \cdot, \cdot \rangle_{\mathcal{H}_{K,\mathbb{P}}})$.
- (2) We have the direct sum decomposition $\mathcal{H}_{K,\mathbb{P}} = \mathbb{P} \oplus \mathcal{F}_{K,\mathbb{P}}$.

To study the interpolation problem through a \mathbb{P} -CPD kernel, we need to introduce the following definition:

Definition 2.8 (\mathbb{P} -unisolvent sets). A finite set $E = \{\omega_1, \dots, \omega_l\} \subset \mathbb{R}^d$ of l distinct points is said to be \mathbb{P} -unisolvent if the application

$$I_E : \begin{array}{ccc} \mathbb{P} & \rightarrow & \mathbb{R}^l \\ p & \mapsto & (p(\omega_1), \dots, p(\omega_l)) \end{array} \quad (2.14)$$

is injective. If it is a bijection, E is said to be a *minimal* \mathbb{P} -unisolvent set. In that case remark that $\dim \mathbb{P} = l$.

Remark 2.9. If $X = \{x_1, \dots, x_n\}$ contains a \mathbb{P} -unisolvent set and (p_1, \dots, p_l) is a \mathbb{P} basis, the application

$$P_X^T : \begin{array}{ccc} \mathbb{R}^n & \rightarrow & \mathbb{R}^l \\ \gamma & \mapsto & (\sum_{k=1}^n \gamma_k p_i(x_k), i \in \llbracket 1, l \rrbracket) \end{array} \quad (2.15)$$

is a surjection.

Example 2.10. Let us introduce \mathcal{P}_m the space of polynomial functions with degree at most m .

- If $\mathbb{P} = \mathcal{P}_0$, then any non-empty set is \mathcal{P}_0 -unisolvent, and the minimal ones are the singletons.
- If $\mathbb{P} = \mathcal{P}_1$, in \mathbb{R}^d , a set of points contains a \mathcal{P}_1 -unisolvent set if and only if it is not contained in a hyperplane.
- If $\mathbb{P} = \mathcal{P}_m$ with $m \leq d-1$, then a set E fails to be \mathcal{P}_m -unisolvent if and only if there exists a nonzero polynomial $p \in \mathcal{P}_m$ such $p(\omega) = 0$, for all $\omega \in E$. The null set of p is an algebraic hypersurface of Lebesgue measure zero, and the union of such sets as p ranges over \mathcal{P}_m still has zero Lebesgue measure.

This allows to write the following result about interpolation in $\mathcal{H}_{K,\mathbb{P}}$:

Proposition 2.11. Let X be a finite subset of Ω containing a \mathbb{P} -unisolvent set. Then the problem

$$\min_{g \in \mathbb{P} \oplus \mathcal{F}_{K,\mathbb{P}}(X)} \|f - g\|_{\mathcal{H}_{K,\mathbb{P}}} \quad (2.16)$$

has a unique solution that interpolates f on X , denoted by $g = S_{K,\mathbb{P},X}(f)$. Moreover, $S_{K,\mathbb{P},X}(f)$ is the interpolator of f on X with minimal $\mathcal{H}_{K,\mathbb{P}}$ semi-norm.

A Representer Theorem, analogous to Proposition 2.3 can be proven:

Proposition 2.12. *Let $X = (x_1, \dots, x_n) \in \Omega^n$ be a collection of n distinct points containing a \mathbb{P} -unisolvent set, and let $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$. Consider the following regularized regression problem*

$$\min_{f \in \mathcal{H}_{K, \mathbb{P}}} \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \|f\|_{\mathcal{H}_{K, \mathbb{P}}}^2, \quad (2.17)$$

where $\lambda > 0$. Then, its solution lies in $\mathbb{P} \oplus \mathcal{F}_{K, \mathbb{P}}(X)$. Let (p_1, \dots, p_l) be a \mathbb{P} basis, K_X the $n \times n$ matrix with entries $K(x_i, x_j)$ and P_X the $n \times l$ matrix with entries $p_j(x_i)$. The solution f has the explicit form

$$f = \sum_{i=1}^n \gamma_i K_{x_i} + \sum_{j=1}^l \alpha_j p_j, \quad (2.18)$$

where $\gamma = (\gamma_1, \dots, \gamma_n)$ and $\alpha = (\alpha_1, \dots, \alpha_l)$ are given by

$$\begin{cases} \gamma &= (K_X + \lambda \text{Id}_n)^{-1} (Y - P_X \alpha), \\ \alpha &= (P_X^T (K_X + \lambda \text{Id}_n)^{-1} P_X)^{-1} P_X^T (K_X + \lambda \text{Id}_n)^{-1} Y. \end{cases} \quad (2.19)$$

2.2. Case of the Energy-Distance kernel, induced norm on $\mathcal{H}_{K, \mathbb{P}}$. The semi-norm defined on $\mathcal{H}_{K, \mathbb{P}}$ induces a semi-metric on \mathbb{R}^n that depends on the interpolation points $X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$. More precisely, we can state the following proposition:

Proposition 2.13. *Let K be the Energy-Distance kernel. Let \mathbb{P} be a finite-dimensional space of polynomial functions from \mathbb{R}^d to \mathbb{R} that contains the set \mathcal{P}_0 of constant polynomial functions. Let us denote (p_1, \dots, p_l) a \mathbb{P} basis. Let $X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$ be a tuple of n pairwise distinct points containing a \mathbb{P} -unisolvent set. Then the application*

$$L_X : \begin{array}{ccc} \mathcal{H}_{K, \mathbb{P}}(X) & := \mathcal{F}_{K, \mathbb{P}}(X) \oplus \mathbb{P} & \rightarrow \mathbb{R}^N \\ v & & \mapsto (v(x_1), \dots, v(x_n)) \end{array} \quad (2.20)$$

is a linear isomorphism. Moreover, using the same matrices K_X and P_X as in 2.12 and given interpolation values $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$, there exists a unique couple $(\gamma_X(Y), \alpha_X(Y)) \in \mathbb{R}^n \times \mathbb{R}^l$ such that

$$\begin{cases} Y = K_X \gamma_X(Y) + P_X \alpha_X(Y), \\ P_X^T \gamma_X(Y) = 0. \end{cases} \quad (2.21)$$

As a consequence, L_X induces a semi-metric $g_{K, \mathbb{P}, X}$ on \mathbb{R}^n through

$$g_{K, \mathbb{P}, X}(Y, Y) := \gamma_X(Y)^T K_X \gamma_X(Y). \quad (2.22)$$

The null space of this semi-metric is exactly the set $\{P_X \alpha \mid \alpha \in \mathbb{R}^l\} \subset \mathbb{R}^n$.

Proof. First, we prove that $\mathcal{H}_{K, \mathbb{P}}(X)$ is a n -dimensional space. More precisely, by defining the constraint set $\Omega_{X, \mathbb{P}} := \{\gamma \in \mathbb{R}^n \mid P_X^T \gamma = 0\}$, we show that the application

$$I_X : \begin{array}{ccc} \Omega_{X, \mathbb{P}} \times \mathbb{R}^m & \rightarrow \mathcal{H}_{K, \mathbb{P}}(X) := \mathcal{F}_{K, \mathbb{P}}(X) \oplus \mathbb{P} \\ (\gamma, \alpha) & \mapsto v(X, \gamma, \alpha) := \sum_{i=1}^n \gamma_i K_{x_i} + \sum_{j=1}^m \alpha_j p_j \end{array} \quad (2.23)$$

is a linear isomorphism. Remark that the application I_X is surjective by construction. Now, let $(\gamma, \alpha) \in \Omega_{X, \mathbb{P}} \times \mathbb{R}^l$ such that $I_X(\gamma, \alpha) = 0$. Since the sum is direct, this implies $\sum_{i=1}^n \gamma_i K_{x_i} = \sum_{j=1}^l \alpha_j p_j = 0$. As (p_1, \dots, p_l) is a \mathbb{P} basis, we get $\alpha = 0$. Next, the equation $\sum_{i=1}^n \gamma_i K_{x_i} = 0$ implies:

$$\left\| \sum_{i=1}^n \gamma_i K_{x_i} \right\|_{\mathcal{H}_{K, \mathbb{P}}}^2 = \sum_{i, j=1}^n \gamma_i \gamma_j K(x_i, x_j) = 0. \quad (2.24)$$

However, since $\mathcal{P}_0 \subset \mathbb{P}$, we have $\sum_i \gamma_i = 0$, so that the equality above holds if and only if $\gamma = 0$, proving that I_X is a linear isomorphism. Finally, since P_X^T has rank l as X contains a unisolvent set, we obtain $\dim \Omega_{X, \mathbb{P}} = n - l$. Consequently, $\dim \mathcal{H}_{K, \mathbb{P}}(X) = n$. From [AB09, Lemma 5.1, 5.2], this proves that L_X is a linear isomorphism. As an additional corollary of the isomorphism I_X , there exists a unique couple $(\gamma_X(Y), \alpha_X(Y)) \in \Omega_{X, \mathbb{P}} \times \mathbb{R}^l$ such that Equation (2.21) is satisfied, given

values $Y = (y_1, \dots, y_n) \in \mathbb{R}^N$. Consequently, the semi-metric $g_{K, \mathbb{P}, X}$ is well-defined, and we get the equivalence:

$$g_{K, \mathbb{P}, X}(Y, Y) = 0 \Leftrightarrow \gamma_X(Y) = 0 \Leftrightarrow \exists \alpha_X(Y) \in \mathbb{R}^l \text{ such that } Y = P_X \alpha_X(Y). \quad (2.25)$$

□

This provides a way to induce a non-degenerate metric on \mathbb{R}^n , although other choices are possible:

Corollary 2.14. *With the previous notations, if $\lambda > 0$ the formula:*

$$\tilde{g}_{K, \mathbb{P}, X, \lambda}(Y, Y) := \gamma_X(Y)^T K_X \gamma_X(Y) + \lambda \alpha_X(Y)^T P_X^T P_X \alpha_X(Y) \quad (2.26)$$

defines a metric on \mathbb{R}^n .

When interpolating velocity fields, extending this choice by tensorization to a metric on $\mathbb{R}^{N \times d}$, using the same λ for each dimension, ensures rotation invariance. Additionally, it naturally identifies the space of translations with Euclidean space, yielding an isotropic metric on \mathbb{P} that better preserves geometric structure.

2.3. Regularity and embeddings. In order to characterize the regularity of these spaces, let us introduce the Sobolev spaces, which are defined, for any $s \in \mathbb{R}$, as:

$$H^s := \left\{ v \in \mathcal{S}' \mid \|v\|_s < \infty \right\}, \quad (2.27)$$

where \mathcal{S}' is the space of tempered distributions, $\|\cdot\|_s$ is the classical Sobolev norm, defined via

$$\|v\|_{s,2}^2 := \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{v}(\xi)|^2 d\xi. \quad (2.28)$$

Now, the classical Sobolev embedding theorem states [AF03]:

Theorem 2.15. *Let $s > d/2$. Then*

$$H^s \hookrightarrow \mathcal{C}^0, \quad (2.29)$$

where \mathcal{C}^0 is the space of continuous functions endowed with the uniform norm $\|\cdot\|_\infty$. Moreover, there exists $\lambda > 0$ such that

$$H^s \hookrightarrow \mathcal{C}^{0,\lambda}. \quad (2.30)$$

where $\mathcal{C}^{0,\lambda}$ is the set of Hölder continuous functions equipped with the norm $\|\cdot\|_{\lambda,\infty} = \|\cdot\|_\infty + |\cdot|_\lambda$, and $|\cdot|_\lambda$ is the Hölder semi-norm.

Now, following [AST04], let us define the space

$$\dot{H}^s := \left\{ v \in \mathcal{S}' \mid |v|_s < \infty \right\}, \quad (2.31)$$

where \mathcal{S}' is the set of tempered distributions on \mathbb{R}^d , and $|\cdot|_s$ is the Sobolev semi-norm, defined as

$$|v|_s^2 := \int_{\mathbb{R}^d} |\xi|^{2s} |\hat{v}(\xi)|^2 d\xi. \quad (2.32)$$

The L^2 Beppo-Levi spaces are defined as

$$X^{m,s} := \left\{ v \in \mathcal{D}' \mid \forall \alpha \in \mathbb{N}^d, |\alpha| = m, \partial^\alpha v \in \dot{H}^s \right\}, \quad (2.33)$$

endowed with the semi-norm:

$$|v|_{m,s} := \left(\sum_{|\alpha|=m} |\partial^\alpha v|_s^2 \right)^{1/2}. \quad (2.34)$$

Now, let us suppose the following condition on $m \in \mathbb{N}^*$ and $s \geq 0$:

$$-m + \frac{d}{2} < s < \frac{d}{2}. \quad (2.35)$$

Then the following embedding theorem is valid [AST04, Sec. I-2]:

Theorem 2.16. *Let us suppose that (2.35) holds.*

- *Let $E \subset \mathbb{R}^d$ be a finite subset, and define*

$$\|v\|_{E,m,s} := \left(\sum_{\omega \in E} |v(\omega)|^2 + |v|_{m,s}^2 \right)^{1/2}. \quad (2.36)$$

Then, if E is a \mathcal{P}_{m-1} -unisolvent set, $\|\cdot\|_{E,m,s}$ defines a Hilbertian norm on $X^{m,s}$, whose topology is independent of E . Moreover, we have the continuous embedding:

$$X^{m,s} \hookrightarrow \mathcal{C}^0. \quad (2.37)$$

- *Let Ω be an open subset of \mathbb{R}^d with Lipschitz boundary. Then the operator R_Ω of restriction to Ω is linear and continuous from $X^{m,s}$ to $H^{m+s}(\Omega)$.*

The following corollary allows for a different parametrization of a Hilbertian norm on $X^{m,s}$ when $m = 1$:

Corollary 2.17 (Parametrization through the mean on a design set). *Suppose that (2.35) holds, with $m = 1$. Let X be a non-empty finite subset of n points in \mathbb{R}^d , and define*

$$\|v\|_{\Sigma X,m,s} := \left(\left| \frac{1}{n} \sum_{x \in X} v(x) \right|^2 + |v|_{m,s}^2 \right)^{1/2}. \quad (2.38)$$

Then $\|\cdot\|_{\Sigma X,m,s}$ defines a Hilbertian norm on $X^{m,s}$, whose topology is independent of X . Moreover, it is equivalent to $\|\cdot\|_{\{\omega\},m,s}$ from Theorem 2.16, for any $\omega \in \mathbb{R}^d$, so that the continuous embedding $X^{m,s} \hookrightarrow \mathcal{C}^0$ still holds with the norm $\|\cdot\|_{\Sigma X,m,s}$.

