T.E.R : Energy Distance Kernel for Shape Registration

SOUTARSON Bruny

June 2024

Supervised by François-Xavier Vialard

Table des matières

1 Presentation

During the second semester of my master's 1 in mathematics and applications at the Université Gustave Eiffel (UGE), I undertook a Supervised Research Project (TER in french). This academic module involves weekly or biweekly meetings to assess the progress of a research project supervised by a professor. Wishing to explore a field at the intersection of computer science and mathematics, I sought the supervision of François-Xavier Vialard, a faculty member at the Gaspard Monge computer science laboratory of UGE. He kindly agreed to supervise me during this semester on a project focused on the energy distance kernel for shape registration. We will first detail the theoretical foundations of this concept, followed by the numerical implementations I developed.

2 Acknowledgments

I would like to express my deepest gratitude to François-Xavier Vialard for agreeing to guide me through this research project, for his insightful responses to my questions, and for his valuable advice. My thanks also go to Siwan Boufadene, PhD candidate, for his significant contribution to my in-depth understanding of the subject. This TER has been an immensely rewarding experience, not only by broadening my academic horizons in areas previously unexplored during my mathematics studies, but also by allowing me to appreciate the dynamics and challenges of the research world from the inside

3 Introduction

3.1 Motivations : medical image registration

The field of medical imaging encompasses various techniques for acquiring and rendering images of the human body. In practice, medical imaging plays a crucial role in several areas :

— **Prevention** : For example, lung X-rays are performed to detect certain diseases.

— **Diagnosis** : It enables diagnoses by visualizing the inside of the human body without the need for surgical intervention.

— **Therapy** : It is used to guide treatments, such as using video images during a therapeutic procedure.

The modeling obtained from these images is therefore extremely important at each stage of the process. One of the major challenges in this field is the ability to compare two images or geometric objects of the same nature. This comparison could not only save time for medical staff but also help detect differences that are invisible or very difficult to discern with the naked eye. For example, consider the case of a patient whose cancer progression is being assessed through X-rays.

We will focus particularly on the technique of image registration, which involves finding the best transformation to align one image to another of the same nature by superimposing them. To achieve this, it is possible to identify correspondences between the two images, a process known as matching.

3.2 Registration methods using diffeomorphic flow

To effectively implement this technique, it is essential to mathematically formalize the notion of deformation. Let us consider an image I, assumed to be defined on a continuous domain Ω. A deformation involves repositioning the points in Ω. This modification is formalized by a transformation $\phi : \Omega \to \Omega$, which maps each point x to its initial position, thus transforming the image I into $\tilde{I} = I \circ \phi^{-1}$. To prevent information loss or ambiguities, it is crucial that ϕ is bijective. We also impose a certain regularity by requiring that ϕ and its inverse be continuous, or that these functions are continuously differentiable, which qualifies ϕ as a diffeomorphism.

A natural approach to combining diffeomorphisms is composition, a relation through which a group is obtained. However, there is no efficient technique for working in such a space. One could then think of using linear combinations of diffeomorphisms ; however, such a combination has no reason to remain a diffeomorphism. Driven by the popularity of linear techniques, we nevertheless wanted to pursue this approach by representing ϕ in the form $\phi = \mathrm{id} + u$, where u is a displacement field whose variations are assumed to be small so that ϕ remains a diffeomorphism. How can we construct this u ?
By adopting an approach from physics, we can take $v(t, x)$, which gives the momentum associated with a particle located at position x at time t. By setting

$\phi(t, x)$ as the position at time t of the particle that was at x at time 0, we obtain the equation,

$$\frac{\partial \phi}{\partial t}(t, x) = v(t, \phi(t, x)).$$

Now that we know we need diffeomorphisms to perform our transformations, we want to obtain the optimal diffeomorphism. We aim to find the diffeomorphism such that the "transport" of the first object is as close as possible to the second one. How can we achieve this ?

The problem is formulated as follows : given two images I_0 and I_1 (as functions on Ω), we seek to find a diffeomorphism $\phi : \Omega \to \Omega$ such that $I_1 \circ \phi \approx I_0$. A common method is to minimize

$$\int_\Omega |I_0 \circ \phi^{-1} - I_1|^2 \, dx$$

over the set of all possible diffeomorphisms. This problem is often ill-posed, as the minimum is not necessarily achieved, but gradient descent algorithms applied to this functional can produce viable solutions, provided they are executed correctly and stopped in time. Traditionally, a regularization term is added to this functional to minimize

$$J(\phi) = \int_\Omega |I_0 \circ \phi^{-1} - I_1|^2 \, dx + \lambda R(\phi), \tag{1}$$

where R quantifies the regularity of ϕ as a function.

To determine an appropriate measure of regularity, it is crucial to understand the underlying mathematical space. An ideal framework would be a Hilbert space V of vector fields, whose norm measures the cost of deformations while ensuring sufficient regularity of the vector fields.

