A New Geometric Metric in the Shape and Size Space of Curves in $\mathbb{R}^n$

Irene Epifanio [1,†] , **Vicent Gimeno** [2,†] , **Ximo Gual-Arnau** [3,†] **and M. Victoria Ibáñez-Gual** [2,*,†]

[1] Department of Mathematics-IF, Universitat Jaume I, 12071 Castelló, Spain; epifanio@uji.es
[2] Department of Mathematics-IMAC, Universitat Jaume I, 12071 Castelló, Spain; gimenov@uji.es
[3] Department of Mathematics-INIT, Universitat Jaume I, 12071 Castelló, Spain; gual@uji.es
[*] Correspondence: mibanez@uji.es
[†] All authors contributed equally to this work.

Abstract: Shape analysis of curves in $\mathbb{R}^n$ is an active research topic in computer vision. While shape itself is important in many applications, there is also a need to study shape in conjunction with other features, such as scale and orientation. The combination of these features, shape, orientation and scale (size), gives different geometrical spaces. In this work, we define a new metric in the shape and size space, $\mathcal{S}_2$, which allows us to decompose $\mathcal{S}_2$ into a product space consisting of two components: $\mathcal{S}_4 \times \mathbb{R}$, where $\mathcal{S}_4$ is the shape space. This new metric will be associated with a distance function, which will clearly distinguish the contribution that the difference in shape and the difference in size of the elements considered makes to the distance in $\mathcal{S}_2$, unlike the previous proposals. The performance of this metric is checked on a simulated data set, where our proposal performs better than other alternatives and shows its advantages, such as its invariance to changes of scale. Finally, we propose a procedure to detect outlier contours in $\mathcal{S}_2$ considering the square-root velocity function (SRVF) representation. For the first time, this problem has been addressed with nearest-neighbor techniques. Our proposal is applied to a novel data set of foot contours. Foot outliers can help shoe designers improve their designs.

Keywords: shape space; square-root velocity function (SRVF); outliers

1. Introduction

Shape analysis of curves in $\mathbb{R}^n$, where $n \geq 2$, is an important branch in many applications, including computer vision and medical imaging. Using the landmark representation of objects, Dryden et al. [1] studied the joint shape and size features of objects. However, an over-abundance of digital data, especially image data, is prompting the need for a different kind of shape analysis. In particular, the representation of shapes as elements of infinite-dimensional Riemannian manifolds with a given metric is of interest at this time and has important applications [2,3]. More recently, Srivastava et al. [4] presented a special representation of curves, called the square-root velocity function, or SRVF, under which a specific elastic metric becomes an L^2 metric and simplifies the shape analysis. This approach was analyzed by Kurtek et al. [5] on different scenarios, corresponding to different combinations of physical properties of the curves: shape, size, location and orientation. When the metric used in these infinite-dimensional spaces is invariant with respect to scaling, translation, rotation and reparameterization, the Riemannian manifold that represents the space of curves is known as the shape space, and following the notation of [5] will be denoted as $\mathcal{S}_4$. However, in many of the applications, other features such as orientation or size (scale) also play important roles and need to be incorporated into the underlying framework. A prime example is medical imaging, where the size of

the anatomical structure of interest can provide important diagnostic information. In the case where the curve length (size) is considered, the feature space is called the shape and size (or shape and scale) space and will be denoted as $\mathcal{S}_2$ [5]. Other spaces denoted as $\mathcal{S}_1$ and $\mathcal{S}_3$ consider also changes in the orientations of the curves [5]. In this work, we will focus on the shape space $\mathcal{S}_4$ and the shape and size space $\mathcal{S}_2$, which are in general, completely different infinite-dimensional Riemannian manifolds. Curves in $\mathcal{S}_2$ are different elements of the space if their shape or scale are different. Curves in S_4 are different elements of the space if their shape is different.

It seems natural, however, that the distance between two curves in space $\mathcal{S}_2$ should be related to the distance of these same curves in space $\mathcal{S}_4$. We can thereby discern whether the distance between the curves in space $\mathcal{S}_2$ is due, to a greater extent, to their difference in size or to their difference in shape.

In this sense, in [6], the Sobolev-type metric given in [3] for the shape space of planar closed curves is extended to the space of all planar closed curves where the metric considered exhibits a decomposition of the space of closed planar curves into a product space consisting of three components; that is, centroid translations, scale changes and curves in the shape space.

In this approach, we will consider representations of curves in $\mathbb{R}^n$ from square-root velocity functions (SRVF). Using these representations, we will consider two feature spaces studied in [5]: the shape space $\mathcal{S}_4$ and the shape and size space $\mathcal{S}_2$. The metric in $\mathcal{S}_4$ will be the same as in [5]; however, we propose a new metric in $\mathcal{S}_2$, which is completely different to the metric considered in [5]. This metric enjoys the property that $\mathcal{S}_2$ can be decomposed into a product space consisting of two components: $\mathcal{S}_4 \times \mathbb{R}$, where the second space is related to the length (size) of the curve.