Proof. Let $\omega \in \mathbb{R}^d$. Then the set $\{\omega\}$ is \mathcal{P}_0 -unisolvent, so that we can apply Theorem 2.16. This means there exists $C_\omega > 0$ so that, for all $x \in X$:

$$|v(x)| \leq C_\omega \|v\|_{\{\omega\},m,s}, \quad (2.39)$$

so that

$$\|\cdot\|_{\Sigma X,m,s} \leq (1 + C_\omega^2)^{1/2} \|\cdot\|_{\{\omega\},m,s}. \quad (2.40)$$

Now let us prove that $(X^{m,s}, \|\cdot\|_{\Sigma X,m,s})$ is a Banach space. Let u_j be a Cauchy sequence in this space. Then:

$$\exists \tilde{u} \in X^{m,s}, \lim_j |u_j - \tilde{u}|_{m,s} = 0, \quad (2.41)$$

$$\exists \eta \in \mathbb{R}, \lim_j \frac{1}{n} \sum_{x \in X} u_j(x) = \eta. \quad (2.42)$$

Moreover, there exists $\lambda \in \mathbb{R}$ so that

$$\frac{1}{n} \sum_{x \in X} \tilde{u}(x) + \lambda = \eta. \quad (2.43)$$

This proves that u_j converges to $\tilde{u} + \lambda$ with respect to the $\|\cdot\|_{\Sigma X,m,s}$ norm, proving $(X^{m,s}, \|\cdot\|_{\Sigma X,m,s})$ is a Banach space. We can apply the Banach Isomorphism Theorem [Bre99, p. II.6] to obtain the equivalence of the two norms. The rest of the statement is a simple application of Theorem 2.16. \square

Now, let us define:

Definition 2.18 ((m, s) splines kernels). Let m, s satisfying (2.35), and denote $\nu := 2m + 2s - d$. The associated (m, s) spline kernel is defined as

$$G_{m,s}(x) = \begin{cases} (-1)^{\lceil \nu \rceil} |x|^{2\nu}, & \nu \notin \mathbb{N}^*, \\ (-1)^{\nu+1} |x|^{2\nu} \log |x|, & \nu \in \mathbb{N}^*. \end{cases} \quad (2.44)$$

From [Wen04, Th 10.43] and [AST04, Sec. II], we obtain the following characterization :

Theorem 2.19. *Let m, s satisfying (2.35), and define \mathcal{P}_{m-1} the space of d -variate polynomials of degree at most $m-1$. Then the kernel $G_{m,s}$ is \mathcal{P}_{m-1} -CPD. Moreover, the associated semi-Hilbert space is a Beppo-Levi space, given by:*

$$\mathcal{H}_{G_{m,s}, \mathcal{P}_{m-1}} = \mathcal{P}_{m-1} \oplus \mathcal{F}_{G_{m,s}, \mathcal{P}_{m-1}} = X^{m,s}. \quad (2.45)$$

Furthermore, $X^{m,s}$ being endowed with the semi-Hilbert norm $|\cdot|_{m,s}$, the semi-inner products of these spaces are the same, in the sense that, for all $f \in X^{m,s}$:

$$|v|_{m,s} = |v|_{\mathcal{H}_{G_{m,s}, \mathcal{P}_{m-1}}}. \quad (2.46)$$

2.4. A few possible choices for the Energy-Distance kernel. The Energy-Distance kernel $x \mapsto -|x|$ corresponds to an (m, s) spline kernel, when $2m + 2s - d = 1$. This leads to several possible choices.

2.4.1. The Energy-Distance as a pseudo-polynomial spline. Choosing $m = 1$ and $s = \frac{d-1}{2}$, the kernel $G_{1, \frac{d-1}{2}}$ is referred to as a *pseudo-polynomial* [AST04, Sec II-3] or *multiquadric* [Har90] spline. We can apply Proposition 2.13, with $\mathbb{P} = \mathcal{P}_0$. The null space of the semi-metric $g_{K, \mathbb{P}, X}$ is a one-dimensional set, equal to $\mathbb{R}\mathbf{1}_N$, where $\mathbf{1}_N = (1, \dots, 1) \in \mathbb{R}^N$. In the context of velocity field interpolation, this corresponds to imposing no penalty on translations. Specifically, translations are the null-space of the tensorized semi-metric.

2.4.2. Affine transformation invariance. Choosing $m = 2$ and $s = \frac{d-3}{2}$ leads to $\mathbb{P} = \mathcal{P}_1$ and applying Proposition 2.13, this time the null space of the associated metric $g_{K, \mathbb{P}, X}$ is exactly the set of vectors $Y_0 = (f(x_i) + b)_{i \in [1, N]}$, where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is a linear form and $b \in \mathbb{R}$. In the context of velocity field interpolation, this corresponds to imposing no penalty on affine deformations, which are the null-space of the tensorized semi-metric.

2.4.3. The Energy-Distance as a Thin-Plate Spline kernel (TPS). The TPS kernels [Boo89; CY01] are defined as (m, s) spline kernels for $s = 0$. Choosing $m = \frac{d+1}{2}$ if d is odd, condition (2.35) is verified, and all the embedding theorems above are true. Applying Proposition 2.13 with $\mathbb{P} = \mathcal{P}_{m-1}$, the set $\{P_X \alpha, \alpha \in \mathbb{R}^m\}$ is the null space of the induced semi-metric $g_{K, \mathbb{P}, X}$. As in this paragraph $\mathbb{P} = \mathcal{P}_{\frac{d-1}{2}}$, for $d \geq 2$, this space contains, for example, all the vectors that can be written as $Y_0 = (f(x_i) + b)_{i \in [1, N]}$, where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is a linear form and $b \in \mathbb{R}$. However, it contains many more if $d \geq 5$.

3. BI-LIPSCHITZ HOMEOMORPHIC MATCHING USING THE ENERGY-DISTANCE KERNEL.

We now show that regularization is necessary to obtain invertible and regular maps by proving that the so-called Landmark space endowed with the Energy distance kernel as the co-metric is not a complete Riemannian space. In other words, by solving the standard matching problem with the ED kernel, one can obtain non-invertible solutions of the variational problem.

3.1. The variational problem on points. In this section, we study the variational problem on the space of landmarks (i.e. n distinct points), **without** any regularization on the vector field other than the norm induced by the kernel. Specifically, we aim to match n distinct points $(x_i)_{i=1, \dots, n} \in \mathbb{R}^d$ with n distinct target points $(y_i)_{i=1, \dots, n} \in \mathbb{R}^d$. The displacement is modeled through velocity fields v_t , which are regularized through a norm $\|\cdot\|_{V_{m,s}}$. Here, $\|\cdot\|_{V_{m,s}}$ is any of the equivalent norms defined in Theorem 2.16, or in Corollary 2.17 if $m = 1$. We assume that condition (2.35) is satisfied, and denote by $V_{m,s}$ the corresponding Beppo-Levi space endowed with the norm $\|\cdot\|_{V_{m,s}}$.

Lemma 3.1 (Existence of trajectories). *Let $v \in L^2([0, 1], V_{m,s})$. Then, there exists a solution to the equation*

$$\frac{d}{dt} x_i(t) = v_t(x_i(t)), \quad (3.1)$$

with the initial condition $x_i(t) = x_i$.

Proof. This is a simple application of the Caratheodory existence theorem. Indeed, the space $V_{m,s}$ is continuously embedded in $(\mathcal{C}^0, \|\cdot\|_\infty)$ by Theorem 2.16 or Corollary 2.17. \square

Remark 3.2. The trajectories are not necessarily unique for general v_t , as illustrated in [AC14].

However, they are uniformly continuous in time:

Proposition 3.3. *Let x_i be a solution to Equation (3.1). Then $t \in [0, 1] \mapsto x_i(t)$ is 1/2-Hölder continuous.*

Proof. First, let us remark that any solution to Equation (3.1) verifies, for all $t \geq 0$:

$$x_i(t) = x_i(0) + \int_0^t v_s(x_i(s)) ds. \quad (3.2)$$

This proves, using the Cauchy-Schwartz inequality and the continuous embedding $V_{m,s} \hookrightarrow \mathcal{C}^0$:

$$|x_i(t) - x_i(s)| \leq C_{m,s} \sqrt{|t - s|} \|v\|_{L^2([0,1], V_{m,s})}. \quad (3.3)$$

for a constant $C_{m,s} > 0$. \square

Now, we study the following problem:

$$\min_{v \in L^2([0,1], V_{m,s})} \int_0^1 \|v_t\|_{V_{m,s}}^2 dt, \quad (3.4)$$

under the constraint that $\psi(t = 1, x_i) = y_i$ for $(x_i)_{i=1, \dots, n} \in \mathbb{R}^d$, n distinct points as well as $(y_i)_{i=1, \dots, n} \in \mathbb{R}^d$.

Proposition 3.4. *There exist solutions to the variational problem (3.4).*

Proof. Let v^k be a minimizing sequence for Problem (3.4). We can assume that $\|v^k\|_{L^2([0,1], V_{m,s})}$ is uniformly bounded in k , and by applying Proposition (3.3), they are uniformly continuous. We can apply the Arzela-Ascoli Theorem, proving the existence of a limit velocity v^∞ that is a solution to the matching problem (3.4). Now, passing to the limit in (3.2), the limiting curves x_i^∞ are solutions to the ODE equation associated with v^∞ . Moreover, by uniform convergence, $x_i^\infty = y_i$. Finally, v^∞ has minimal $\|\cdot\|_{L^2([0,1], V_{m,s})}$ norm by lower semi-continuity. \square

Remark 3.5. Due to the non-uniqueness of solutions to the ODE, it may be possible to find non-diffeomorphic solutions. In other words, from n distinct points, fewer points can be obtained during the evolution after a local collapse. Indeed, the non-uniqueness of solutions enables branching trajectories.

While the space $V_{m,s}$ is infinite dimensional, the solutions to Problem (3.4) are sparse, in the following sense:

Proposition 3.6 (Moment representation). *Let $v \in L^2([0, 1], V_{m,s})$ be a solution to Problem (3.4), and denote $X_t = (x_i(t), 1 \leq i \leq n)$. Let us denote $G_{m,s}$ as the corresponding (m, s) -spline kernel from Definition 2.18. Then, v has the following form:*

$$v_t(x) = \sum_{i=1}^n \gamma_i(t) G_{m,s}(x - x_i(t)) + p_t(x), \quad (3.5)$$

where $p_t \in \mathcal{P}_{m-1}$ at all times and $\gamma_i(t) \in \mathbb{R}^d$ for all i verifies the constraint:

$$\forall p \in \mathcal{P}_{m-1}, \sum_{i=1}^n \gamma_i(t) p(x_i(t)) = 0. \quad (3.6)$$

Proof. From Proposition 2.11, for all times $t \geq 0$, the projection $S_{G_{m,s}, \mathcal{P}_{m-1}, X_t}(v_t)$ is written as in Equation (3.5) with $\gamma_i(t)$ verifying the constraints (3.6), interpolates v_t on X_t , and verifies

$$|S_{G_{m,s}, \mathcal{P}_{m-1}, X_t}(v_t)|_{m,s} \leq |v_t|_{m,s},$$

with equality if and only if $v_t = S_{G_{m,s}, \mathcal{P}_{m-1}, X_t}(v_t)$. This completes the proof. \square

Example 3.7 (Two particles case). Let us illustrate through a simple but instructive example the non-completeness of the landmark manifold endowed with this metric. In dimension 2, we consider the Energy-Distance kernel as a $(1, \frac{1}{2})$ -spline (see 2.4.1). We denote V the associated Beppo-Levi space. Let $x_+ = (\frac{1}{2}r_0, 0)$ and $x_- = (-\frac{1}{2}r_0, 0)$ be two points in \mathbb{R}^2 . We consider the target points $x_+^\varepsilon = (\frac{1}{2}\varepsilon, 0)$ and $x_-^\varepsilon = (-\frac{1}{2}\varepsilon, 0)$, where $r_0 > \varepsilon > 0$. Let us suppose that there exists $v \in L^2([0, 1], V)$ solution to the corresponding minimal energy matching problem (3.4). Thanks to the symmetries of the problem, the constant translation part of the optimal velocity field is zero, and the optimal moments belong in $\mathbb{R}(1, 0)$. This means that the flow only depends on a time-continuous moment $t \mapsto \gamma_t \in \mathbb{R}$, in the sense that the optimal velocity field v_t has the following form on \mathbb{R}^2 :

$$\forall t \in [0, 1], x \in \mathbb{R}^2, \exists \gamma_t \in \mathbb{R} \mid v_t(x) = \begin{pmatrix} \gamma_t \\ 0 \end{pmatrix} \cdot (|x - x_+(t)| - |x - x_-(t)|). \quad (3.7)$$