Definition 3.1. A vector space V of vector fields on $\mathbb{R}^d$ is said to be admissible if it satisfies the following conditions :

1. V is a Hilbert space. Its norm will be denoted by $\| \cdot \|_V$ and its inner product by $\langle \cdot, \cdot \rangle_V$.
2. $(V, \| \cdot \|_V)$ is continuously embedded in $C_0^1(\mathbb{R}^d, \mathbb{R}^d)$, the space of C^1 vector fields on $\mathbb{R}^d$ that vanish at infinity, along with their partial derivatives. Thus, there exists a constant $c_V > 0$ such that for all $v \in V$,

$$\|v\|_\infty \leq c_V \|v\|_V.$$

According to [1], in this new context, minimizing with respect to ϕ is equivalent to minimizing with respect to v, a vector field. Thus, (1) becomes :

$$J(v) = \int_\Omega |I_0 \circ \phi^{-1} - I_1|^2 \, dx + \lambda \int_0^1 \|v_t\|^2 \, dt. \tag{2}$$

3.3 Reproducing kernel Hilbert space

With the aim of constructing the ideal framework, we focus on kernels and establish the connection between their regularity properties and those of the space in which they operate.

Definition 3.2. A **reproducing kernel Hilbert space** (RKHS) is a Hilbert space $\mathcal{H}$ of functions defined on a set $\mathcal{X}$ such that for all $x \in \mathcal{X}$, the evaluation at x is a continuous linear map on $\mathcal{H}$. Mathematically, this means that there exists a unique function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (or $\mathbb{C}$), called the *reproducing kernel*, such that for all $x \in \mathcal{X}$ and for all $f \in \mathcal{H}$, we have :

$$f(x) = \langle f, K(x, \cdot) \rangle_{\mathcal{H}}.$$

Definition 3.3. A **positive kernel** is a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ (or $\mathbb{C}$) that satisfies the following two properties :

1. K is symmetric : $K(x, y) = K(y, x)$ for all $x, y \in \mathcal{X}$.
2. K is positive : for any finite set $\{x_1, x_2, \ldots, x_n\} \subset \mathcal{X}$ and any vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n$ (or $\mathbb{C}^n$), we have :

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0.$$

   Equivalently, for all $n \in \mathbb{N}$ and all $x_1, \ldots, x_n \in \mathcal{X}$, the Gram matrix $\mathbf{K}$ is a symmetric, positive semi-definite matrix.

Positive kernels have several interesting properties :

1. **Symmetry** : $K(x, y) = K(y, x)$ for all $x, y \in X$.
2. **Cone Property** : If K_1 and K_2 are positive definite kernels, then $aK_1 + bK_2$ is also a positive definite kernel for $a, b \geq 0$.
3. **Product** : If K_1 and K_2 are positive definite kernels, then their product $K(x, y) = K_1(x, y)K_2(x, y)$ is also a positive definite kernel.
4. **Exponentiation** : If K is a positive definite kernel, then $e^{K(x,y)}$ is also a positive definite kernel.

Example 3.4. The linear kernel is defined by :

$$K(x, y) = \langle x, y \rangle.$$

The Gaussian kernel is defined by :

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

where $\sigma > 0$ is a parameter that controls the width of the kernel.

Let us also introduce a weaker notion than positivity, which will be used later.

Definition 3.5. A symmetric kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be **conditionally positive (or CP)** if, for any set of distinct points $\{x_1, x_2, \ldots, x_n\} \subseteq \mathcal{X}$ and any non-zero vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n$ such that :

$$\sum_{i=1}^{n} \alpha_i = 0,$$

we have :

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j K(x_i, x_j) \geq 0.$$

Example 3.6. The kernel function is given by :

$$\text{K(x, y)} = -\|x-y\|^2 = -\|x\|^2 - \|y\|^2 + 2\langle x, y\rangle. \tag{3}$$

Démonstration. Let $\{x_1, \ldots, x_n\} \subseteq \mathcal{X}$ with $\sum_{j=1}^{n} c_j = 0$. We have :

$$\sum_{j,k=1}^{n} c_j c_k K(x_j, x_k) = -\sum_{j,k=1}^{n} c_j c_k \left(\|x_j\|^2 + \|x_k\|^2 - 2\langle x_j, x_k\rangle\right) \tag{4}$$

$$= -\sum_{k=1}^{n} c_k \left(\sum_{j=1}^{n} c_j \|x_j\|^2\right) - \sum_{j=1}^{n} c_j \left(\sum_{k=1}^{n} c_k \|x_k\|^2\right) \tag{5}$$

$$+ 2 \sum_{j,k=1}^{n} c_j c_k \langle x_j, x_k\rangle \tag{6}$$

$$= 2 \sum_{j,k=1}^{n} c_j c_k \langle x_j, x_k\rangle \geq 0. \tag{7}$$

Thus, this kernel satisfies the necessary property to be conditionally positive. $\square$

The interest of this definition lies in the following fundamental theorem :

Theorem 3.7. *The positive vector-valued kernels of dimension m on $\mathcal{X}$ are exactly the reproducing kernels of the RKHS of functions defined on $\mathcal{X}$ and taking values in $\mathbb{R}^m$. More precisely,*

i) The reproducing kernel of an RKHS is a positive kernel.

ii) For any positive kernel k on $\mathcal{X}$ of dimension m, there exists a unique RKHS on $\mathcal{X}$, with values in $\mathbb{R}^m$, for which k is the reproducing kernel.