The outline of the paper is as follows: In Section 2, we review the SRVF representation of curves and the standard elastic metrics. In Section 3, we introduce the new metric in $\mathcal{S}_2$. The mean shape and geodesics with this new metric are introduced in Sections 4 and 5, respectively. A comparison of the proposed metric and the standard elastic metric is carried out in Section 6 in a controlled setting with simulated curves, where we show the advantages of our proposal. We propose a procedure to detect outlier contours in $\mathcal{S}_2$ considering the SRVF representation. To the best of our knowledge, this is the first time this problem has been addressed with nearest-neighbor (NN) techniques. This is introduced in Section 7. Not only that, but so far outlier detection (in the multivariate context) in Anthropometry has only been used as a cleaning technique, for correcting or removing the outliers before analyzing data in the multivariate context [7,8]. However, outliers report very valuable information in the footwear design process: outliers can indicate which kinds of feet are more different from the rest and could therefore cause fitting problems in footwear if the design is not appropriate. In Section 8, our proposal is applied to a novel data set of foot contours. Finally, Section 9 contains the conclusions.

The code and data for reproducing the results are available at http://www3.uji.es/~epifanio/RESEARCH/metric.zip.

2. Classical Spaces of Curves in $\mathbb{R}^n$ for the SRVF Representation

In this section, we review some results from [9]. In particular, we consider the SRVF representation of curves in $\mathbb{R}^n$ and we summarize the main results for the shape space, $\mathcal{S}_4$, and for the shape and size space, $\mathcal{S}_2$, with the standard elastic metrics.

Let $\beta : [0,1] \longrightarrow \mathbb{R}^n$ be a parameterized curve that is absolutely continuous on $[0,1]$. The square-root velocity (SRVF) of β is defined as the function $q : [0,1] \longrightarrow \mathbb{R}^n$ given by:

$$q(t) = \frac{\beta'(t)}{\sqrt{|\beta'(t)|}}. \tag{1}$$

As it can be seen in [9], this representation exists even where $|\beta'(t)| = 0$.

For every $q \in \mathbb{L}^2([0,1], \mathbb{R}^n)$, there is a curve β (unique up to translation) such that the given q is the SRVF function of that β. In fact,

$$\beta(t) = \int_0^t q(s)|q(s)| \, ds. \tag{2}$$

If a curve β is of length one, then $\int_0^1 |q(t)|^2 dt = 1$. Furthermore, the hypersphere

$$\mathcal{C}^0 = \left\{ q : [0,1] \longrightarrow \mathbb{R}^n \mid \int_0^1 |q(t)|^2 dt = 1 \right\}, \tag{3}$$

is a Hilbert manifold.

One way to study the shape and size (scale) space of open curves is to consider as a pre-shape space $\mathbb{L}^2 = \mathbb{L}^2([0,1], \mathbb{R}^n)$ with the usual inner product. To take care of the rotation and reparameterization of the curve β, we remember that a rotation is an element of $SO(n)$, the special orthogonal group of $n \times n$ matrices; and a reparameterization is an element of $\Gamma = \{\gamma : [0,1] \rightarrow [0,1] \mid : \gamma(0) = 0, \gamma(1) = 1, \gamma$ is a diffeomorphism$\}$.

The action of a reparameterization $\gamma \in \Gamma$ transforms the curve $\beta : [0,1] \rightarrow \mathbb{R}^n$ to the curve $t \mapsto \beta(\gamma(t))$. Hence, by the definition of the SRVF of a curve, we define the action of $\gamma \in \Gamma$ in $\mathcal{C}_0$ by

$$\gamma(q(t)) := q(\gamma(t))\sqrt{\gamma'(t)}.$$

Likewise, the action of $O \in SO(n)$ on $q \in \mathcal{C}^0$ is just $O(q)(t) = O(q(t))$. We shall denote the combined action

$$(O, \gamma)q(t) := O(q(\gamma(t)))\sqrt{\gamma'(t)}, \quad O \in SO(n), \quad \gamma \in \Gamma.$$

The orbit of a function $q \in \mathbb{L}^2$ is

$$[q] = \{(O, \gamma)(q) \mid (O, \gamma) \in SO(n) \times \Gamma\}. \tag{4}$$

If we consider the metric in $\mathbb{L}^2$ given by the usual inner product

$$\langle v_1, v_2 \rangle_{L^2} = \int_0^1 \langle v_1(t), v_2(t) \rangle \, dt, \tag{5}$$

the feature space of interest is:

$$\mathcal{S}_2 = \left\{ [q] \ / \ q \in \mathbb{L}^2 \right\}, \tag{6}$$

and the distance in $\mathcal{S}_2$ is:

$$d_2([q_1], [q_2]) = \inf_{O \in SO(n), \gamma \in \Gamma} |q_1 - (O, \gamma)(q_2)|_{L^2}. \tag{7}$$