Denoting $r_t := |x_+(t) - x_-(t)|$, the energy at time t is given by:

$$\|v_t\|_V^2 = \gamma_t^2 r_t. \quad (3.8)$$

Through a simple reparametrization argument [Gla05, Lemme 6], the optimal velocity fields are such that $t \mapsto \|v_t\|_V^2$ is constant in time, equal to $\lambda := \gamma_0^2 r_0$. Let us suppose for now that $t \mapsto \gamma_t$ and $t \mapsto r_t$ are smooth functions on $(0, 1)$, strictly positive with non-zero derivatives. Then, as

$$\begin{cases} x_+(t) = x_+ - \int_0^t \gamma_s r_s ds, \\ x_-(t) = x_- + \int_0^t \gamma_s r_s ds, \end{cases} \quad (3.9)$$

we get:

$$r_t = r_0 - 2 \int_0^t \gamma_s r_s ds, \quad (3.10)$$

so that:

$$\dot{r}_t = -2\gamma_t r_t. \quad (3.11)$$

Moreover, the following conservation equation is verified:

$$\gamma_t^2 \dot{r}_t + 2r_t \gamma_t \dot{\gamma}_t = 0, \quad (3.12)$$

so that the final differential equations read:

$$\begin{cases} \dot{\gamma}_t = \gamma_t^2, \\ \dot{r}_t = -2\sqrt{\lambda}\sqrt{r_t}. \end{cases} \quad (3.13)$$

The first equation is a well-known diverging equation, whose solution with initial conditions γ_0 exists on the open time interval $[0, \frac{1}{\gamma_0})$, and is written as:

$$\gamma_t = \frac{\gamma_0}{1 - \gamma_0 t}. \quad (3.14)$$

The corresponding r_t , thanks to the constant value of $\|v_t\|_V^2$, is given by:

$$r_t = r_0(1 - \gamma_0 t)^2. \quad (3.15)$$

Choosing $\gamma_0 = 1 - \sqrt{\frac{\varepsilon}{r_0}}$ solves the matching problem, and justifies all the computations above. The final matching energy to match (x_+, x_-) with $(x_+^\varepsilon, x_-^\varepsilon)$ is then equal to $\lambda_{r_0, \varepsilon} = \left(1 - \sqrt{\frac{\varepsilon}{r_0}}\right)^2 r_0$. Let us make a few remarks. First, $\lambda_{r_0, \varepsilon}$ has a limit as $\varepsilon \rightarrow 0+$, equal to r_0 . This proves that the landmark manifold is not complete for this metric: there exists a sequence $v^n \in L^2([0, 1], V)$ with uniformly bounded energy such that $\lim_{n \rightarrow \infty} x_+^{v^n}(1) = \lim_{n \rightarrow \infty} x_-^{v^n}(1) = 0$. Due to the non-uniqueness of solutions to the flow, there exists a solution to the matching Problem (3.4) under the constraints $x_+(1) = x_-$ and $x_-(1) = x_+$, even in dimension one. Indeed, the two points can collapse at 0 in finite time. Then, by reversing the solution in time, one can exchange the position of the two particles.

However, this situation is specific to this configuration, as a small perturbation in the positions induces a smooth path. Below is an illustration of the deformation generated on \mathbb{R}^2 , and of a matching between the perturbed landmarks $(x_+ + (\frac{0}{\varepsilon}), x_- - (\frac{0}{\varepsilon}))$ and $(x_- + (\frac{0}{\varepsilon}), x_+ - (\frac{0}{\varepsilon}))$, where $\varepsilon = 10^{-3}$.

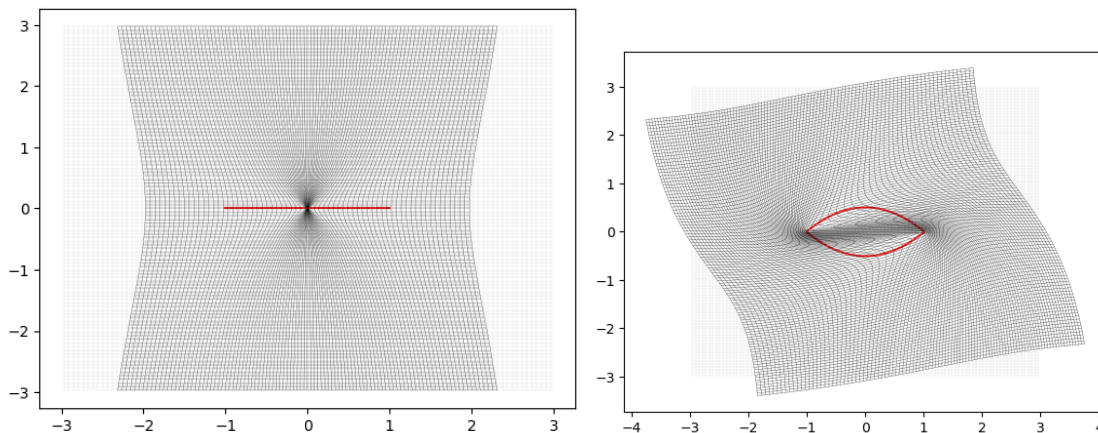


FIGURE 1. Obtained deformation map while matching two particles to 0 (left), and between the landmarks $(x_+ + (\frac{0}{\varepsilon}), x_- - (\frac{0}{\varepsilon}))$ and $(x_- + (\frac{0}{\varepsilon}), x_+ - (\frac{0}{\varepsilon}))$ (right).

In summary, achieving homeomorphisms for the landmark problem requires regularization. We will examine a potential regularization approach in the following section.

3.2. Total variation and regularity. In this section, we show that a bound on the total variation of the dual of the velocity field is sufficient to ensure Lipschitz regularity in the case of the ED kernel. Hereafter, K denotes the Energy-Distance kernel, and $V_{m,s}$ is the corresponding Beppo-Levi space (see subsection 2.4). In Example 3.7, v_t is given by $v_t = K \star \Gamma_t$, where Γ_t is the zero sum vector-valued measure defined by:

$$\Gamma_t := \begin{pmatrix} \gamma_t \\ 0 \end{pmatrix} (\delta_{x_-(t)} - \delta_{x_+(t)}) . \quad (3.16)$$

When matching (x_+, x_-) to $(0_2, 0_2)$, the total variation map $t \mapsto |\Gamma_t|_{TV}$ is not in $L^1([0, 1])$. However, a regularity result can be obtained if we have such a bound on the total variation:

Proposition 3.8 (Lipschitz regularity under TV control). *Let $t \mapsto v_t$ be a time-dependent velocity field written as:*

$$v_t = K \star \Gamma_t , \quad (3.17)$$

where K is the Energy-Distance kernel and Γ_t is a measure in $\mathcal{M}_{\mathcal{P}_{m-1}}$ such that:

$$\int_0^1 |\Gamma_t|_{TV} dt < +\infty . \quad (3.18)$$

Then v_t is Lipschitz for almost every $t \in [0, 1]$ and there exists a bi-Lipschitz homeomorphism curve $t \mapsto \psi_t$ such that ψ_t is the unique solution to the flow equation

$$\begin{cases} \partial_t \psi_t = v_t \circ \psi_t \\ \psi_0 = \text{Id} . \end{cases} \quad (3.19)$$

Proof. Since $\Gamma_t \in L^1([0, 1], |\cdot|_{TV})$, the quantity $|\Gamma_t|_{TV}$ is bounded for almost every $t \in [0, 1]$. Thus, for any $x, y \in \mathbb{R}^d$ and almost every t , we have:

$$|v_t(x) - v_t(y)| = \left| \int (|x - z| - |y - z|) d\Gamma_t(z) \right| \quad (3.20)$$

$$\leq \int |x - y| d|\Gamma_t|_{TV}(z) \quad (3.21)$$

$$|v_t(x) - v_t(y)| \leq |\Gamma_t|_{TV} |x - y|. \quad (3.22)$$

This shows that v_t is Lipschitz for almost every $t \in [0, 1]$, with a Lipschitz constant such that:

$$\int_0^1 \text{Lip}(v_t) dt < +\infty. \quad (3.23)$$

Under these conditions, the existence of the flow ψ_t for all $t \in [0, 1]$ is well known, applying the Cauchy-Lipschitz theorem [AC14]. We can prove that this map is also Lipschitz. Indeed, if $x, y \in \mathbb{R}^d$, then:

$$|\psi_t(x) - \psi_t(y)| = \left| x - y + \int_0^t v_s(\psi_s(x)) - v_s(\psi_s(y)) ds \right| \quad (3.24)$$

$$\leq |x - y| + \int_0^t |v_s(\psi_s(x)) - v_s(\psi_s(y))| ds \quad (3.25)$$

$$|\psi_t(x) - \psi_t(y)| \leq |x - y| + \int_0^t |\Gamma_s|_{TV} |\psi_s(x) - \psi_s(y)| ds, \quad (3.26)$$

so that a simple application of the Grönwall Lemma yields:

$$|\psi_t(x) - \psi_t(y)| \leq |x - y| \exp\left(\int_0^t |\Gamma_s|_{TV} ds\right). \quad (3.27)$$

Finally, defining the inverse flow ψ^{-1} by:

$$\partial_t(\psi_t^{-1}) = -v_{1-t} \circ \psi_t^{-1}, \quad (3.28)$$

we can apply the same analysis, proving that ψ_t is a bi-Lipschitz homeomorphism. \square

Example 3.7 above shows that the MMD functional $\Gamma \in \mathcal{M}_{\mathcal{P}_0} \mapsto \langle \Gamma, K \star \Gamma \rangle \in \mathbb{R}$ is not coercive (in general) with respect to the total variation norm. However, in the case of a sum of distinct Dirac masses, a bound on the TV norm can be obtained that depends on the minimal distance between points.

Proposition 3.9. *Let $\sigma, r > 0$, and let $B_{\sigma r}$ be the closed ball of radius σr centered at 0 in \mathbb{R}^d . Consider the subspace of $\mathcal{M}_{\mathcal{P}_{m-1}}(B_{\sigma r})$ defined as:*

$$\mathcal{M}_{\mathcal{P}_{m-1}, n, r}(B_{\sigma r}) = \left\{ \Gamma = \sum_{i=1}^n \gamma_i \delta_{x_i} \in \mathcal{M}_{\mathcal{P}_{m-1}}(B_{\sigma r}) \mid \min_{i \neq j} |x_i - x_j| \geq r \right\}. \quad (3.29)$$

Then, there exists $\lambda_\sigma > 0$ such that, for all $\Gamma \in \mathcal{M}_{\mathcal{P}_{m-1}, n, r}(B_{\sigma r})$ we have:

$$|\Gamma|_{TV} \leq \lambda_\sigma \sqrt{\frac{n}{r}} \|\Gamma\|_K, \quad (3.30)$$

where $\|\Gamma\|_K = \langle \Gamma, K \star \Gamma \rangle$.

Proof. Let us denote $L_{n, r} = \left\{ X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n \mid \min_{i \neq j} |x_i - x_j| \geq r \right\}$. First, remark that if $\Gamma \in \mathcal{M}_{\mathcal{P}_{m-1}, n, r}(B_{\sigma r})$, then the scaled measure $d_{\lambda \#} \Gamma$, where $d_\lambda : x \mapsto \lambda x$, belongs to $\mathcal{M}_{\mathcal{P}_{m-1}, n, \lambda r}(B_{\lambda \sigma r})$.

Next, define K_X the matrix with entries $K(x_i - x_j)$ and consider the linear space:

$$C_X = \left\{ \gamma = (\gamma_1, \dots, \gamma_n) \in \mathbb{R}^n \mid \forall p \in \mathcal{P}_{m-1}, \sum_{i=1}^n \gamma_i p(x_i) = 0 \right\}. \quad (3.31)$$

Since the application $\gamma \in C_X \cap \mathcal{S}_{n-1} \mapsto \langle \gamma, K_X \gamma \rangle \in \mathbb{R}$ is continuous and takes strictly positive values, since $C_X \cap \mathcal{S}_{n-1}$ is compact, there exists $\rho_X > 0$ such that:

$$\forall \gamma \in C_X, \rho_X |\gamma|_2^2 \leq \langle \gamma, K_X \gamma \rangle, \quad (3.32)$$

where $|\cdot|_2$ denotes the Euclidean distance in \mathbb{R}^n . This proves:

$$\forall \gamma \in C_X, |\gamma|_2 \leq \sqrt{\frac{1}{\rho_X}} \sqrt{\langle \gamma, K_X u \rangle}. \quad (3.33)$$

Now, for $\sigma, r > 0$, define the set:

$$\mathcal{S}(r, \sigma) := L_{n,r} \cap (B_{\sigma r})^n. \quad (3.34)$$

The mapping $\rho : X \mapsto \rho_X$ is continuous on $\mathcal{S}(r, \sigma)$, which is a compact set. Therefore, there exists $\rho_r(\lambda) > 0$ such that for all $X \in \mathcal{S}(r, \sigma)$, $\rho_r(\sigma) \leq \rho_X$. Now, let us prove, given $\sigma > 0$, that $r \mapsto \rho_r(\sigma)$ is 1-homogeneous. From the equalities, for $r > 0$: $\mathcal{S}(r, \sigma) = r\mathcal{S}(1, \sigma)$ and $K_{rX} = rK_X$ for any collection of distinct points X , we get the equivalences:

$$\forall X \in \mathcal{S}(1, \sigma), \forall \gamma \in C_X, \rho_1(\sigma) |\gamma|_2^2 \leq \langle \gamma, K_X \gamma \rangle \quad (3.35)$$

$$\Leftrightarrow \forall X \in \mathcal{S}(1, \sigma), \forall \gamma \in C_{rX}, \rho_1(\sigma) |\gamma|_2^2 \leq \frac{1}{r} \langle \gamma, K_{rX} \gamma \rangle \quad (3.36)$$

$$\Leftrightarrow \forall Y \in \mathcal{S}(r, \sigma), \forall \gamma \in C_Y, r\rho_1(\sigma) |\gamma|_2^2 \leq \langle \gamma, K_Y \gamma \rangle. \quad (3.37)$$