Thus, from an RKHS, one can obtain a positive kernel, and the converse is also true. Moreover, the following proposition shows the regularity connections between the two.

Proposition 3.8. *Let k be a positive vector-valued kernel of dimension m, continuous and bounded on $\mathbb{R}^d$, such that for all $x \in \mathbb{R}^d$, $k(x, \cdot)$ vanishes at infinity. Then the reproducing space of k is continuously embedded in $C_0(\mathbb{R}^d, \mathbb{R}^m)$.*

Démonstration. By hypothesis, the functions $k(x, \cdot)$ all belong to $C_0(\mathbb{R}^d, \mathbb{R}^m)$. Thus, the space H_0 that they generate is included in $C_0(\mathbb{R}^d, \mathbb{R}^m)$. On the other hand, by the definition of the reproducing kernel, for all $f \in H_0$ and $x \in \mathbb{R}^d$, $\alpha \in \mathbb{R}^m$, we have :

$$f(x) \cdot \alpha = \langle f, k(x, \cdot)\alpha \rangle_H = \langle k(x, \cdot)\alpha, f \rangle_H.$$

By applying the Cauchy-Schwarz inequality, we obtain :

$$|f(x) \cdot \alpha| \leq \|k(x, \cdot)\alpha\|_H \|f\|_H = \sqrt{\langle k(x, \cdot)\alpha, k(x, \cdot)\alpha \rangle_H} \|f\|_H,$$

and therefore :

$$\|f\|_\infty \leq \sup_{x \in \mathbb{R}^d} \|k(x, \cdot)\|_H \|f\|_H.$$

Consequently, the Cauchy sequences in H_0 converge in the sense of the uniform norm, and thus the elements of H (simple limits of Cauchy sequences) belong to $C_0(\mathbb{R}^d, \mathbb{R}^m)$ and satisfy :

$$\|f\|_\infty \leq \|k\|_\infty \|f\|_H.$$

Thus, H is continuously embedded in $C_0(\mathbb{R}^d, \mathbb{R}^m)$. □

The construction of the Hilbert space V of vector fields can therefore be achieved using a simple method : choose a sufficiently regular kernel and then define V as the unique RKHS corresponding to it.

According to [1], in the case where $\mathcal{X} = \{x^1, \ldots, x^n\}$, with the points x^i being distinct, for any element $v \in H$, there exist vectors $\alpha^i \in \mathbb{R}^d$ such that :

$$v(x) = \sum_{i=1}^{n} k_V(x, x^i)\alpha^i. \tag{8}$$

The trajectories $x_t^i = \phi_t(x^i)$ satisfy

$$x_t^i = x^i + \int_0^t \sum_{i=1}^n k_V(x_s^i, x_s^j)\alpha^j \, ds.$$

We can thus parameterize the minimization space using the vectors α^i. However, to define a new minimization space over these variables, we need to control the norm of the α^i. This is possible if we assume the strict positivity of the kernel k_V. Indeed, we have the equality :

$$\|v\|_V^2 = \sum_{i,j=1}^n \alpha^i \cdot k_V(x_t^i, x_t^j)\alpha^j.$$

We then obtain a new version of (1) :

Proposition 3.9. *Suppose that the kernel k_V is strictly positive. Then, when $\mathcal{X} = \{x^1, \ldots, x^n\}$, with all points x^i being distinct, the matching functional previously considered can be written as a function of the parameters α^i in the space $L^2([0,1], (\mathbb{R}^d)^n)$:*

$$J(\{\alpha^i\}) = \lambda \int_0^1 \sum_{i,j} \alpha^i \cdot k_V(x_t^i, x_t^j)\alpha^j \, dt + A(\{\alpha^i\}), \tag{9}$$

where the trajectories $x_t^i \in \mathbb{R}^d$ are calculated by solving the previous integral system. And

$$A((x_i)_{1\leq i\leq n}) = \sum_{i=1}^n |x_i - y_i|^2.$$

Note that the sum in (9) is performed, in principle, for any positive kernel in $O(n^2)$.

Recently, the work of [2] has focused on the following kernel due to its complexity.

Definition 3.10. Energy Distance Kernel (ED Kernel)
Let $x, y \in X$, where X is a metric space. The *Energy Distance Kernel* between x and y is defined by the following formula :

$$K(x, y) = -\|x - y\|,$$

where $\|x - y\|$ represents the norm induced by the metric of the space X.

The proof of the complexity in [2] for this kernel is non-trivial, so we will simply present the underlying idea. In dimension 1, for a sorted list $\{x_i\}_{i \in \mathbb{N}}$, we obtain

$$\sum_{\substack{i \in [1,N] \\ j \in [0,N]}} |x_i - x_j| = \sum_{i=1,\ldots,N} (2i - 1 - N)\, x_i. \tag{10}$$

Sorting a list is performed in $O(n \log(n))$, and computing the sum takes linear time, resulting in a total complexity of $O(n \log(n))$.
However, this kernel is not positive ; it is only conditionally positive. To demonstrate this, we introduce the following notions.