Then, the geodesic between q_1 and the optimal reparameterization of q_2, which is denoted as $q_2^* = (O, \gamma^*)(q_2)$, is

$$\Psi_\tau = (1 - \tau)q_1 + \tau q_2^*, \ \tau \in \mathbb{R}. \tag{8}$$

On the other hand, the shape space (without considering the size (scale) of the curves) is

$$\mathcal{S}_4 = \{[q] \mid q \in \mathcal{C}^0\}. \tag{9}$$

The distance in $\mathcal{S}_4$ is given by

$$d_4([q_1], [q_2]) = \inf_{O \in SO(n), \gamma \in \Gamma} \cos^{-1}(\langle q_1, (O, \gamma)(q_2) \rangle_{\mathbb{L}^2}), \tag{10}$$

and the geodesic between q_1 and q_2^* is

$$\alpha(\tau) = \frac{1}{\sin\theta} \left(\sin(\theta(1-\tau))q_1 + \sin(\theta\tau)q_2^* \right), \tag{11}$$

where $\theta = \cos^{-1}\langle q_1, q_2^* \rangle_{\mathbb{L}^2}$.

3. A New Metric in the Shape and Size Space of Curves in $\mathbb{R}^n$

When $\mathcal{S}_2$ is considered as shape and size space, it is difficult to distinguish whether the distance between two shapes $[q_1]$ and $[q_2]$ is due to the difference in shape or to the difference in size between the corresponding curves β_1 and β_2. We are therefore going to consider another shape-size space for curves that will be isometric to $\mathcal{S}_2$ with another appropriate product metric.

Instead of considering in $\mathbb{L}^2$ the usual $\mathbb{L}^2$-metric given in Equation (5), if $q \in \mathbb{L}^2([0,1], \mathbb{R}^n)$, for any two vectors v_1, v_2 in $T_q\mathbb{L}^2([0,1]) \equiv \mathbb{L}^2([0,1], \mathbb{R}^n)$, we will consider the following metric to endow $\mathbb{L}^2([0,1])$ with a Riemannian structure,

$$\hat{g}(v_1, v_2) := \frac{1}{R^2(q)} \langle v_1, v_2 \rangle_{L^2}. \tag{12}$$

where

$$R(q) := \left(\int_0^1 |q(t)|^2 dt \right)^{\frac{1}{2}}.$$

The case $R(q) = 0$ will be excluded, which will mean that curves of length 0 are not considered in our space.

From this metric, we will endow (Theorem 1) $\mathcal{C}^0 \times \mathbb{R}$ with a Riemannian structure in such a way that $\mathcal{C}^0 \times \mathbb{R}$ will be isometric to $\left(\mathbb{L}^2([0,1]), \hat{g} \right)$. This isometry will be exported (Theorem 2) to an isometry between $\mathcal{S}_2$ and $\mathcal{S}_4 \times \mathbb{R}$.

Therefore, we obtain a new metric which enjoys the property that $\mathcal{S}_2$ can be decomposed into a product space consisting of two components: $\mathcal{S}_4 \times \mathbb{R}$, where the second space is related with the length (size) of the curve. This new metric is associated with a distance function, see Corollary 3, given by

$$d_{\text{new}}([p], [q]) = \sqrt{d_4^2\left(\left[\frac{p}{R(p)}\right], \left[\frac{q}{R(q)}\right]\right) + \ln^2\left(\frac{R(p)}{R(q)}\right)}. \tag{13}$$

This distance is invariant under rotations and under rescaling in the sense that $d_{\text{new}}(O([p]), O([q])) = d_{\text{new}}([p], [q])$ and $d_{\text{new}}(\lambda[p], \lambda[q]) = d_{\text{new}}([p], [q])$ for any O in $SO(n)$ and any $\lambda > 0$.

3.1. An Isometry between $\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}$ and $\mathcal{C}^0 \times \mathbb{R}$

We will begin this section defining a function F which will provide an isometry between $\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}$ and $\mathcal{C}^0 \times \mathbb{R}$. We consider the smooth map

$$F : \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\} \to \mathcal{C}^0 \times \mathbb{R},$$

$$q \mapsto F(q) := \left(\frac{q}{R(q)}, \ln R(q) \right).$$

Observe that F is well defined for any $\mathbb{L}^2([0,1], \mathbb{R}^n)) \setminus \{\vec{0}\}$ where

$$\{\vec{0}\} := \left\{ q \in \mathbb{L}^2([0,1], \mathbb{R}^n) \ / \ \int_0^1 |q(t)|^2 dt = 0 \right\} = R^{-1}(0).$$

The function F has immediate smooth inverse given by

$$F^{-1} : \mathcal{C}^0 \times \mathbb{R} \to \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\},$$
$$(q,t) \mapsto F^{-1}(q,t) = e^t q.$$

The Functions R, π and F and Their Properties

Now we state some properties of the function R, some properties of $\pi(q) := \frac{1}{R(q)} q$ from $\mathbb{L}^2$ to $\mathcal{C}^0$, that is the natural projection given by the normalization of an SRVF using its norm, and some properties of the function F:

Proposition 1. *Properties of the functions R, π and F:*

1. *Given a curve $\beta(t)$ and $q(t) = \frac{\beta'(t)}{\sqrt{|\beta'(t)|}}$. Then, $R(q) = \sqrt{L(\beta)}$ is the square root of the length of the curve β.*
2. *For any $\gamma \in \Gamma$,*
    $$R\left(q(\gamma(t)) \sqrt{\gamma'(t)} \right) = R(q(t)).$$
3. *For any $O \in \text{SO}(n)$,*
    $$R(O(q(t))) = R(q(t)).$$
4. *For any $O \in \text{SO}(n)$,*
    $$\pi(Oq) = O\pi(q).$$
    Namely, the rotation group commutes with the projection π.
5. *For any $\lambda > 0$*
    $$\pi(\lambda q) = \pi(q)$$
6. *Given $(O, \gamma) \in \text{SO}(n) \times \Gamma$, and $q \in \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}$ then*
    $$F(O(q(\gamma(t)))\sqrt{\gamma'(t)}) = (O(\pi(q(\gamma(t)))), \ln R(q)).$$
7. *F is smooth and admits smooth inverse with never vanishing differential map.*

From the properties of π, R and F, we can conclude the following diffeomorphisms.

Corollary 1. *From the function F we have that*

1. *$\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\} \overset{diff}{\sim} \mathcal{C}^0 \times \mathbb{R}$.*
2. *Since $\mathcal{S}_2 = \mathbb{L}^2/\Gamma \times \text{SO}(n)$, then*
    $$\mathcal{S}_2 \setminus \{\vec{0}\} \overset{diff}{\sim} \left(\mathcal{C}^0/\Gamma \times \text{SO}(n) \right) \times \mathbb{R} \overset{diff}{\sim} \mathcal{S}_4 \times \mathbb{R}.$$

As already mentioned at the beginning of the section, if $q \in \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{0\}$, the tangent space at q can be identified with $\mathbb{L}^2([0,1], \mathbb{R}^n)$ itself,

$$T_q \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{0\} \equiv \mathbb{L}^2([0,1], \mathbb{R}^n)$$

and for any two vectors v_1, v_2 in $T_q \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{0\}$ we will use the following metric to endow $\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{0\}$ with a Riemannian structure,

$$\hat{g}(v_1, v_2) := \frac{1}{R^2(q)} \langle v_1, v_2 \rangle_{L^2}. \tag{14}$$

Therefore, using $F^{-1} : \mathcal{C}^0 \times \mathbb{R} \to \mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}$, we can pullback the metric $\hat{g}$ to $\hat{g}^*$ in $\mathcal{C}^0 \times \mathbb{R}$ by the differential map $dF^{-1} : T_{(q,t)}\mathcal{C}^0 \times \mathbb{R} \to T_{e^t p}\mathbb{L}^2([0,1],, \mathbb{R}^n) \setminus \{0\}$ in order to endow $\mathcal{C}^0 \times \mathbb{R}$ with a Riemannian structure in such a way that $(\mathcal{C}^0 \times \mathbb{R}, \hat{g}^*)$ will be isometric to $(\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{0\}, \hat{g})$.

Theorem 1. *There is an isometry* $F : (\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}, \hat{g}) \longrightarrow \mathcal{C}^0 \times \mathbb{R}$ *given by*

$$F(p) = \left(\frac{p}{R(p)}, \ln R(p) \right), \tag{15}$$

where the usual product metric is considered in $\mathcal{C}^0 \times \mathbb{R}$.

Proof. We have shown in Corollary 1 that F is a diffeomorphism; therefore, we only have to prove that the pullback $\hat{g}^*$ of the metric $\hat{g}$ is the usual product metric in $\mathcal{C}^0 \times \mathbb{R}$.

$$T_{(q,t)}\mathcal{C}^0 \times \mathbb{R} = T_q\mathcal{C}^0 \oplus T_t\mathbb{R}$$

Given two vectors in $v_1, v_2 \in T_{(q,t)}\mathcal{C}^0 \times \mathbb{R}$ the pullback $\hat{g}^*(v_1, v_2)$ is given by

$$\begin{aligned} \hat{g}^*(v_1, v_2) &= \hat{g}(dF^{-1}(v_1), dF^{-1}(v_2)) = \\ &= \frac{1}{R^2(e^t q)} \langle dF^{-1}(v_1), dF^{-1}(v_2) \rangle_{\mathbb{L}_2}. \end{aligned}$$

But

$$R^2(e^t q) = \int_0^1 |e^t q(s)|^2 ds = e^{2t} \int_0^1 |q|^2 ds = e^{2t},$$

because $q \in \mathcal{C}^0$. Therefore, for any two vectors $v_1, v_2 \in T_{(q,t)}\mathcal{C}^0 \times \mathbb{R}$

$$\hat{g}^*(v_1, v_2) = e^{-2t} \langle dF^{-1}(v_1), dF^{-1}(v_2) \rangle_{\mathbb{L}_2}.$$