Thus, for any $X \in L_{n,r} \cap (B_{\sigma r})^n = \mathcal{S}(r, \sigma)$ and $\gamma \in C_X$, we have:

$$r\rho_1(\sigma) |\gamma|_2^2 \leq \langle \gamma, K_X \gamma \rangle. \quad (3.38)$$

Now, denoting $\lambda_\sigma := \rho_1(\sigma)^{-1/2}$, we get:

$$\forall X \in L_{n,r} \cap (B_{\sigma r})^n, \forall \gamma \in C_X, |\gamma|_2 \leq \lambda_\sigma \sqrt{\frac{1}{r}} \sqrt{\langle \gamma, K_X \gamma \rangle}. \quad (3.39)$$

The result follows from the fact that for any $\Gamma \in \mathcal{M}_{\mathcal{P}_{m-1,n,r}(B_{\sigma r})}$, there exists $X = (x_1, \dots, x_n) \in L_{n,r} \cap (B_{\sigma r})^n$ and $\gamma \in C_X$ such that: $\Gamma = \sum_{i=1}^n \gamma_i \delta_{x_i}$, and by remarking that $|\Gamma|_{TV} = \sum_{i=1}^n |\gamma_i| \leq \sqrt{n} |\gamma|_2$ and $\|\Gamma\|_K = \sqrt{\langle \gamma, K_X \gamma \rangle}$. \square

Remark 3.10. Note that this result immediately generalizes to vector-valued measures. If Γ is written as: $\Gamma = \sum_{i=1}^n \gamma_i \delta_{x_i}$, where $\gamma_i = (\gamma_{i,1}, \dots, \gamma_{i,d}) \in \mathbb{R}^d$ for all $1 \leq i \leq n$, then:

$$|\Gamma|_{TV} = \sum_{i=1}^n \sqrt{\sum_{k=1}^d |\gamma_{i,k}|^2} \leq \lambda_\sigma \sqrt{\frac{n}{r}} \sqrt{\sum_{k=1}^d \langle \gamma_{\cdot,k}, K_X \gamma_{\cdot,k} \rangle}. \quad (3.40)$$

Example 3.11. In the limit $\varepsilon \rightarrow 0$ in Example 3.7, the solution moment measure Γ_t satisfies $|\Gamma_t|_{TV} = \frac{1}{1-t}$ and $\|\Gamma_t\|_K = r_0$ for all times $t \in [0, 1)$. Additionally, the distance between the two particles at time $0 \leq t < 1$ is given by $r_t = r_0(1-t)^2$, so that Γ_t is in the measure set $\mathcal{M}_{\mathcal{P}_0, 2, r_t}(B_{r_t}) = \mathcal{M}_{\mathcal{P}_0, 2, r_0(1-t)^2}(B_{r_0(1-t)^2})$. Applying Proposition 3.9 to Γ_t provides

$$|\Gamma_t|_{TV} \leq \lambda_1 \sqrt{\frac{2}{r_0(1-t)^2}} r_0 = \frac{\sqrt{2}\lambda_1}{1-t}, \quad (3.41)$$

corresponding with the computations, so that the estimate (3.30) seems sharp.

3.3. Two different regularizations ensuring bi-Lipschitz homeomorphisms. To define models that guarantee bi-Lipschitz homeomorphisms, we propose two different approaches, which both lead to a control on the TV norm of Γ . The first model consists in directly adding the total variation in the formulation. This model can be applied to any type of data, such as clouds of points, measures, etc... The second model is tailored to probability measures and is easier to solve numerically.

3.3.1. *TV regularization on the dual space.* Let us define the measure space:

$$\mathcal{B}_{0,TV} := \{ \Gamma \in \mathcal{M}_0^d(\Omega) \mid |\Gamma|_{TV} < +\infty \}. \quad (3.42)$$

Similarly, we define \mathcal{B}_{TV} without the total mass constraint. Here, the total variation norm $|\cdot|_{TV}$ is defined through the duality with $(\mathcal{C}^0(\Omega), \|\cdot\|_\infty)$ the space of bounded continuous function on $\Omega \subset \mathbb{R}^d$, a compact bounded domain. In this section, we consider the energy:

$$E^\beta : \Gamma \in L^2([0, 1], \mathcal{B}_{TV}) \mapsto \mathcal{L}(\psi_\Gamma(1) \# \mu_0, \nu) + \beta \int_0^1 |\Gamma_t|_{TV}^2 dt, \quad (3.43)$$

where $\psi_\Gamma(1)$ is the flow at time 1 defined by $\psi_\Gamma(0) = \text{Id}$ and $\partial_t \psi_{\gamma_t} = v_t \circ \psi_{\gamma_t}$ with $v_t = K \star \gamma_t$. The goal of this subsection is to prove, provided that the loss \mathcal{L} is lower semi-continuous for the weak-* topology, the following existence result:

Theorem 3.12. *The matching problem*

$$\inf_{\Gamma \in L^2([0, 1], \mathcal{B}_{TV})} E^\beta(v) = \mathcal{L}(\psi_\Gamma(1) \# \mu_0, \nu) + \beta \int_0^1 |\Gamma_t|_{TV}^2 dt \quad (3.44)$$

admits a solution in $L^2([0, 1], \mathcal{B}_{TV})$. Moreover, the matching solution $\psi_\Gamma(1)$ is bi-Lipschitz homeomorphism.

First, let us state a few facts:

Proposition 3.13. *Bounded balls in \mathcal{B}_{TV} are compact for the weak-* topology.*

Proof. Since $|\cdot|_{TV} \leq \frac{1}{\beta} \|\cdot\|_{\mathcal{B}_{TV}}$, bounded balls in \mathcal{B}_{TV} are bounded for the total variation norm, so that they are compact for the weak-* topology by the Banach-Alaoglu theorem. \square

Next, the space $L^2([0, 1], (\mathcal{M}_0^d(\Omega), |\cdot|_{TV}))$ is defined through the duality with $L^2([0, 1], (\mathcal{C}^0(\Omega), \|\cdot\|_\infty))$. We still refer to the induced topology as the weak-* topology, and we get the result:

Proposition 3.14. *Bounded balls in $L^2([0, 1], (\mathcal{M}^d(\Omega), \|\cdot\|_{\mathcal{B}_{TV}}))$ are compact for the weak-* topology.*

Proof. Same proof as above. \square

Proposition 3.15. *The map $\mu \in \mathcal{M}(\Omega) \mapsto \langle \mu, K \star \mu \rangle$ is continuous for the weak-* topology.*

Proof. If Ω is compact, then $K \star \mu$ is bounded by $|\mu|_{TV} \text{diam}(\Omega)$ for any μ . The identity $\langle \mu, K \star \mu \rangle - \langle \nu, K \star \nu \rangle = \langle \mu - \nu, K \star (\mu - \nu) \rangle + 2\langle \nu, K \star (\mu - \nu) \rangle$, combined with the Cauchy-Schwarz inequality, proves the result. \square

Let us consider the flow application

$$\psi(1) : \begin{array}{ccc} L^2([0, 1], \mathcal{B}_{TV}) & \rightarrow & \mathcal{C}^0(\Omega, \mathbb{R}^d) \\ \Gamma & \mapsto & \psi_\Gamma(1). \end{array} \quad (3.45)$$

Proposition 3.16. *The application $\psi(1)$ is continuous from $L^2([0, 1], \mathcal{B}_{TV})$ to $(\mathcal{C}^0(\Omega, \mathbb{R}^d), \|\cdot\|_\infty)$.*

Proof. Let $(\Gamma_k) \in L^2([0, 1], \mathcal{B}_{TV})^{\mathbb{N}}$ such that $\Gamma_k \xrightarrow[k \rightarrow \infty]{} \Gamma$ for the \mathcal{B}_{TV} topology. Denote ψ_s^k and ψ_s the associated flows, corresponding to velocity fields v_s^k and v_s . These flows exist by Proposition 3.8, as $|\cdot|_{TV} \leq \frac{1}{\beta} \|\cdot\|_{\mathcal{B}_{TV}}$. Let $x \in \Omega$. Then:

$$|\psi_t^k(x) - \psi_t(x)| = \left| \int_0^t v_s^k(\psi_s^k(x)) - v_s(\psi_s(x)) ds \right| \quad (3.46)$$

$$\leq \int_0^t |v_s^k(\psi_s^k(x)) - v_s^k(\psi_s(x))| ds + \left| \int_0^t v_s^k(\psi_s(x)) - v_s(\psi_s(x)) ds \right|. \quad (3.47)$$

For the first term:

$$\int_0^t |v_s^k(\psi_s^k(x)) - v_s^k(\psi_s(x))| ds \leq \int_0^t \text{Lip}(v_s^k) |\psi_s^k(x) - \psi_s(x)| ds \quad (3.48)$$

$$\leq \frac{1}{\beta} \|\psi_s^k - \psi_s\|_\infty \int_0^t \|\Gamma_s^k\|_{\mathcal{B}_{TV}} ds. \quad (3.49)$$

For the second term, we define

$$m_x(\Gamma) := \int_0^t K \star \gamma_s(\psi_s(x)) ds. \quad (3.50)$$

which is a continuous linear application from \mathcal{B}_{TV} to \mathbb{R}^d . From estimates (3.27), the family $\{m_x; x \in \Omega\}$ is equicontinuous. Using the Arzelà-Ascoli theorem, this ensures uniform convergence, in the sense

$$\lim_{k \rightarrow \infty} \lambda_k := \sup_{x \in \Omega} |m_x(v) - m_x(v^k)| = 0. \quad (3.51)$$

We get

$$\|\psi_s^k - \psi_s\|_\infty \leq \lambda_k + \frac{1}{\beta} \|\psi_s^k - \psi_s\|_\infty \int_0^t \|\Gamma_s^k\|_{\mathcal{B}_K^\lambda} ds. \quad (3.52)$$

Thus, an application of Grönwall's Lemma ends the proof. \square

Then, Theorem 3.12 is a direct corollary of this proposition and Proposition 3.14.

Representation theorem. The regularization with the TV norm promotes sparse solutions to the variational problem. In this case, a Representer Theorem similar to 2.12 can be established: certain solutions to the corresponding minimization problem can be expressed as linear combinations of extremal points of the unit ball (see [Boy+18; BC19]). First, let us consider a soft interpolation problem:

Proposition 3.17. *Let Ω be a bounded subset of \mathbb{R}^d and $X = (x_1, \dots, x_n) \in (\Omega)^N$ be a tuple of n distinct points in \mathbb{R}^d . Given values $Y = (y_1, \dots, y_n)$, consider the problem*

$$\inf_{(\Gamma, p) \in \mathcal{M}_0 \times \mathcal{P}_0} \sum_{i=1}^n |K \star \Gamma(x_i) + p(x_i) - y_i|^2 + |\gamma|_{TV} \quad (3.53)$$

admits a sparse solution, i.e. Γ can be written as

$$\Gamma = \sum_{j=1}^m \alpha_j \delta_{z_j}, \quad (3.54)$$

where $m \leq Nd + d$, $z_j \in \Omega$ for all j and $\sum_{j=1}^m \alpha_j = 0$.

Proof of Proposition 3.17. Consider an extension, denoted \tilde{K} of the map $K\mathcal{M}_{0,TV} \rightarrow \mathcal{C}^0(X)$ defined by $\mu \mapsto K \star \mu$ to the space of measures \mathcal{P}_{TV} . Then, we apply [BC19, Theorem 3.3]. The function $z, p \in \Omega^n \times \mathcal{P}_0 \mapsto \sum_i |z_i + p(x_i) - y_i|^2$ is coercive, convex, and l.s.c for the usual topology in the Euclidean space. Moreover, the operator $\mathcal{A} : \gamma, p \in \mathcal{M} \times (\mathbb{R}^d)^n \times \mathcal{P}_0 \mapsto (\tilde{K} \star \Gamma(x_i) + p(x_i), 1 \leq i \leq n) \in (\mathbb{R}^d)^n$ is a linear, continuous and surjective. In addition, we add the constraint $\mu(\Omega) = 0 \in \mathbb{R}^d$. By [BC19, pp. 4.1, 3], there exists a solution which is a linear combination of $nd + d$ Dirac masses (the extremal points of the TV ball of $\mathcal{M}(\Omega, \mathbb{R}^d)$). \square

For more details on sparse representation theorems, see [BC19, pp. 4.1, 3] and [Boy+18].

Remark 3.18 (About the proof). To directly apply the result from [BC19], the extremal points of the ball of $\mathcal{M}_0(\Omega, \mathbb{R}^d)$ need to be characterized. These extremal points are sums of at most $d + 1$ Dirac masses, as shown in the Appendix. However, this approach would give an $Nd(d + 1)$ upper bound on the number of Dirac masses. Thus, a more effective approach is to incorporate the total volume constraint directly into the functional.

Remark 3.19. The above result does not provide information about the localization of the Dirac masses δ_{z_i} . Recall that the RKHS representation result guarantees N Dirac masses for an optimal solution, which stands in sharp contrast with the result we have $(N + 1)d$. However, this bound is only an upper estimate, and it remains unclear whether it is tight.