Definition 3.11. A metric space $(\mathcal{X}, d)$ is said to be of negative type if, for any set of distinct points $\{x_1, x_2, \ldots, x_n\} \subseteq \mathcal{X}$ and any non-zero vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}$ such that :

$$\sum_{i=1}^{n} \alpha_i = 0,$$

we have :

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j d(x_i, x_j) \leq 0.$$

According to [4], we have :

Theorem 3.12. *A metric space $\mathcal{X}$ is of negative type if there exists a Hilbert space H and a function $\phi : \mathcal{X} \to H$ such that for all $x, x' \in \mathcal{X}$, we have $d(x, x') = \|\phi(x) - \phi(x')\|^2$.*

Let us prove, in the simple case where $\mathcal{X} = \mathbb{R}$, that the energy distance kernel is conditionally positive by verifying that $(\mathbb{R}, |\cdot|)$ is of negative type.
We use $\phi(x) := \mathbb{1}_{[0,\infty[} - \mathbb{1}_{[x,\infty[}$ in $L^2(\mathbb{R}, \lambda)$, where λ is the Lebesgue measure. We consider several cases to demonstrate the equality with this function and calculate the norm in $L^2(\mathbb{R}, \lambda)$ of the difference between the indicator functions $\mathbb{1}_{[0,x]}$ and $\mathbb{1}_{[0,y]}$.

Case 1 : $x = y$

If $x = y$, then $\mathbb{1}_{[0,x]} = \mathbb{1}_{[0,y]}$, and therefore :

$$\mathbb{1}_{[0,x]} - \mathbb{1}_{[0,y]} = 0,$$

and :

$$\|\mathbb{1}_{[0,x]} - \mathbb{1}_{[0,y]}\|_{L^2}^2 = 0.$$

Case 2 : $x \neq y$

Sub-case 2.1 : $x < y$

In this case, we have :

$$\mathbb{1}_{[0,x]}(t) - \mathbb{1}_{[0,y]}(t) = \begin{cases} 0 & \text{si } t \in [0,x], \\ -1 & \text{si } t \in (x,y], \\ 0 & \text{si } t > y. \end{cases}$$

The squared L^2 norm of this function is calculated as follows :

$$\|\mathbb{1}_{[0,x]} - \mathbb{1}_{[0,y]}\|_{L^2}^2 = \int_x^y 1 \, d\lambda(t) = y - x.$$

Sub-case 2.2 : $x > y$

In this case :

$$\mathbb{1}_{[0,x]}(t) - \mathbb{1}_{[0,y]}(t) = \begin{cases} 0 & \text{si } t \in [0,y], \\ 1 & \text{si } t \in (y,x], \\ 0 & \text{si } t > x. \end{cases}$$

The squared L^2 norm of this function is :

$$\|\mathbb{1}_{[0,x]} - \mathbb{1}_{[0,y]}\|_{L^2}^2 = \int_y^x 1 \, d\lambda(t) = x - y.$$

Finally, we obtain :

$$\|\mathbb{1}_{[0,x]} - \mathbb{1}_{[0,y]}\|_{L^2}^2 = |x - y|,$$

from which the expected result follows.

The entire theoretical construction discussed in the previous section relies on two essential aspects of the kernel : its positivity and its regularity. Given the complexity advantages that the energy distance kernel offers, we would like to be able to use it in this context.

Two approaches are then considered :

- **Mathematically :** This approach involves examining the regularity of the deformations generated by the flows. This is a complex problem that falls outside the scope of my TER.
- **Numerically :** The fundamental question is : does it work ? Indeed, the entire purpose of the shape registration approach ultimately relies on numerical applications. If it does not work numerically, there is little point in studying the question theoretically.

In the next section, we will explore numerical applications to determine whether this kernel is viable. We will compare it with the gaussian kernel and an extended version of the energy distance kernel, which we define as follows.

Definition 3.13 (Modified ED Kernel)**.** Let $x, y \in \mathbb{R}$ (or the appropriate space). The *Modified ED Kernel* $K(x, y)$ is defined by the formula :

$$K(x, y) := -|x - y| + |x| + |y|,$$

where $|x|$ and $|y|$ denote the norms of x and y, respectively, and $|x - y|$ is the distance between x and y.

The advantage of this modified version is that it remains computable in $O(n \log n)$, as the added elements only require linear computation time. Furthermore, we obtain a positive kernel.

Démonstration. Note that :

$$K(x, y) = -|x - y| + |x| + |y| = 2 \min(x, y),$$

which is twice the covariance kernel of Brownian motion. More precisely, let $(W_t)_{t>0}$ be a Brownian motion and $s, t > 0$. Then, we have

$$\mathrm{Cov}(W_s, W_t) = \min(s, t) = \frac{1}{2} K(x, y).$$

Covariance matrices are always symmetric and positive semi-definite. Since $K(x, y)$ can be expressed as twice the covariance of Brownian motion, this implies that $K(x, y)$ is a positive semi-definite kernel.