Now, first of all, we need to prove that in the tangent space to $\mathcal{C}^0 \times \mathbb{R}$ at (q,t), $T_q\mathcal{C}^0$ is orthogonal to $T_t\mathbb{R}$. In order to do that, consider two vectors $v_1 \in T_q\mathcal{C}^0$ and $v_2 \in T_t\mathbb{R}$ and two curves $\gamma_1, \gamma_2 : (-\epsilon, \epsilon) \to \mathcal{C}^0 \times \mathbb{R}$, such that

$$\begin{cases} \gamma_1(s) = (\gamma_1^1(s), t), & \gamma_1(0) = (q,t), & \frac{d}{ds}\gamma_1(s)|_{s=0} = v_1 \\ \gamma_2(s) = (q, \gamma_2^2(s)), & \gamma_1(0) = (q,t), & \frac{d}{ds}\gamma_2(s)|_{s=0} = v_2 \end{cases}$$

Then,

$$dF^{-1}(v_1) = \frac{d}{ds}F^{-1}(\gamma_1(s))|_{s=0} = \frac{d}{ds}\left(e^t \gamma_1^1(s) \right)|_{s=0} = e^t v_1$$

with $v_1 \in T_q\mathcal{C}^0$. Likewise,

$$\begin{aligned} dF^{-1}(v_2) &= \frac{d}{ds}F^{-1}(\gamma_2(s))|_{s=0} = \frac{d}{ds}\left(e^{\gamma_2^2(s)} q \right)|_{s=0} \\ &= v_2 e^t q \end{aligned}$$

where $v_2 \in \mathbb{R}, q \in \mathcal{C}^0$. Hence,

$$g^*(v_1, v_2) = e^{-2t} \langle e^t v_1, v_2 e^t q \rangle_{\mathbb{L}_2} = v_2 \langle v_1, q \rangle_{\mathbb{L}_2} = 0$$

because $\langle v_1, q \rangle_{\mathbb{L}_2} = 0$ for any $v_1 \in T_q\mathcal{C}^0$.

Similarly if $v_1, v_2 \in T_q\mathcal{C}^0$,

$$\hat{g}^*(v_1, v_2) = e^{-2t} \langle e^t v_1, e^t v_2 \rangle_{\mathbb{L}_2} = \langle v_1, v_2 \rangle_{\mathbb{L}_2}$$

and if $v_1, v_2 \in T_t \mathbb{R}$

$$\hat{g}^*(v_1, v_2) = e^{-2t} \langle v_1 e^t q, v_2 e^t q \rangle_{\mathbb{L}_2} = v_1 v_2 \langle q, q \rangle_{\mathbb{L}_2} = v_1 v_2.$$

□

Since we are using the usual product metric in $\mathcal{C}^0 \times \mathbb{R}$, we conclude the following result:

Corollary 2. *Let* $\mathrm{d}_{\hat{g}}$ *denote the distance function in* $(\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}, \hat{g})$, *then*

$$
\begin{aligned}
\mathrm{d}_{\hat{g}}(p, q) &= \sqrt{\mathrm{d}_{\mathcal{C}^0}^2(\pi(p), \pi(q)) + \mathrm{d}_{\mathbb{R}}^2(\ln R(p), \ln R(q))} \\
&= \sqrt{\mathrm{d}_{\mathcal{C}^0}^2 \left(\frac{p}{R(p)}, \frac{q}{R(q)} \right) + \ln^2 \left(\frac{R(p)}{R(q)} \right)}
\end{aligned}
\tag{16}
$$

where $\mathrm{d}_{\mathcal{C}^0}$ *is the usual distance in* $\mathcal{C}^0$.

From this explicit expression of the distance, it is easy to see that the metric $\hat{g}$ is invariant under the action of reparameterizations and rotations.

Proposition 2. *The group* $SO(n) \times \Gamma$ *acts isometrically on* $(\mathbb{L}^2([0,1], \mathbb{R}^n) \setminus \{\vec{0}\}, \hat{g})$.

Proof. We need to prove that

$$
\begin{aligned}
\mathrm{d}_{\hat{g}}((O, \gamma)(p), (O, \gamma)(q)) &= \sqrt{\mathrm{d}_{\mathcal{C}^0}^2 \left(\frac{(O, \gamma)(p)}{R((O, \gamma)(p))}, \frac{(O, \gamma)(q)}{R((O, \gamma)(q))} \right) + \ln^2 \left(\frac{R((O, \gamma)(p))}{R((O, \gamma)(q))} \right)} \\
&= \mathrm{d}_{\hat{g}}(p, q), \quad \forall (O, \gamma) \in SO(n) \times \Gamma.
\end{aligned}
$$