3.3.2. L^2 regularization on the momentum. In this section, we combine the regularity properties of the previous model with the simpler representation of the model from Section 3.1. Specifically, let $\mu \in \mathcal{P}(\Omega)$, where Ω is a bounded domain. If $\gamma \in L^2(\mu, \mathbb{R}^d)$, then the measure $\Gamma := \gamma\mu$ is finite.

Lemma 3.20. *Let $\mu \in \mathcal{P}(\Omega)$, where Ω is a bounded domain, and let $\gamma \in L^2(\mu, \mathbb{R}^d)$. Define $\Gamma := \gamma\mu$. Then, the quantity $\langle \Gamma, K \star \Gamma \rangle$ is well-defined. Furthermore, the function $K \star \Gamma$ is also in $L^2(\mu, \mathbb{R}^d)$, with the inequalities:*

$$\|K \star \Gamma\|_{L^2(\mu)} \leq \text{diam}(\Omega) \|\gamma\|_{L^2(\mu)}. \quad (3.55)$$

$$|\Gamma|_{TV} \leq \|\gamma\|_{L^2(\mu)}. \quad (3.56)$$

Proof. Since Ω is a bounded domain, the Energy-Distance kernel is bounded by $\text{diam}(\Omega)$ on the support of μ . This proves that $\|K \star \Gamma\|_{L^2(\mu)} \leq \text{diam}(\Omega) \|\gamma\|_{L^2(\mu)}$ and $\langle \Gamma, K \star \Gamma \rangle \leq \text{diam}(\Omega) \|\gamma\|_{L^2(\mu)}^2$. For the second part, we apply the Cauchy-Schwarz inequality:

$$|\Gamma|_{TV} = \int |\gamma| d\mu \leq \left(\int |\gamma|^2 d\mu \right)^{1/2} = \|\gamma\|_{L^2(\mu)}. \quad (3.57)$$

□

Now, if μ is an absolutely continuous measure curve [AGS05, Sec I.1] in $AC^2([0, 1], \mathcal{P}(\Omega))$, we define the μ -dependant function space:

$$L^2([0, 1], L^2(\mu_t)) = \{\gamma : t \mapsto \gamma_t \in L^2(\mu_t) \mid \int_0^1 \|\gamma_t\|_{L^2(\mu_t)}^2 dt < +\infty\}. \quad (3.58)$$

Define the vector-valued measure curve $\Gamma : t \mapsto \Gamma_t := \gamma_t \mu_t$, and the associated velocity $v_t := K \star \Gamma_t \in L^2(\mu_t)$. Then, the system

$$\begin{cases} \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0, \\ v_t = K \star \Gamma_t, \\ \Gamma_t = \gamma_t \mu_t, \\ \gamma_t \in L^2(\mu_t), \end{cases} \quad (3.59)$$

where the first equation holds in the sense of distributions [AGS05, Chap. 8], is well defined, and admits solutions $(\mu_t, \gamma_t) \in AC^2([0, 1], \mathcal{P}(\Omega)) \times L^2([0, 1], L^2(\mu_t))$. Now, we define the matching problem:

$$\inf_{(\mu, \gamma) \text{ solves (3.59)}} E^\lambda(\mu, \gamma) = \mathcal{L}(\mu_1, \nu) + \int_0^1 \langle \Gamma_t, K \star \Gamma_t \rangle dt + \int_0^1 \|\gamma_t\|_{L^2(\mu_t)}^2 dt, \quad (3.60)$$

where $(\mu, \gamma) \in AC^2([0, 1], \mathcal{P}(\Omega)) \times L^2([0, 1], L^2(\mu_t))$ is a solution to the system 3.59, μ_1 is the measure μ_t at time 1, and K is the Energy-Distance kernel. Here, ν is a reference target measure, and \mathcal{L} is any weakly-continuous loss between μ_1 and ν (see Example 1.10). Let us start with the following lemma, which allows to control the growth of the support of μ_t over time:

Lemma 3.21 (Support growth bound). *Let (μ, γ) be a solution to the system (3.59). Let us suppose that $\text{supp}(\mu_0)$ is bounded, and that:*

$$\int_0^1 \|\gamma_t\|_{L^2(\mu_t)}^2 dt \leq C, \quad (3.61)$$

for some positive constant C . Then, for all $t \in [0, 1]$:

$$\text{diam}(\text{supp } \mu_t) \leq \text{diam}(\text{supp } \mu_0) \exp \left(\left(\int_0^t \|\gamma\|_{L^2(\mu_t)}^2 dt \right)^{1/2} \right). \quad (3.62)$$

Proof. For almost every $t \in (0, 1)$, for μ_t -almost any x , we have:

$$|v_t(x)| \leq \text{diam}(\text{supp } \mu_t) \|\gamma\|_{L^2(\mu_t)}, \quad (3.63)$$

so that

$$\frac{d}{dt} \text{diam}(\text{supp } \mu_t) \leq \text{diam}(\text{supp } \mu_t) \|\gamma\|_{L^2(\mu_t)}. \quad (3.64)$$

Using Grönwall's Lemma and the Cauchy-Schwarz inequality, we obtain

$$\text{diam}(\text{supp } \mu_t) \leq \text{diam}(\text{supp } \mu_0) \exp \left(\int_0^t \|p\|_{L^2(\mu_t)} dt \right) \quad (3.65)$$

$$\leq \text{diam}(\text{supp } \mu_0) \exp \left(\left(\int_0^t \|p\|_{L^2(\mu_t)}^2 dt \right)^{1/2} \right). \quad (3.66)$$

□

The following result completely characterizes solutions to the system (3.59) in the case of atomic initial measures with finite E^λ energy, and provides a straightforward parametrization of the flow.

Proposition 3.22 (*L^2 parametrization*). *Let $X = (x_1, \dots, x_N) \in (\mathbb{R}^d)^n$, $p \in L^2([0, 1], (\mathbb{R}^d)^n)$. Then, there exists an absolutely continuous measure curve μ written as*

$$\mu_t = \sum_{i=1}^n \frac{1}{n} \delta_{x_i(t)}, \quad (3.67)$$

such that, defining $\gamma_t \in L^2(\mu_t) : x_i(t) \mapsto p_t^i$, (μ, γ) is a solution to the system (3.59) and such that the $x_i(t)$ are distinct at all times. Furthermore, provided initial distinct points $X = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, the application

$$p \in L^2([0, 1], (\mathbb{R}^d)^n) \mapsto \mu_t \in \mathcal{P}(\mathbb{R}^d) \quad (3.68)$$

is weakly continuous for all $t \in [0, 1]$, and the support of μ_t is uniformly bounded in time. Conversely, if μ_t is an absolutely continuous curve written as in (3.67), where the points are distinct at all times, with support contained in a uniform compact set Ω , then there exists $p \in L^2([0, 1], (\mathbb{R}^d)^n)$ such that $\mu_t = \mu_t(p)$ as constructed in the previous statement.

Proof. Let μ be a probability measure with finite support (x_1, \dots, x_n) and $p : x_i \mapsto p_i \in L^2(\mu)$. Define the velocity fields $v = K \star (p\mu)$. Then, v is Lipschitz, with a Lipschitz constant bounded by $\|p\|_{L^2(\mu)}$, which is independent from the support of μ . This ensures, by the Cauchy-Lipschitz Theorem, that the Lagrangian equations

$$\begin{cases} \dot{x}_i(t) = v_t(x_i(t)), \\ x_i(0) = x_i, \end{cases} \quad \text{for } 1 \leq i \leq n, \quad (3.69)$$

have a unique solution $(t \in [0, 1] \mapsto x_i(t), 1 \leq i \leq n)$. Denote

$$\mu_t := \sum_{i=1}^n \frac{1}{n} \delta_{x_i(t)}. \quad (3.70)$$

From Lemma 3.21, the support of μ_t is uniformly bounded in time. Using Proposition 3.8, there exists a bi-Lipschitz homeomorphism curve $t \in [0, 1] \mapsto \psi_t$ such that ψ_t is the unique solution of the flow equation

$$\begin{cases} \partial_t \psi_t = v_t \circ \psi_t \\ \psi_0 = \text{Id}. \end{cases} \quad (3.71)$$

Applying estimate (3.27) applied to $\psi_t^{-1}(x)$ and $\psi_t^{-1}(y)$ provides, for all $x \in \mathbb{R}^d$:

$$|\psi_t(x) - \psi_t(y)| \geq |x - y| \exp \left(- \int_0^1 \|p\|_{L^2(\mu_t)} dt \right) \geq |x - y| \exp \left(- \int_0^1 \|p\|_{L^2(\mu_t)}^2 dt \right). \quad (3.72)$$

Furthermore, since the homeomorphism ψ_t satisfies, at all times:

$$x_i(t) = \psi_t(x_i), \quad (3.73)$$

the points $(x_i(t), 1 \leq i \leq n)$ are distinct at all times, with a uniform bound from below of the minimal distance between points. For the last statement, remark that the ODE system (3.69) implies that, for all $1 \leq i \leq n$:

$$x_i(t) = x_i + \int_0^1 \sum_{j=1}^n p_{i,s} K(x_i(s) - x_j(s)) ds, \quad (3.74)$$

which is a weakly continuous application in p for the $L^2([0, 1])$ topology.

For the converse part of the proposition, let $\mu_t \in AC^2([0, 1], \mathcal{P}(\Omega))$. Then, from [AGS05, Theorem 8.3.1], there exists a Borel velocity field $v_t \in L^2(\mu_t, \mathbb{R}^d)$ such that the continuity equation

$$\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \quad (3.75)$$

holds in the sense of distribution. This amounts to the existence, for almost all times $t \in [0, 1]$, of a velocity $v_{t,i} \in \mathbb{R}^d$ such that

$$\dot{x}_i(t) = v_{t,i}. \quad (3.76)$$

Let us denote $X(t) = (x_1(t), \dots, x_n(t))$ and $K_{X(t)}$ the matrix with general entries $-|x_i(t) - x_j(t)|$. By Proposition 2.13, for almost all t , there exists a unique couple $(p_t, \alpha_t) \in (\mathbb{R}^d)^n \times \mathbb{R}^d$ such that $\sum_i p_{t,i} = 0_d$ and such that, for all $1 \leq i \leq n$:

$$v_{t,i} = (K_{X(t)} p_t)_i + \alpha_t. \quad (3.77)$$

Define

$$\Gamma_t(p) = \sum_{i=1}^n p_{t,i} \delta_{x_i(t)}. \quad (3.78)$$

Since the trajectories $x_i(t)$ are distinct at all times, there exists $r, \sigma > 0$ such that $\Gamma_t \in \mathcal{M}_{\mathcal{P}_{m-1}, n, r}(B_{\sigma r})$ at all times. This proves, using estimate (3.39) from Proposition 3.9 with the same $\lambda_\sigma > 0$, that:

$$|p_t|_2^2 \leq \frac{\lambda_\sigma^2}{r} |K_{X(t)} p_t|_2^2. \quad (3.79)$$

As $\int_0^1 |K_{X(t)} p_t|^2 dt = \int_0^1 |v_t|^2 d\mu_t < +\infty$, this proves that $t \mapsto p_t$ is in $L^2([0, 1], (\mathbb{R}^d)^n)$, concluding the proof. \square

This property allows for a reparametrization of the matching problem (3.60) by the set $L^2([0, 1], (\mathbb{R}^d)^n)$ in the case of n landmarks. We consider the following problem:

$$\inf_{p \in L^2([0, 1], (\mathbb{R}^d)^n)} \tilde{E}^\lambda(p) = \mathcal{L}(\mu_1, \nu) + \int_0^1 \langle \Gamma_t, K \star \Gamma_t \rangle dt + \int_0^1 |p_t|^2 dt, \quad (3.80)$$

where

$$\mu_0 = \sum_{i=1}^n \frac{1}{n} \delta_{x_i}, \quad (3.81)$$

the x_i are all distinct, μ_t is the (unique) absolutely continuous curve provided by Proposition 3.22,

$$\Gamma_t = \sum_{i=1}^n p_{t,i} \delta_{x_i(t)}, \quad (3.82)$$

and \mathcal{L} is a general loss. We are able to prove the following result:

Theorem 3.23. *Suppose that \mathcal{L} is a lower semicontinuous loss. Then, there exist solutions to the variational problem (3.80).*

Proof. The level sets $\left\{ p \in L^2([0, 1], (\mathbb{R}^d)^n) \mid \int_0^1 |p_t|^2 dt \leq C \right\}$ are weakly compact, and $p \mapsto \tilde{E}^\lambda(p)$ is weakly l.s.c. by Proposition 3.22 and the hypothesis on \mathcal{L} . This proves the result. \square

We can prove the existence of minimizers to Problem (3.60) for very general initial measures:

Theorem 3.24. *Let \mathcal{L} be a lower semicontinuous loss. Then, there exists a solution to the variational problem (3.60).*

Proof. First, note that if $\gamma \in L^2([0, 1], L^2(\mu_t))$, then the moment measure $\Gamma_t := \gamma_t \mu_t$ is well defined for almost all times, and satisfies:

$$E^\lambda(\mu, \gamma) = \tilde{E}^\lambda(\mu, \Gamma) := \mathcal{L}(\mu_1, \nu) + \int_0^1 \langle \Gamma_t, K \star \Gamma_t \rangle dt + \int_0^1 \int \left| \frac{d\Gamma_t}{d\mu_t} \right|^2 d\mu_t dt. \quad (3.83)$$