□

4 Numerical Implementation

4.1 Euler Scheme on Particles

Recall that we have $(v(t,x))_{t\in[0,1]}$, a family of vector fields that gives the momentum associated with a particle x at time t, and $\phi(t,x)$, the position at time t of the particle that was located at x at time 0. Then, according to the flow equation, we have the identity defined by the flow equation :

$$\frac{\partial\phi}{\partial t}(t,x) = v(t,\phi(t,x)).$$

To simulate the motion of particles under the influence of forces, we perform the explicit Euler scheme associated with the previous differential equation. This allows us to obtain an approximate value of ϕ_{t+1} at time $t+1$ from its value at time t as follows :

$$\phi(t+1,x) = \phi(t,x) + v(t,\phi(t,x)).$$

Furthermore, recall that according to (8), we have $v(x) = \sum_{i=1}^{n} k_V(x,x^i)\alpha^i$. We then implement the following algorithm.

Algorithm 1 Trajectory optimization between points using an explicit Euler scheme

Initialize the initial positions q_0, the initial momenta p_0, and the target positions z.
Set the number of time steps nt.
Define the kernel h to be used.
Create the optimizer for p_0.
for $i = 1$ **to** nt **do**
    for all *member* **in** $q[i]$ **do**
        $q_l(i+1) = q_l(i) + 1/nt \times v_i(q(i))$
    end for
    Append the new position to the list of positions.
end for
Optimize p_0 to minimize the loss relative to z.

Here is an illustration of the results obtained.