By Proposition 1 we know that $R((O, \gamma)(p)) = R(p)$ (and $R((O, \gamma)(q)) = R(q)$). Hence, the proposition follows because

$$
\begin{aligned}
\mathrm{d}_{\mathcal{C}^0}^2 \left(\frac{(O, \gamma)(p)}{R((O, \gamma)(p))}, \frac{(O, \gamma)(q)}{R((O, \gamma)(q))} \right) &= \mathrm{d}_{\mathcal{C}^0}^2 \left(\frac{(O, \gamma)(p)}{R(p)}, \frac{(O, \gamma)(q)}{R(q)} \right) \\
&= \int_0^1 \left\| \frac{(O, \gamma)(p)(t)}{R(p)} - \frac{(O, \gamma)(q)(t)}{R(q)} \right\|^2 dt \\
&= \int_0^1 \left\| O\left(\left(\frac{p(\gamma(t))}{R(p)} - \frac{q(\gamma(t))}{R(q)} \right) \sqrt{\gamma'(t)} \right) \right\|^2 dt \\
&= \int_0^1 \left\| \frac{p(\gamma(t))}{R(p)} - \frac{q(\gamma(t))}{R(q)} \right\|^2 \gamma'(t) dt = \int_0^1 \left\| \frac{p(\tau)}{R(p)} - \frac{q(\tau)}{R(q)} \right\|^2 d\tau \\
&= \mathrm{d}_{\mathcal{C}^0}^2 \left(\frac{p}{R(p)}, \frac{q}{R(q)} \right)
\end{aligned}
$$

□

3.2. The Isometry between $\mathcal{S}_2$ and $\mathcal{S}_4 \times \mathbb{R}$

Using the isometry given by Theorem 1, an isometry between $\mathcal{S}_2$ and $\mathcal{S}_4 \times \mathbb{R}$ can be constructed as stated in the following theorem:

Theorem 2. *The isometry F can be exported to an isometry [F] by using the following commutative diagram*

$$
\begin{array}{ccc}
\mathbb{L}^2([0,1],\mathbb{R}^n) \setminus \{\vec{0}\} & \xrightarrow{\ F\ } & \mathcal{C}^0 \times \mathbb{R} \\
\Big\downarrow \Pi_1 & & \Big\downarrow \Pi_2 \\
\mathcal{S}_2 \setminus \{0\} & \xrightarrow{\ [F]\ } & \mathcal{S}_4 \times \mathbb{R}
\end{array}
$$

where $\Pi_1(q) = [q]$ *and* $\Pi_2(q,t) = ([q],t)$. *Namely ,*

$$
[F][q] = \Pi_2(F(q))
$$

for any $q \in \Pi_1^{-1}([q])$.

Proof. From Proposition 2 we know that $SO(n) \times \Gamma$ acts by isometries on $(\mathbb{L}^2 \setminus \{\vec{0}\}, \hat{g})$. Therefore, since the action of the group $\Gamma \times SO(n)$ on $(\mathbb{L}^2 \setminus \{\vec{0}\}, \hat{g})$ and on $\mathcal{C}^0$ is by isometries, and bearing in mind the diffeomorphisms in Corollary 1, we obtain the result. $\square$

The isometry $[F]$ of the above theorem can be used to obtain the expression of the new distance function.

Corollary 3. *Let* d_{new} *denote the distance function in* $\mathcal{S}_2 \setminus \{\vec{0}\}$ *when the isometry* $[F]$ *of Theorem 2 is considered, then*

$$
d_{new}([p],[q]) = \sqrt{d_4^2([\pi(p)],[\pi(q)]) + d_{\mathbb{R}}^2(\ln R(p), \ln R(q))}
$$

$$
= \sqrt{d_4^2\left(\left[\frac{p}{R(p)}\right], \left[\frac{q}{R(q)}\right]\right) + \ln^2\left(\frac{R(p)}{R(q)}\right)}. \tag{17}
$$

In the following proposition we are proving that d_{new} is a well defined distance function.

Proposition 3. *Let* d_{new} *denote the distance function in* $\mathcal{S}_2 \setminus \{\vec{0}\}$ *when the isometry* $[F]$ *of Theorem 2 is considered, then*

1. $d_{new}([p],[q]) = d_{new}([q],[p])$ *for all* $[p],[q] \in \mathcal{S}_2 \setminus \vec{0}$.
2. $d_{new}([p],[q]) = 0$, *if and only if,* $d_4\left(\left[\frac{p}{R(p)}\right], \left[\frac{q}{R(q)}\right]\right) = 0$ *and* $R(p) = R(q)$.
3. *For any* $[p],[q],[r] \in \mathcal{S}_2 \setminus \vec{0}$

$$
d_{new}([p],[r]) \leq d_{new}([p],[q]) + d_{new}([q],[r]).
$$

4. $d_{new}(\lambda[p],\lambda[q]) = d_{new}([q],[p])$ *for all* $[p],[q] \in \mathcal{S}_2 \setminus \vec{0}, \lambda \in \mathbb{R}, \quad \lambda \neq 0$.