Let (μ^k, γ^k) be a minimizing sequence to problem (3.60). Note that there exists a uniform constant $C > 0$ such that, for all $k \geq 0$:

$$\int_0^1 |\gamma_t^k|_{L^2(\mu_t^k)}^2 dt \leq C. \quad (3.84)$$

By Lemma 3.21, the support of μ_t^k is uniformly bounded in time and in k . This proves that $\mu^k \in AC^2([0, 1], \mathcal{P}(\Omega))$ for some compact set Ω , so that there exists an accumulation point $\mu \in AC^2([0, 1], \mathcal{P}(\Omega))$. Now, let us define, for any $k \geq 0$, the momentum measure:

$$\Gamma_t^k := \gamma_t^k \mu_t^k. \quad (3.85)$$

Since, for any k :

$$\int_0^1 |\Gamma_t^k|_{TV} dt \leq C, \quad (3.86)$$

there exists a subsequence such that Γ^k weakly converges to an accumulation point $\Gamma \in \mathcal{M}([0, 1]; \mathbb{R}^d)$, in the sense that for any test function $f \in \mathcal{C}^0([0, 1] \times \Omega)$,

$$\int_0^1 \int_\Omega f_t d\Gamma_t^k \xrightarrow{k \rightarrow +\infty} \int_0^1 \int_\Omega f_t d\Gamma_t. \quad (3.87)$$

Now, define the following velocity field:

$$v_t^k := K \star \Gamma_t^k, \quad (3.88)$$

which is defined for almost any t on Ω . The Energy-Distance kernel K is continuous, so that if $g \in L^2([0, 1], L^2(\mu_t))$ is a test function:

$$\int_0^1 \int g_t v_t^k d\mu_t dt = \int_0^1 \int g_t(x) \left(\int K(x-y) d\Gamma_t^k(y) \right) d\mu_t(x) dt \quad (3.89)$$

$$\xrightarrow{k \rightarrow \infty} \int_0^1 \int g_t(x) \left(\int K(x-y) d\Gamma_t(y) \right) d\mu_t(x) dt \quad (3.90)$$

$$= \int_0^1 \int g_t v_t d\mu_t dt. \quad (3.91)$$

This proves that, for almost all times t , the sequence v_t^k converges to $v_t := K \star \Gamma_t$ in $L^2(\mu_t)$. Moreover, the continuity equation is weakly verified at the limit. Indeed, for any $\varphi \in \mathcal{C}^\infty([0, 1] \times \Omega)$:

$$0 = \int_0^1 (\partial_t \varphi + v_t^k \cdot \nabla \varphi_t) d\mu_t^n dt \xrightarrow{k} \int_0^1 (\partial_t \varphi + v_t \cdot \nabla \varphi_t) d\mu_t^n dt. \quad (3.92)$$

Now, since the Lagrangian

$$(\mu, \Gamma) \mapsto \int_0^1 \langle \Gamma_t, K \star \Gamma_t \rangle dt + \int_0^1 \int \left| \frac{d\Gamma_t}{d\mu_t} \right|^2 d\mu_t \quad (3.93)$$

is convex, it is lower semi-continuous, so that by hypothesis on \mathcal{L} the functional \tilde{E}^λ is lower semi-continuous. This proves:

$$\tilde{E}^\lambda(\mu, \Gamma) \leq \inf_{(\mu, \gamma) \text{ solves (3.59)}} E^\lambda(\mu, \gamma). \quad (3.94)$$

Finally, let us define

$$\gamma_t = \frac{d\Gamma_t}{d\mu_t}, \quad (3.95)$$

the Radon-Nikodym derivative of Γ_t by μ_t . Then $\gamma \in L^2([0, 1], L^2(\mu_t))$, and $\tilde{E}^\lambda(\mu, \Gamma) = E^\lambda(\mu, \gamma)$, ending the proof. \square

This solution has additional structure: at all times, the measure μ_t is the pushforward of μ_0 through a bi-Lipschitz homeomorphism.

Proposition 3.25. *Let (μ, γ) be a solution to the system (3.59) with finite E^λ energy. Then, there exists a bi-Lipschitz homeomorphism curve ψ_t such that ψ_t is the unique solution to the flow equation*

$$\begin{cases} \partial_t \psi_t = v_t \circ \psi_t \\ \psi_0 = \text{Id}. \end{cases} \quad (3.96)$$

Furthermore, at all times $t \in [0, 1]$, we have: $\mu_t = \psi_{t\#}\mu_0$.

Proof. This is a direct consequence of Proposition 3.8 and of the inequality $|\Gamma|_{TV} \leq \|\gamma\|_{L^2(\mu)}$. \square

4. SLICED ENERGY DISTANCE.

In this section, we show how various interaction functionals involving the Energy-Distance can be statistically estimated in $O(n \log n)$ time, where n is the number of points. This almost-linear complexity is a significant improvement over the $O(n^2)$ time of more naive methods, making it suitable for large-scale applications. The approach follows the same principle as the Sliced Wasserstein distance [Jul+11; Bon+15; Nad21]: the original metric is replaced with a sliced approximation over 1D distributions. Applications of this principle to the Energy-Distance kernel build on ideas from [Her+24], where an efficient method for computing the gradient of the Energy-Distance loss between two sets of n Dirac masses was established. We extend these results, computing general convolutions of the kernel matrix with 1D atomic distributions. Specifically, the convolution of the Energy-Distance kernel with any 1D measure composed of n Dirac masses is computable in quasi-linear time using a sorting procedure, leveraging particularly efficient computations when the 1D points are sorted. We prove that the convolution of atomic measures in higher dimensions can be efficiently estimated using a tractable number of projections of the measures over 1D lines, with all the computations being easily parallelizable.

4.1. Fast estimators. First, we establish that, in the case of sorted 1D data, the convolution can be computed in linear time.

Proposition 4.1 (Linear time complexity for sorted data). *Let $x_0 \leq x_1 \leq \dots \leq x_N \in \mathbb{R}$ and $\gamma_0, \dots, \gamma_N \in \mathbb{R}$. Then, the quantity*

$$K \star \gamma := \left(\sum_{j=0}^n \gamma_j |x_i - x_j|; i \in [1, n] \right) \in \mathbb{R}^n \quad (4.1)$$

can be computed in $O(n)$ time complexity.

Proof. Let $i \in [0, n]$. We rewrite the sum as follows:

$$\sum_{j=0}^n \gamma_j |x_i - x_j| = \sum_{j=0}^{i-1} \gamma_j (x_i - x_j) + \sum_{j=i+1}^n \gamma_j (x_j - x_i) \quad (4.2)$$

$$= x_i \left(\sum_{j=0}^{i-1} \gamma_j - \sum_{j=i+1}^n \gamma_j \right) - \left(\sum_{j=0}^{i-1} \gamma_j x_j - \sum_{j=i+1}^n \gamma_j x_j \right) \quad (4.3)$$

$$\sum_{j=0}^n \gamma_j |x_i - x_j| = a_i x_i - b_i, \quad (4.4)$$

where a_i and b_i are scalar terms that satisfy the recurrence relations:

$$\begin{cases} a_0 &= -\sum_{j=1}^n \gamma_j \\ a_{i+1} &= a_i + \gamma_i + \gamma_{i+1} \end{cases} \quad \text{and} \quad \begin{cases} b_0 &= -\sum_{j=1}^n \gamma_j x_j \\ b_{i+1} &= b_i + \gamma_i x_i + \gamma_{i+1} x_{i+1} \end{cases}. \quad (4.5)$$

Computing $((a_i, b_i); i \in [0, n])$ requires $O(n)$ operations, which concludes the proof. \square

The convolution of the Energy-Distance kernel with any finite measure in \mathbb{R}^d is equal to its sliced version.

Proposition 4.2 (Sliced Energy-Distance). *Let $\gamma \in \mathcal{M}(\mathbb{R}^d)$ be a finite signed measure. For $\theta \in \mathcal{S}_{d-1}$, define the 1D projection on the line $\mathbb{R}\theta$ via $\pi_\theta(x) := \langle x, \theta \rangle \theta$. Then, the convolution with the Energy-Distance kernel is sliced, meaning that there exists a positive constant c_d such that, for all $x \in \mathbb{R}^d$:*

$$K \star \gamma(x) = c_d \int_{\mathcal{S}_{d-1}} K \star (\pi_{\theta\#}\gamma)(x) d\sigma(\theta), \quad (4.6)$$

and

$$\langle \gamma, K \star \gamma \rangle = c_d \int_{\mathcal{S}_{d-1}} \langle \pi_{\theta\#}\gamma, K \star (\pi_{\theta\#}\gamma) \rangle d\sigma(\theta), \quad (4.7)$$

where σ is the standard uniform probability measure on the sphere \mathcal{S}_{d-1} , and

$$c_d = \frac{\sqrt{\pi} \Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}. \quad (4.8)$$

Proof. The proof follows directly from adapting the result of [Her+24, Theorem 1] to a general finite measure γ . \square

We can obtain a bound on the error resulting from approximating the integral above through a Monte-Carlo method.

Proposition 4.3 (Statistical error). *Let $\gamma \in \mathcal{M}(\mathbb{R}^d)$ be a finite signed measure with support in a ball of radius R , and let $x \in \mathbb{R}^d$. Let $\theta_1, \dots, \theta_P$ be P i.i.d. random variables uniformly distributed on the sphere \mathcal{S}_{d-1} . Then, the following estimators*

$$\widehat{K \star \gamma}(x) := \frac{c_d}{P} \sum_{p=1}^P K \star (\pi_{\theta_p\#}\gamma)(x), \quad (4.9)$$

and

$$\langle \gamma, \widehat{K \star \gamma} \rangle := \frac{c_d}{P} \sum_{p=1}^P \langle \pi_{\theta_p\#}\gamma, K \star (\pi_{\theta_p\#}\gamma) \rangle \quad (4.10)$$

satisfy

$$\mathbb{E}_{\theta_1, \dots, \theta_P \sim \mathcal{U}(\mathcal{S}_{d-1})} (\|\widehat{K \star \gamma}(x) - K \star \gamma(x)\|) \leq AR|\gamma| \sqrt{\frac{d}{P}}, \quad (4.11)$$

and

$$\mathbb{E}_{\theta_1, \dots, \theta_P \sim \mathcal{U}(\mathcal{S}_{d-1})} (|\langle \gamma, \widehat{K \star \gamma} \rangle - \langle \gamma, K \star \gamma \rangle|) \leq AR|\gamma|^2 \sqrt{\frac{d}{P}}. \quad (4.12)$$

where $|\gamma|$ denotes the total variation of the measure γ , for some constant $A > 0$.

Proof. In this proof we write \mathbb{E} and \mathbb{P} instead of $\mathbb{E}_{\theta_1, \dots, \theta_P \sim \mathcal{U}(\mathcal{S}_{d-1})}$ and $\mathbb{P}_{\theta_1, \dots, \theta_P \sim \mathcal{U}(\mathcal{S}_{d-1})}$. Let $x \in \mathbb{R}^d$. First, remark that if $\theta_1, \dots, \theta_P \sim \mathcal{U}(\mathcal{S}_{d-1})$ then almost surely:

$$\|\widehat{K \star \gamma}(x) - K \star \gamma(x)\| \leq 2c_d R |\gamma|. \quad (4.13)$$

This directly implies

$$\mathbb{E}(\|\widehat{K \star \gamma}(x) - K \star \gamma(x)\|^2) \leq 4(c_d R |\gamma|)^2. \quad (4.14)$$

Now, we can apply Bernstein's inequality to the zero-mean variables $\widehat{K \star \gamma}(x) - K \star \gamma(x)$, to get:

$$\mathbb{P}(\|\widehat{K \star \gamma}(x) - K \star \gamma(x)\| \geq t) \leq 2 \exp\left(\frac{-\frac{1}{2}P^2 t^2}{4P(c_d R |\gamma|)^2 + \frac{1}{3}|\gamma| c_d R P t}\right) = 2 \exp\left(-\frac{3P}{4c_d R |\gamma|} \frac{t^2}{6c_d R |\gamma| + t}\right). \quad (4.15)$$

For any $\alpha > 0$, integrating the inequality over t : Then:

$$\int_0^\alpha \mathbb{P}(\|\widehat{K \star \gamma}(x) - K \star \gamma(x)\| \geq t) dt \leq \int_0^\alpha 2 \exp\left(-\frac{3P}{4c_d R |\gamma|} \frac{t^2}{6c_d R |\gamma| + \alpha}\right) dt \quad (4.16)$$

$$\leq 2\sqrt{\frac{\pi c_d R |\gamma| (6c_d R |\gamma| + \alpha)}{3P}}. \quad (4.17)$$

Furthermore, we can get the bound:

$$\int_{\alpha}^{\infty} \mathbb{P}(\|K \star \widehat{\gamma}(x) - K \star \gamma(x)\| \geq t) dt \leq \int_{\alpha}^{\infty} 2 \exp\left(-\frac{3P}{4c_d R |\gamma|} \frac{t^2}{6c_d R |\gamma| + t}\right) dt \quad (4.18)$$

$$\leq \int_{\alpha}^{\infty} 2 \exp\left(-\frac{3Pt}{4c_d R |\gamma|} \left(1 - \frac{6c_d R |\gamma|}{t}\right)\right) dt \quad (4.19)$$

$$= 2 \exp\left(\frac{9P}{2}\right) \frac{4c_d R |\gamma|}{3P} \exp\left(-\frac{3P\alpha}{4c_d R |\gamma|}\right). \quad (4.20)$$

Choosing $\alpha = 7c_d R |\gamma|$ for example ensures the existence of a constant A such that:

$$\mathbb{E}(\|K \star \widehat{\gamma}(x) - K \star \gamma(x)\|) \leq AR |\gamma| \frac{c_d}{\sqrt{P}}. \quad (4.21)$$

As $\frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})}$ behaves asymptotically as $O(\sqrt{d})$, we get the estimate (4.11). The second estimate (4.12) is obtained by integrating the first estimate over γ . \square

Below is an illustration of the previous results. We consider an atomic measure μ with support on a set of distinct points $X = (x_1, \dots, x_N) \in (\mathbb{R}^d)^N$, with associated moment γ_i in \mathbb{R}^d at each point. We compute the vector $(K \star \gamma \mu(x_i), 1 \leq iN)$, and compare it with its sliced approximations.