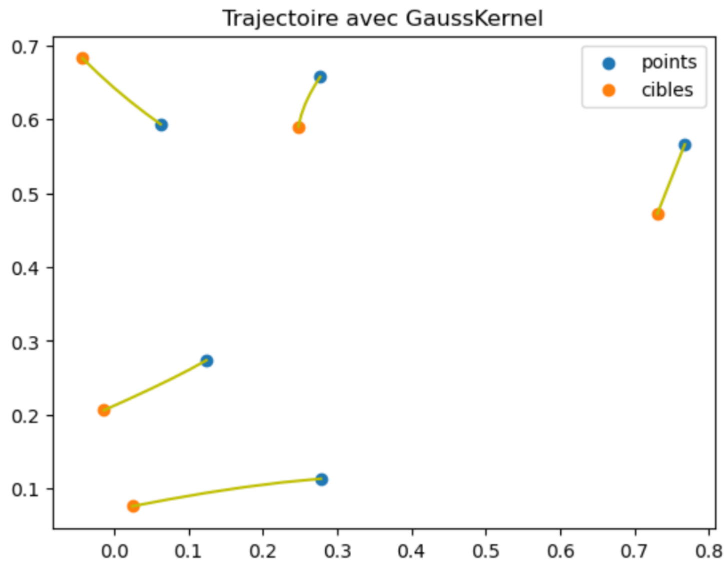

FIGURE 1 – Trajectory with GaussKernel of points with $\sigma = 0.25$

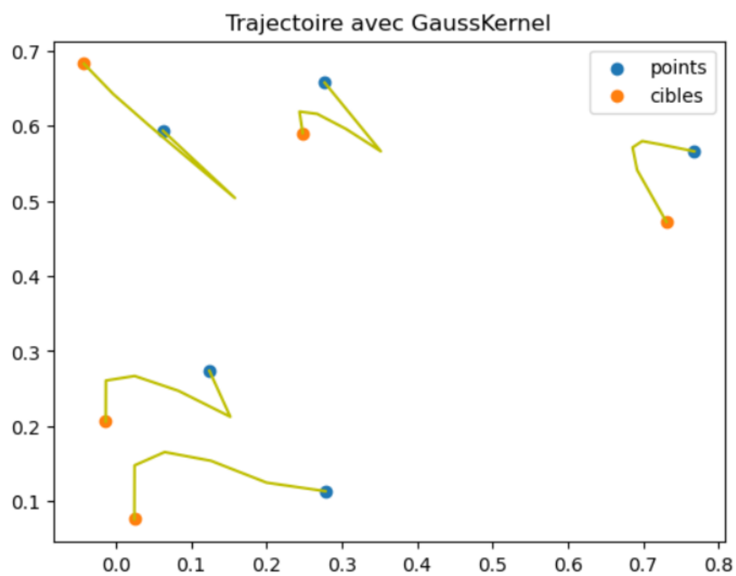

FIGURE 2 – Trajectory with GaussKernel of points with $\sigma == 0.9$

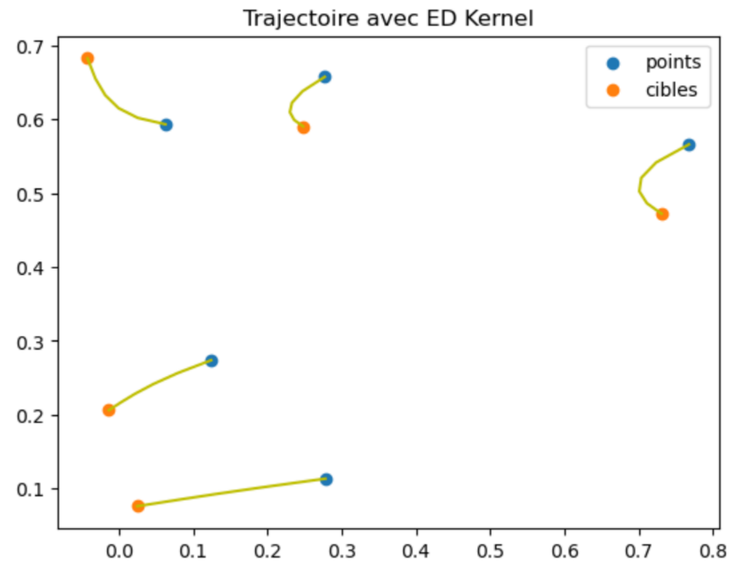

FIGURE 3 – Trajectory with Modified ED Kernel of points

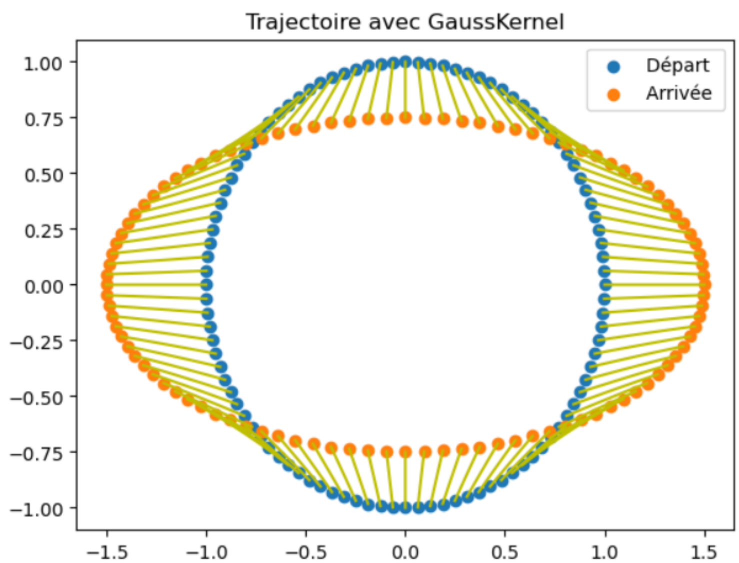

FIGURE 4 – Trajectory with GaussKernel (Circle to Ellipse) $\sigma = 0.25$

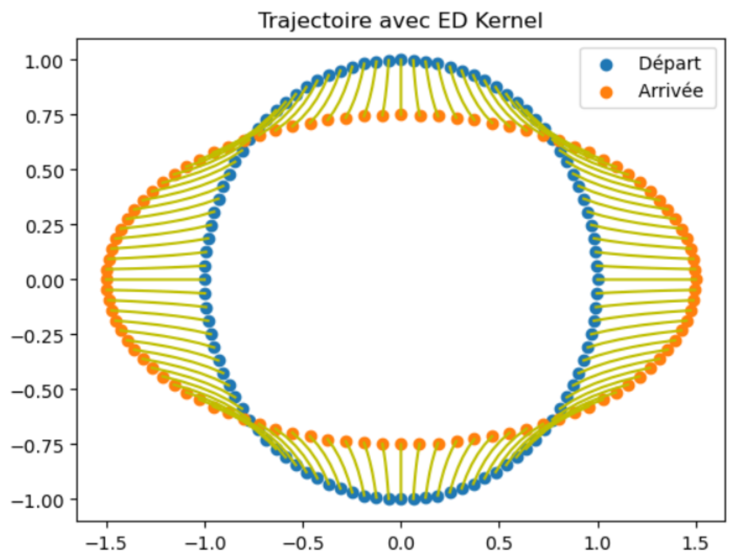

FIGURE 5 – Trajectory with Modified ED Kernel (Circle to Ellipse)