Proof. Most of the statements of the proposition follows directly from the definition and the properties of d_4 and R. We shall prove the triangle inequality for the sake of completeness. Let us denote by $\vec{v}, \vec{w} \in \mathbb{R}^2$ the vectors given by

$$
\vec{v} := \left(d_4\left(\left[\frac{p}{R(p)}\right], \left[\frac{q}{R(q)}\right]\right), \ln\left(\frac{R(p)}{R(q)}\right)\right), \quad \vec{w} := \left(d_4\left(\left[\frac{q}{R(q)}\right], \left[\frac{r}{R(r)}\right]\right), \ln\left(\frac{R(q)}{R(r)}\right)\right)
$$

Then, by applying the triangle inequality for d_4,

$$
\begin{aligned}
(d_{new}([p],[r]))^2 &= d_4^2\left(\left[\frac{p}{R(p)}\right],\left[\frac{r}{R(r)}\right]\right) + \left(\ln\left(\frac{R(p)}{R(r)}\right)\right)^2 \\
&\leq \left(d_4\left(\left[\frac{p}{R(p)}\right],\left[\frac{q}{R(q)}\right]\right) + d_4\left(\left[\frac{q}{R(q)}\right],\left[\frac{r}{R(r)}\right]\right)\right)^2 + \left(\ln\left(\frac{R(p)}{R(q)}\right) + \ln\left(\frac{R(q)}{R(r)}\right)\right)^2 \\
&= \|\vec{v}+\vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\|\vec{v}\|\,\|\vec{w}\|\cos(\theta) \leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\|\vec{v}\|\,\|\vec{w}\| = (\|\vec{v}\| + \|\vec{w}\|)^2 \\
&= (d_{new}([p],[q]) + d_{new}([q],[r]))^2 .
\end{aligned}
$$

□

4. The Mean Shape

Given $\{\beta_1,\cdots,\beta_n\}$, a sample of parameterized curves, and their corresponding SRVF, $\{q_1,\cdots,q_n\}$, the Karcher mean shape regarding the new metric d_{new} is defined as

$$
[\hat{\mu}_{new}] = \arg\min_q \sum_{i=1}^n d_{new}([q],[q_i])^2 = \arg\min_q \sum_{i=1}^n \left(d_4^2\left(\left[\frac{q}{R(q)}\right],\left[\frac{q_i}{R(q_i)}\right]\right) + \ln^2\left(\frac{R(q)}{R(q_i)}\right)\right).
$$

The value of $R(q)$ that minimizes $\sum_{i=1}^n \ln^2\left(\frac{R(q)}{R(q_i)}\right)$ is the geometric mean of the $\{R(q_1),\cdots,R(q_n)\}$, i.e., $\sqrt[n]{\prod_{i=1}^n R(q_i)}$, and a gradient-based approach for finding the value of $q/R(q)$ that minimizes

$$
\sum_{i=1}^n d_4^2\left(\left[\frac{q}{R(q)}\right],\left[\frac{q_i}{R(q_i)}\right]\right)
$$

can be found in [10,11]. The detailed algorithm to find the Karcher mean in the shape space $\mathcal{S}_2$ can be found in [12].

Given $\hat{\mu}_{C_0}$ the Karcher mean of $\{q_i/R(q_i)\}_{i=1,\cdots,n}$ in the shape space, the Karcher mean in the shape and size space with the new metric is obtained as

$$
\hat{\mu}_{new} = \hat{\mu}_{C_0}\sqrt[n]{\prod_{i=1}^n R(q_i)} \quad .
$$

Hence, applying Equation (2), the mean curve is

$$
\hat{\beta}_{new}(t) = \left(\prod_{i=1}^n L(\beta_i)\right)^{\frac{1}{n}} \int_0^t \hat{\mu}_{C_0}(s)|\hat{\mu}_{C_0}(s)|ds \tag{18}
$$

where $L(\beta_i)$ is the length of the curve β_i.

5. Geodesics

Moreover, we can use the isometry given by the Theorem 2 to provide an explicit expression for the geodesics in $(\mathbb{L}^2([0,1])\setminus\{\vec{0}\},\hat{g})$.

Corollary 4. *Any geodesic in $\mathcal{C}^0 \times \mathbb{R}$ is obtained as*

$$
t \mapsto (\alpha(t), a + b\,t)
$$

where α is a geodesic in $\mathcal{C}^0$ and $a, b \in \mathbb{R}$. Therefore any geodesic in $(\mathbb{L}^2([0,1])\setminus\{\vec{0}\},\hat{g})$ can be written as

$$
C(t) = Ae^{bt}\alpha(t)
$$

with $A, b \in \mathbb{R}$ and α a geodesic in $\mathcal{C}^0$.

For any p and q in $\mathbb{L}^2([0,1]) \setminus \{\vec{0}\}$, the above corollary allows us to obtain the geodesic segment (with respect to $\hat{g}$) joining p and q. Namely, we want a geodesic curve (in the new metric) $\gamma : [0,1] \to \mathbb{L}^2([0,1]) \setminus \{\vec{0}\}$ such that $\gamma(0) = p$ and $\gamma(1) = q$. By using the corollary, we only have to consider the geodesic segment in $\mathcal{C}^0$ with $\alpha(0) = \frac{p}{R(p)}$ and $\alpha(1) = \frac{q}{R(q)}$ and the geodesic segment in $\mathbb{R}$ joining $\ln R(p)$ and $\ln R(p)$, i.e.,

$$\tau \mapsto \ln R(p) + \tau \left(\ln R(q) - \ln R(p)\right)$$

Therefore, the geodesic segment joining p and q is

$$\gamma(\tau) = e^{\ln R(p) + \tau(\ln R(q) - \ln R(p))} \alpha(\tau) = R(p)^{1-\tau} R(q)^\tau \alpha(\tau), \quad \tau \in [0,1]$$