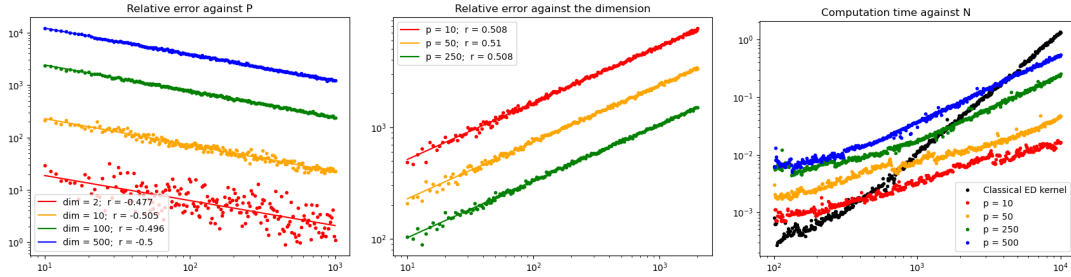


FIGURE 2. In the left figure, we plot the error convergence to zero in $P^{-1/2}$ between the approximated vector and the ground-truth as P grows. In the middle, the error as the dimension increases. In the right-most graph, we plot the computing time against the number of points in dimension 5 on a standard 8-core CPU.

4.2. Use as a loss. Computing the Energy-Distance loss between two points clouds $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ and $(y_1, \dots, y_M) \in (\mathbb{R}^d)^M$ amounts to computing $\langle \rho_{X,Y}, K \star \rho_{X,Y} \rangle$, where

$$\rho_{X,Y} := \sum_{i=1}^N \frac{1}{N} \delta_{x_i} - \sum_{j=1}^M \frac{1}{M} \delta_{y_j}. \quad (4.22)$$

The previous discussion shows that $\langle \rho_{X,Y}, K \star \rho_{X,Y} \rangle$ can be estimated with an $O((N+M) \log(N+M))$ complexity.

5. EXPERIMENTS.

To demonstrate the matching properties of the Energy-Distance kernel, we implement and evaluate the solution to Problem (3.4) across multiple point-cloud datasets. The trajectories are parametrized by their starting points and moment values at all times (see 3.22), and computed using a simple Euler scheme. We have to account for the constraints needed in the case of the Energy-Distance kernel, and parameterize the null-space of the metric explicitly. Let us define the projection matrix

$\Pi_n := \text{Id}_n - \frac{1}{n} \mathbf{1}_n$, where $\mathbf{1}_n$ is the $n \times n$ matrix with 1 entries. If $P_t = (P_{t,l}, 1 \leq l \leq d) \in \mathbb{R}^{n \times d}$, the projection on the constraint set we choose is defined by:

$$\overline{P}_t := (\Pi_n P_{t,l}, 1 \leq l \leq d) \in \mathbb{R}^{n \times d}. \quad (5.1)$$

We detail the computations in the algorithm below (recall that K_X denotes the matrix with general entries $K(x_i, x_j)$ for $X = (x_1, \dots, x_n)$):

Algorithm 1 Regularized trajectories for a kernel K (Euler scheme).

Require: Initial data $X_0 = (x_1, \dots, x_n) \in \mathbb{R}^d$, T_{discr} the number of time discretizations, moment vector $P = (P_t, 0 \leq t < T_{discr}) \in \mathbb{R}^{T_{discr} \times n \times d}$. If K is the Energy-Distance kernel, $(\alpha_t, 0 \leq t < T_{discr}) \in \mathbb{R}^{T_{discr} \times d}$ is a parametrization of translations.

- 1: X_{traj}
- 2: **for** $t = 0, \dots, T_{discr} - 1$ **do**
- 3: Compute the velocity field

$$v_{t,i} \leftarrow \begin{cases} (K_{X_t} \overline{P}_t)_i + \alpha_t & \text{if } K \text{ is the ED kernel,} \\ (K_{X_t} P_t)_i & \text{otherwise.} \end{cases} \quad (5.2)$$

- 4: Compute the positions at time $t + 1$:

$$x_i(t+1) \leftarrow x_i(t) + \frac{1}{T_{discr}} v_t(x_i(t), P_t) \quad (5.3)$$

- 5: **end for**
 - 6: **return** $X(P) = (X_t(P), 0 \leq t \leq T_{discr}) \in \mathbb{R}^{(T_{discr}+1) \times n \times d}$.
-

Note that the computations needed for v_t (5.2) have an $O(n^2)$ time-complexity for general kernels, but $O(n \log n)$ time-complexity for the sliced computations of the Energy-Distance kernel. We propose to solve the constrained optimization problem (3.4) using an Augmented Lagrangian method [BR14] with adaptive penalty update with an inner LBFGS optimization. The algorithm operates with a kernel K as follows:

Algorithm 2 Augmented Lagrangian registration.

Require: Initial data $X_0 = (x_1, \dots, x_n) \in \mathbb{R}^d$, target data $Y_0 = (y_1, \dots, y_m) \in \mathbb{R}^d$, $\mathcal{L}(\cdot, \cdot)$ the loss between point-clouds, T_{discr} the number of time discretizations, k_{max} the number of optimization steps, ε the tolerance to the loss constraint, ρ_{init} a parameter.

- 1: $P \leftarrow 0 \in \mathbb{R}^{T_{discr} \times n \times d}$.
- 2: $\lambda \leftarrow 0$.
- 3: $\rho \leftarrow \rho_{init}$.
- 4: **for** $k = 1, \dots, k_{max}$ **do**
- 5: Define the augmented Lagrangian

$$\mathcal{L}_{\lambda, \rho}(X(P), P) := \frac{\rho}{2} \mathcal{L}(X_1(P), Y_0)^2 + \lambda \mathcal{L}(X_1(P), Y_0) + \frac{1}{T_{discr}} \sum_{t=1}^{T_{discr}} \langle P_t, K_{X_t(P)} P_t \rangle. \quad (5.4)$$

- 6: Compute the deformation curve $X(P) = (X_t(P), 1 \leq t \leq T_{discr})$ using Algorithm 1. Note that the third term of the augmented Lagrangian can be computed at the same time.
- 7: Update P via an LBFGS optimization step of the augmented Lagrangian $\mathcal{L}_{\lambda, \rho}(X(P), P)$.
- 8: Update λ and ρ :

$$\lambda \leftarrow \lambda + \rho \mathcal{L}(X_1(P), Y_0) \quad \text{and} \quad \rho \leftarrow \begin{cases} 1.2\rho & \text{if } \mathcal{L}(X_1(P), Y_0) > \varepsilon, \\ \rho & \text{otherwise.} \end{cases} \quad (5.5)$$

- 9: **end for**
 - 10: **return** $X(P), P$.
-

When implemented with the sliced approximation of the augmented Lagrangian terms, the final algorithm achieves $O(n \log n)$ time complexity. In the figure below, we apply this algorithm with a tuned σ Gaussian kernel and with the Energy-Distance kernel. The loss we choose is a sliced Energy-Distance loss, with $P = 1000$ projections.

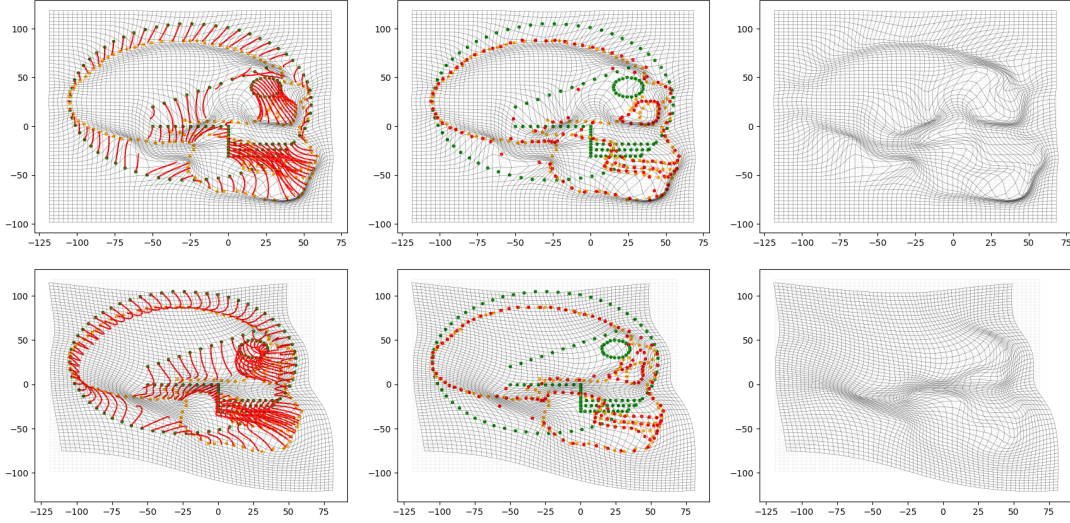


FIGURE 3. Registration with Gaussian kernel at the top, Energy-Distance at the bottom. The final losses are comparable.

In the case of the Energy-Distance kernel, we observe that the resulting deformation exhibits significantly less localized behavior when compared to the Gaussian kernel. Moreover, the algorithm above can be applied with the sliced ED approximation of the particle path. The behavior seems highly stable, matching points with as low as 2 projections, and a convergence to the true deformation as the number of projections grows.

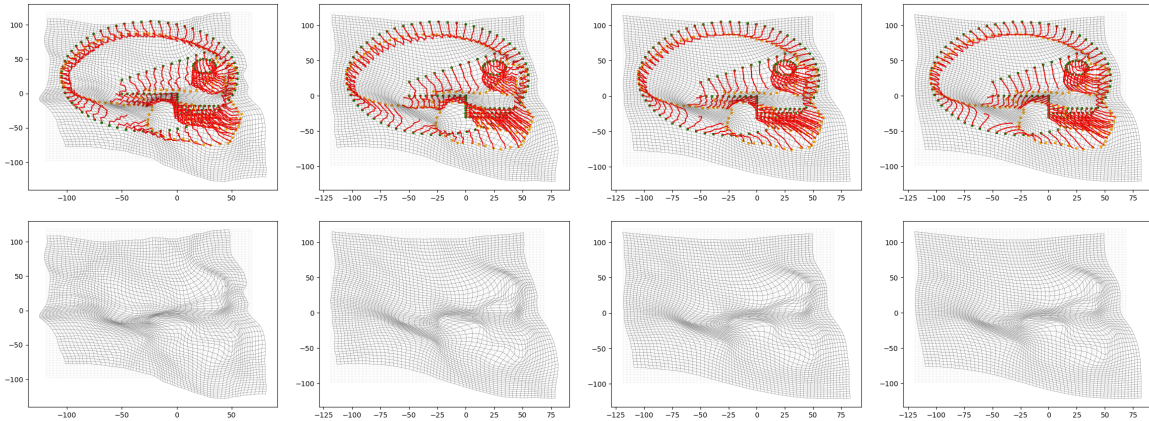


FIGURE 4. Same registration problem using the sliced computation of the kernel. From left to right $P = 2, 10, 50, 500$.

The algorithm successfully recovers accurate matches even for small projection numbers P . Furthermore, the resulting deformations stabilize as P increases, exhibiting convergence toward the true Energy-Distance (ED) registration. Notably, while the deformation paths may become less regular

with fewer projections, the final registration quality remains comparable, suggesting robustness to the choice of P . Moreover, Energy-Distance registration naturally handles translated shapes, as translations lie in the null space of the deformation metric. Our method with sliced computations scales to

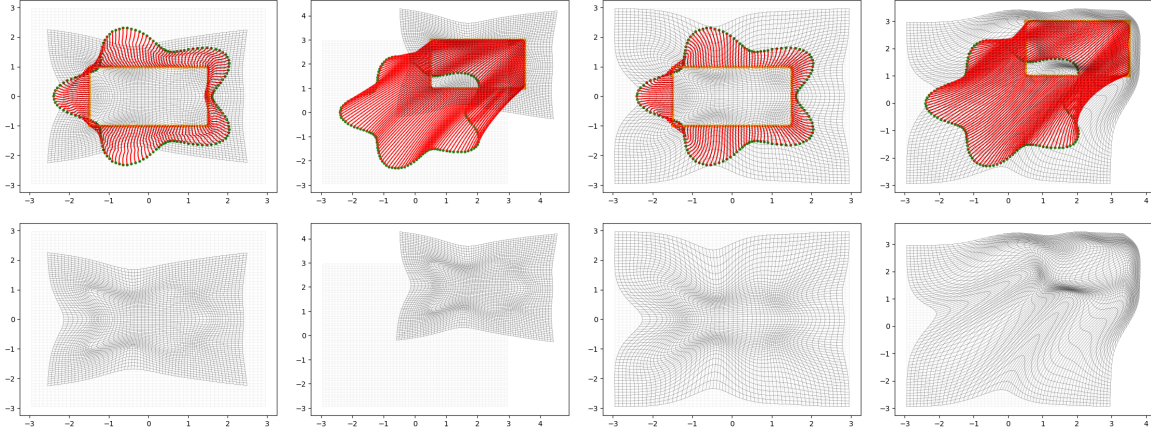


FIGURE 5. Visualization of translated registration results: Energy-Distance kernel (left) and a tuned Gaussian kernel results (right).

a large number of points. The matching accuracy is inherently limited by random sampling, which does not uniformly weight all regions of interest of the shapes. Moreover, the current method does not incorporate additional geometric or contextual information other than the point cloud coordinates. The following experiment can be run in under a minute on a standard recent computer.