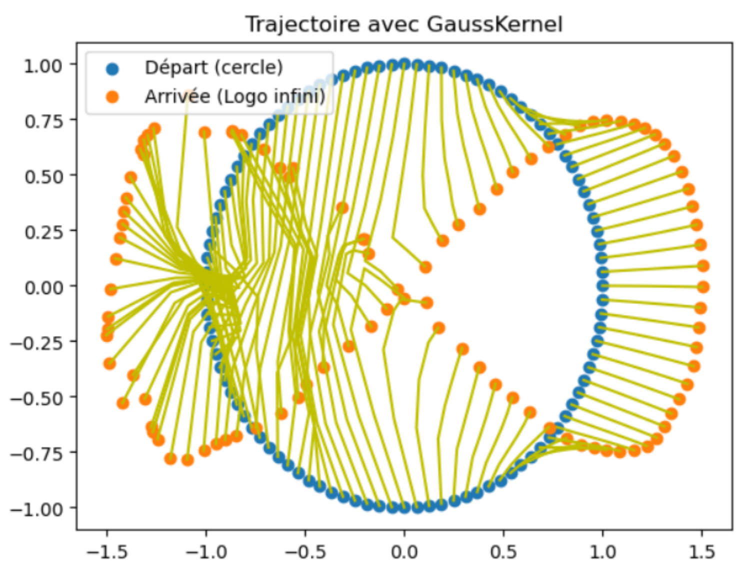

FIGURE 6 – Trajectory with GaussKernel (Circle to Infinity) $\sigma = 0.25$

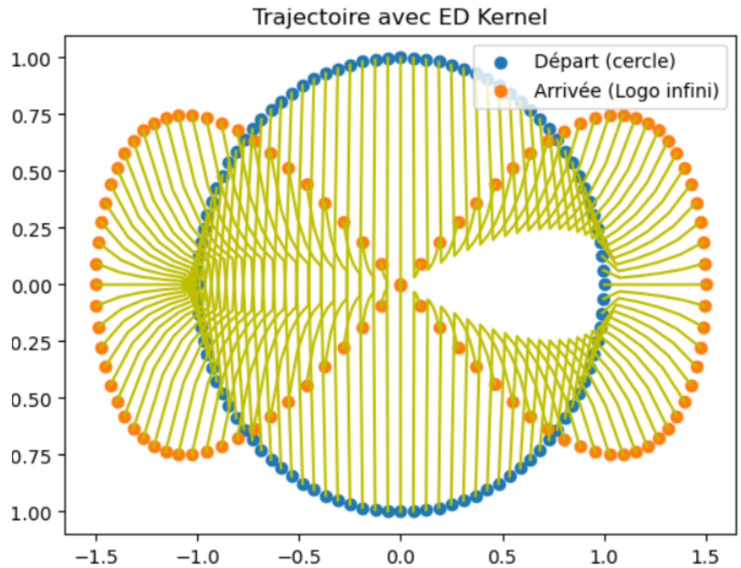

FIGURE 7 – Trajectory with Modified ED Kernel (Circle to Infinity)

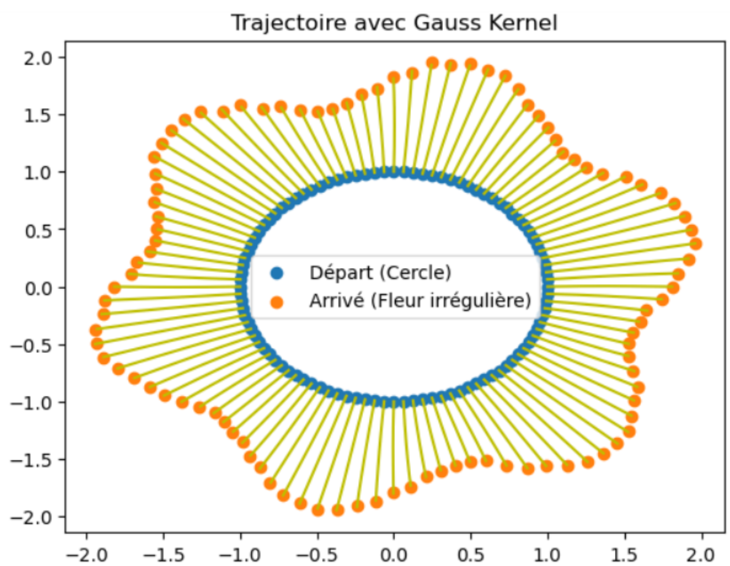

FIGURE 8 – Trajectory with GaussKernel (Circle to Irregular Flower) $\sigma = 0.25$

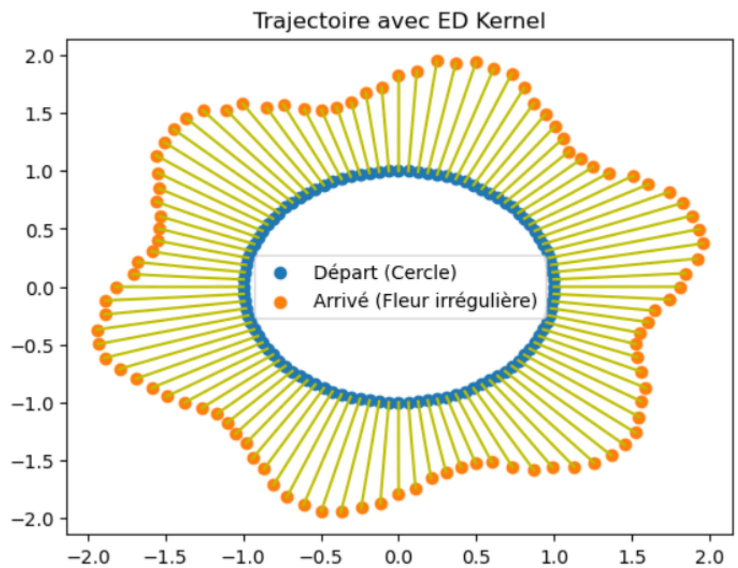

FIGURE 9 – Trajectory with Modified ED Kernel (Circle to Irregular Flower)

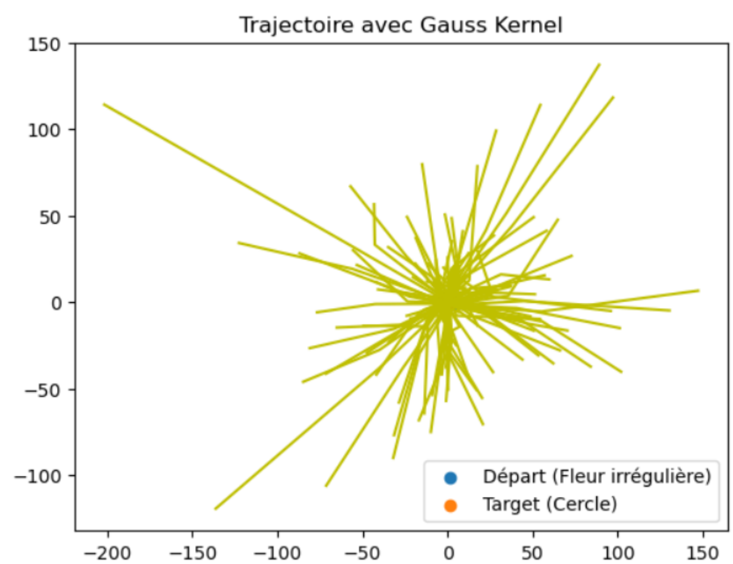

FIGURE 10 – Trajectory with GaussKernel (Irregular Flower to Circle) $\sigma = 0.25$

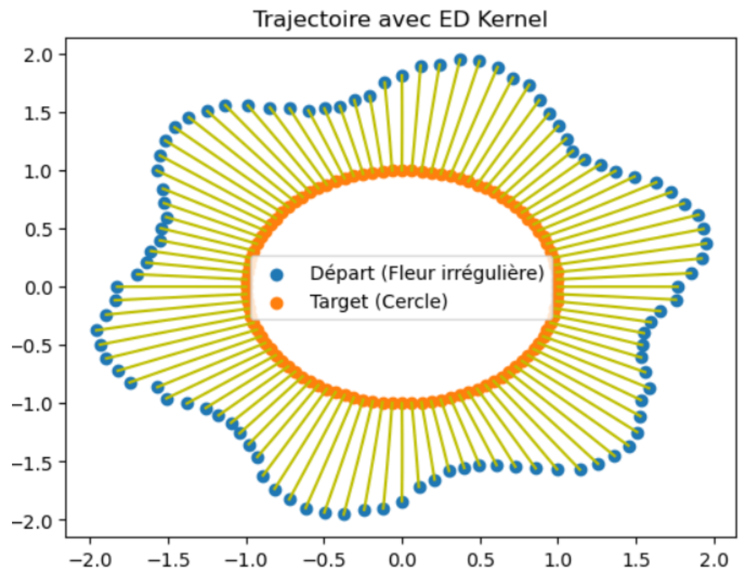

FIGURE 11 – Trajectory with Modified ED Kernel (Irregular Flower to Circle)

In simple point matching (cf. 1, 2, 3), the performance of the Gaussian kernel is strongly influenced by the choice of σ. Indeed, when σ is very small, the trajectories are mainly straight lines. The performance of the Modified ED kernel, on the other hand, is better than that of the Gaussian kernel when σ is large.

In the transition from one shape to another with linear trajectories (cf. 4, 5, 8, 9), no noticeable difference is observed. However, as the shapes become more complex, the advantages of the Modified ED kernel become apparent, producing better results (cf. 6, 7). Furthermore, during the contraction of a shape, a significant difference is observed : the Gaussian kernel proves to be ineffective in this context (cf. 11, 10).

5 Conclusion

Observations from several examples show very encouraging results for the energy distance kernel, revealing advantages that go beyond its lower complexity. For instance, unlike the Gaussian kernel used for comparison, which depends on a specific parameter, the energy distance kernel requires none. These numerical observations therefore motivate further theoretical exploration.
A priori, the ED kernel, being less regular, generates a richer space of deformations than that produced by a Gaussian kernel. However, this could lead to deformations that no longer correspond to diffeomorphisms. The study of this aspect goes beyond the scope of this project.

6 Bibliography

Références

[1] Glaunes. Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique.

[2] Johannes Hertrich, Christian Wald, Fabian Altekrüger, and Paul Hagemann. Generative Sliced MMD Flows with Riesz Kernels. *arXiv*, preprint arXiv :2305.11463, 2023. doi : 10.48550/arXiv.2305.11463.

[3] Russell Lyons. Distance covariance in metric spaces. *Annals of Probability*, 41(5), 2013. doi : 10.1214/12-AOP803.

[4] I. J. Schoenberg. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space. *Ann. of Math.*, 1937.

[5] Alain Trouvé. Cours géométrie et espaces de formes. Cours présenté au master MVA.

[6] Alain Trouvé and Laurent Younes. Mise en correspondance d'objets déformables.