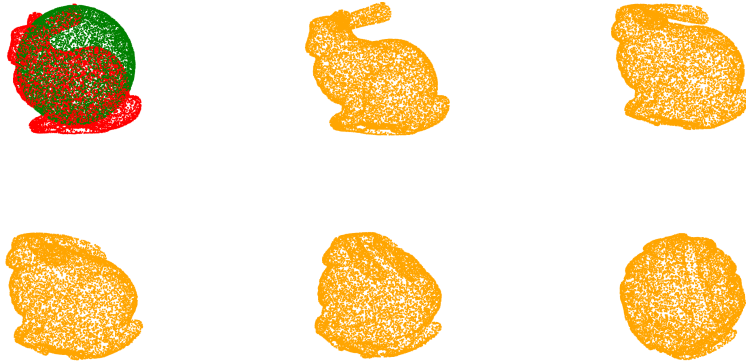


FIGURE 6. Registration with a sliced ED kernel and a sliced ED loss on 10000 points.

ACKNOWLEDGEMENTS

The authors want to warmly thank Stefan Sommer and his co-authors for sharing their work around conditionally positive kernels and stimulating discussions on this topic.

Funding: This work was supported by the Bézout Labex (New Monge Problems), funded by ANR, reference ANR-10-LABX-58.

REFERENCES

- [AF03] R. A. Adams and J. J. F. Fournier. *Sobolev spaces / Robert A. Adams,... and John J.F. Fournier,...* eng. 2nd edition. Pure and applied mathematics. Amsterdam Boston[etc: Elsevier Academic Press, 2003.
- [AC14] L. Ambrosio and G. Crippa. “Continuity equations and ODE flows with non-smooth velocity”. In: *Proceedings of the Royal Society of Edinburgh: Section A Mathematics* 144 (12/2014), pp. 1191–1244.
- [AGS05] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media, 2005.
- [AST04] R. Arcangeli, M. C. Lopez de Silanes, and J. J. Torrens. *Multidimensional Minimizing Splines : Theory and Applications*. eng. Cham: Springer International Publishing., 2004.
- [Aro50] N. Aronszajn. “Theory of Reproducing Kernels”. In: *Transactions of the American Mathematical Society* 68.3 (1950), pp. 337–404.
- [AB09] Y. Auffray and P. Barbillon. *Conditionally positive definite kernels: theoretical contribution, application to interpolation and approximation*. Research Report RR-6835. INRIA, 2009, p. 36.
- [Beg+05] M. F. Beg, M. Miller, A. Trouvé, and L. Younes. “Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms”. In: *International Journal of Computer Vision* 61 (02/2005), pp. 139–157.
- [BR14] D. Bertsekas and W. Rheinboldt. *Constrained Optimization and Lagrange Multiplier Methods*. Computer science and applied mathematics. Academic Press, 2014.
- [Bon+15] N. Bonneel, J. Rabin, G. Peyré, and H. Pfister. “Sliced and Radon Wasserstein Barycenters of Measures”. In: *Journal of Mathematical Imaging and Vision* 1.51 (2015), pp. 22–45.
- [Boo89] F. L. Bookstein. “Principal Warps: Thin-Plate Splines and the Decomposition of Deformations”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 11.6 (06/1989), 567–585.
- [Boy+18] C. Boyer, A. Chambolle, Y. D. Castro, V. Duval, F. D. Gournay, and P. Weiss. *On Representer Theorems and Convex Regularization*. 2018.
- [BC19] K. Bredies and M. Carioni. “Sparsity of solutions for variational inverse problems with finite-dimensional data”. In: *Calculus of Variations and Partial Differential Equations* 59.1 (12/2019).
- [Bre99] H. Brezis. *Analyse fonctionnelle : théorie et applications*. Éditions Dunod, 1999.
- [BV16] M. Bruveris and F.-X. Vialard. *On Completeness of Groups of Diffeomorphisms*. 2016.
- [Bun+24] C. Bunne, G. Schiebinger, A. Krause, A. Regev, and M. Cuturi. “Optimal transport for single-cell and spatial omics”. In: *Nature Reviews Methods Primers* 4 (08/2024).
- [CY01] V. Camion and L. Younes. “Geodesic Interpolating Splines”. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition*. Ed. by M. Figueiredo, J. Zerubia, and A. K. Jain. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 513–527.
- [Cha+21] B. Charlier, J. Feydy, J. A. Glaunès, F. Collin, and G. Durif. “Kernel Operations on the GPU, with Autodiff, without Memory Overflows”. In: *JMLR abs/2004.11127* (2021).
- [CT13] N. Charon and A. Trouvé. “The Varifold Representation of Nonoriented Shapes for Diffeomorphic Registration”. In: *SIAM Journal on Imaging Sciences* 6.4 (01/2013), 2547–2580.
- [Cot08] C. J. Cotter. “The variational particle-mesh method for matching curves”. In: *Journal of Physics A: Mathematical and Theoretical* 41.34 (08/2008), p. 344003.

- [Dal+19] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu. “Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces”. In: *Medical Image Analysis* 57 (10/2019), 226–236.
- [Dav97] C. Davatzikos. “Spatial Transformation and Registration of Brain Images Using Elastically Deformable Models”. In: *Computer Vision and Image Understanding* 66.2 (1997), pp. 207–222.
- [Fey+17] J. Feydy, B. Charlier, F.-X. Vialard, and G. Peyré. *Optimal Transport for Diffeomorphic Registration*. 2017.
- [Fey+20] J. Feydy, A. Glaunès, B. Charlier, and M. Bronstein. “Fast geometric learning with symbolic matrices”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020, pp. 14448–14462.
- [FJ75] S. Fisher and J. Jerome. “Spline solutions to L1 extremal problems in one and several variables”. In: *Journal of Approximation Theory* 13.1 (1975), pp. 73–83.
- [Gla05] J. A. Glaunès. “Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l’anatomie numérique”. Thèse de doctorat dirigée par Younes, Laurent et Trouvé, Alain Mathématiques Paris 13 2005. PhD thesis. 2005, xiv–156 p.
- [GR87] L. Greengard and V. Rokhlin. “A fast algorithm for particle simulations”. In: *Journal of Computational Physics* 73.2 (1987), pp. 325–348.
- [Har90] R. Hardy. “Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988”. In: *Computers & Mathematics with Applications* 19.8 (1990), pp. 163–208.
- [Her+24] J. Hertrich, C. Wald, F. Altekrieger, and P. Hagemann. *Generative Sliced MMD Flows with Riesz Kernels*. 2024.
- [Jul+11] R. Julien, G. Peyré, J. Delon, and B. Marc. “Wasserstein Barycenter and its Application to Texture Mixing”. In: *Proc. SSVM’11*. Israel: Springer, 2011, pp. 435–446.
- [LGSL22] L. de Lara, A. González-Sanz, and J.-M. Loubes. *Diffeomorphic Registration using Sinkhorn Divergences*. 2022.
- [LS15] W. Li and P. Song. “A modified ICP algorithm based on dynamic adjustment factor for registration of point cloud and CAD model”. In: *Pattern Recognition Letters* 65 (2015), pp. 88–94.
- [Nad21] K. Nadjahi. “Sliced-Wasserstein distance for large-scale machine learning : theory, methodology and extensions”. PhD thesis. 11/2021.
- [Rue+99] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. “Nonrigid registration using free-form deformations: application to breast MR images”. In: *IEEE Transactions on Medical Imaging* 18.8 (1999), pp. 712–721.
- [Sej+13] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. “Equivalence of distance-based and RKHS-based statistics in hypothesis testing”. In: *The Annals of Statistics* 41.5 (10/2013).
- [Sri+10] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. G. Lanckriet. *Hilbert space embeddings and metrics on probability measures*. 2010.
- [Tia+24] L. Tian, H. Greer, R. Kwitt, F.-X. Vialard, R. S. J. Estepar, S. Bouix, R. Rushmore, and M. Niethammer. *uniGradICON: A Foundation Model for Medical Image Registration*. 2024.
- [Tro98] A. Trouvé. “Diffeomorphism groups and pattern matching in image analysis”. In: *Int. J. of Comp. Vis.* 28.3 (1998), pp. 213–221.
- [Via+12] F.-X. Vialard, L. Risser, C. J. Cotter, and D. Rueckert. “Diffeomorphic 3D Image Registration via Geodesic Shooting using an Efficient Adjoint Calculation.” In: *International Journal of Computer Vision* (2012), pp. –.
- [Wen04] H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 2004.

- [YBZ04] L. Ying, G. Biros, and D. Zorin. “A kernel-independent adaptive fast multipole algorithm in two and three dimensions”. In: *Journal of Computational Physics* 196.2 (2004), pp. 591–626.
- [You10] L. Younes. *Shapes and diffeomorphisms*. Vol. 171. Springer, 2010.
- [Zhu+21] W. Zhu, Y. Huang, D. Xu, Z. Qian, W. Fan, and X. Xie. *Test-Time Training for Deformable Multi-Scale Image Registration*. 2021.

APPENDIX A. APPENDIX

Lemma A.1. Consider the convex set $B_{0,TV} := \left\{ \Gamma \in \mathcal{M}_0(\Omega, \mathbb{R}^d) \mid |\Gamma|_{TV} \leq 1 \right\}$. Then, its set of extremal points is

$$\text{Ext}(B_{0,TV}) = \left\{ \sum_{i=1}^k v_i \delta_{x_i}, x_i \neq x_j (i \neq j) \in \Omega \text{ s.t. } (v_1, \dots, v_{k-1}) \text{ linearly independent and } \sum_{i=1}^k |v_i| = 1 \right\}. \quad (\text{A.1})$$

Proof. By the result of Fisher and Jerome [FJ75], the extreme points of the solutions to $\min_{\mu} \|\mu\|_{TV}$ such that $\mu(\Omega) = 0$ are necessarily of the form $\sum_{i=1}^k v_i \delta_{x_i}$ with $k \leq d$ and $v_i \neq 0$. Suppose now that v_1, \dots, v_{k-1} are not linearly independent. Then, there exists $\lambda_1, \dots, \lambda_{k-1}$ such that $\sum_{i=1}^{k-1} \lambda_i v_i = 0$. Introduce $\delta = \sum_{i=1}^{k-1} \lambda_i v_i \delta_{x_i}$ and $\tilde{\mu} = \sum_{i=1}^{k-1} v_i$. Consider $\nu_1 = \tilde{\mu} - \varepsilon \delta + \frac{1}{2} v_n \delta_{x_n}$, $\nu_2 = \tilde{\mu} + \varepsilon \delta + \frac{1}{2} v_n \delta_{x_n}$ for $\varepsilon > 0$. Then, by construction $\mu = \nu_1 + \nu_2$ for all ε and since $\delta(\Omega) = 0$, $\nu_2(\Omega) = \nu_1(\Omega) = \mu(\Omega) = 0$. For ε sufficiently small, we also have $\|\nu_1\|_{TV} = \sum_{i=1}^{k-1} |1 - \varepsilon \lambda_i| |v_i| + \frac{1}{2} |v_k|$ and $\|\nu_2\|_{TV} = \sum_{i=1}^{k-1} |1 + \varepsilon \lambda_i| |v_i| + \frac{1}{2} |v_k|$. As a consequence, $\mu = \nu_1 + \nu_2$ and $\|\nu_1\|_{TV} + \|\nu_2\|_{TV} = \|\mu\|_{TV}$, which shows that μ cannot be an extremal point since it is a linear combination of $\frac{1}{\|\nu_1\|_{TV}} \nu_1$ and $\frac{1}{\|\nu_2\|_{TV}} \nu_2$.

We now show that the measures in Formula (A.1) are indeed extremal. Assume that one of these measures μ can be written $\mu = t\nu_1 + (1-t)\nu_2$ where $t \in (0, 1)$ and ν_1, ν_2 of unit TV norm. Consider the decomposition of ν_1 (also ν_2) w.r.t. μ which can be written as $\nu_1 = \tilde{\nu}_1 + \nu_1^\perp$ with ν_1^\perp singular w.r.t. μ . Then, since $\mu = \tilde{\nu}_1 + \tilde{\nu}_2$ we get $\|\mu\| = 1$ implies $\nu_1^\perp = \nu_2^\perp = 0$. Now, we have $\nu_1 = \sum_{i=1}^k u_i \delta_{x_i}$ and $\nu_2 = \sum_{i=1}^k w_i \delta_{x_i}$ for some vectors u_i, w_i such that $\sum_{i=1}^k w_i = \sum_{i=1}^k u_i = 0$ and $u_i + w_i = v_i$ for $i = 1, \dots, k$. The TV norm of μ reads $1 = \sum_{i=1}^k |v_i| \leq \sum_{i=1}^k (|w_i| + |u_i|) = 1$ by the previous equalities. Therefore, by the Cauchy-Schwarz inequality, necessarily w_i and u_i are colinear and since $w_i + u_i = v_i$ they are colinear to v_i . Write $w_i = \alpha_i v_i$ and $u_i = \beta_i v_i$. Then, $\sum_{i=1}^k \alpha_i v_i = 0$ and $\sum_{i=1}^k \beta_i v_i = 0$. Since the rank of (v_1, \dots, v_k) is $k-1$, we get that $(\alpha_1, \dots, \alpha_k)$ and $(\beta_1, \dots, \beta_k)$ are necessarily colinear to $(1, \dots, 1)$. This implies that ν_1, ν_2 are colinear to μ and it leads to $\nu_1 = \nu_2 = \mu$. \square