This geodesic segment in $\mathbb{L}^2([0,1]) \setminus \{\vec{0}\}$ can be understood as a family of deformations of curves in the following sense: if we have two curves $\beta_1 : [0,1] \to \mathbb{R}^n$ and $\beta_2 : [0,1] \to \mathbb{R}^n$, we obtain two points in $\mathbb{L}^2([0,1]) \setminus \{\vec{0}\}$

$$p(t) = \frac{\beta_1'(t)}{\sqrt{|\beta_1'(t)|}} \quad \text{and} \quad q(t) = \frac{\beta_2'(t)}{\sqrt{|\beta_2'(t)|}}$$

with $R(p(t)) = \sqrt{\mathrm{L}(\beta_1)}$ and $R(q(t)) = \sqrt{\mathrm{L}(\beta_2)}$. The geodesic segment α in $\mathcal{C}_0$ joining $\frac{p(t)}{R(p(t))}$ and $\frac{q(t)}{R(q(t))}$ is a family of length-one curves which can be labeled with $\tau \in [0,1]$,

$$\alpha_\tau : [0,1] \to \mathbb{R}^n, \quad \alpha_0(t) = \frac{p(t)}{R(p(t))}, \quad \alpha_1(t) = \frac{q(t)}{R(q(t))}$$

Finally, we obtain the family of curves

$$\tilde{\gamma}_\tau(t) = \mathrm{L}^{1-\tau}(\beta_1) \mathrm{L}^\tau(\beta_1) \int_0^t \alpha_\tau(s)|\alpha_\tau(s)|ds, \quad \tau \in [0,1]$$

with

$$\tilde{\gamma}_0(t) = \beta_1(t), \quad \tilde{\gamma}_1(t) = \beta_2(t).$$

6. Application to a Simulated Data Set

In order to check the performance of the new metric (Equation (13)), we have simulated several curves with different shapes and sizes. In particular, we have simulated ten 3D cylindric spirals, $\beta_{1i}(t)$ $i = 1, \cdots, 10$, $t \in [0,1]$, and ten circumferences, $\beta_{2i}(t)$ $i = 1, \cdots, 10$, $t \in [0,1]$, from:

$$x_{1i} = a_i \cos(8\pi t); \quad y_{1i} = a_i \sin(8\pi t); \quad z_{1i} = b_i t;$$
$$x_{2i} = r_i \cos(2\pi t); \quad y_{2i} = r_i \sin(2\pi t); \quad z_{2i} = 0;$$

where $t \in [0,1]$ and $\forall i \in \{1, \cdots, 10\}$, $b_i = 10 + e_{1i}$, $e_{1i} \sim N(0,1)$, $a_i = \sqrt{\frac{L^2 - b_i^2}{64\pi^2}}$, $L = 270$, $r_i = 15 + e_{2i}$, $e_{2i} \sim N(0, 2.5)$ (so all these spirals will have different shape and the same length, and all the circumferences will have the same shape but different lengths). Figure 1a,b shows the simulations obtained.

The Karcher means of the ten spirals and of the ten circumferences are computed with the new metric ($\hat{\beta}_{new}$, Equation (18)) and by using the distance proposed by [5] in the shape and size space $\mathcal{S}_2$. These means are shown in Figure 2, where the original curves are plotted in light blue; $\hat{\beta}_{new}$ is plotted in black color and $\hat{\beta}_2$, the Karcher-mean using the distance d_2, is plotted in red color. As can be seen, in Figure 2a,b, there is a very slight difference between the means $\hat{\mu}_{new}$ and $\hat{\mu}_2$ of the ten spirals. Figure 2c,d show that the means coincide in the case of the circumferences.
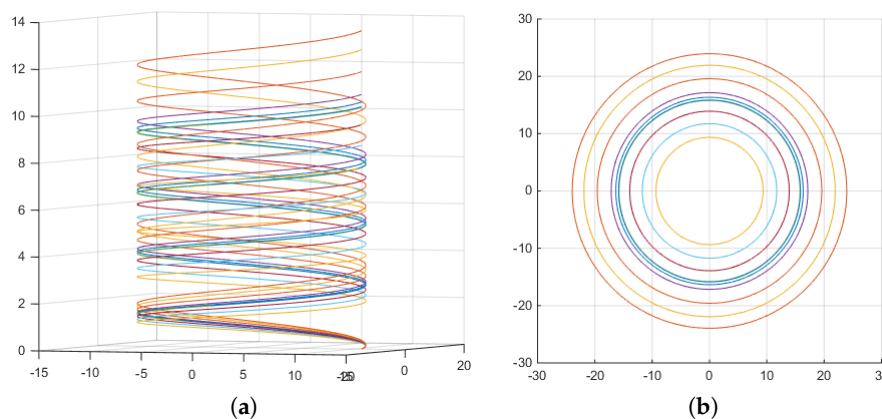
Figure 1. Simulated curves. (**a**) Ten spirals with different shapes and common length. (**b**) Ten circumferences with common shape and different lengths.
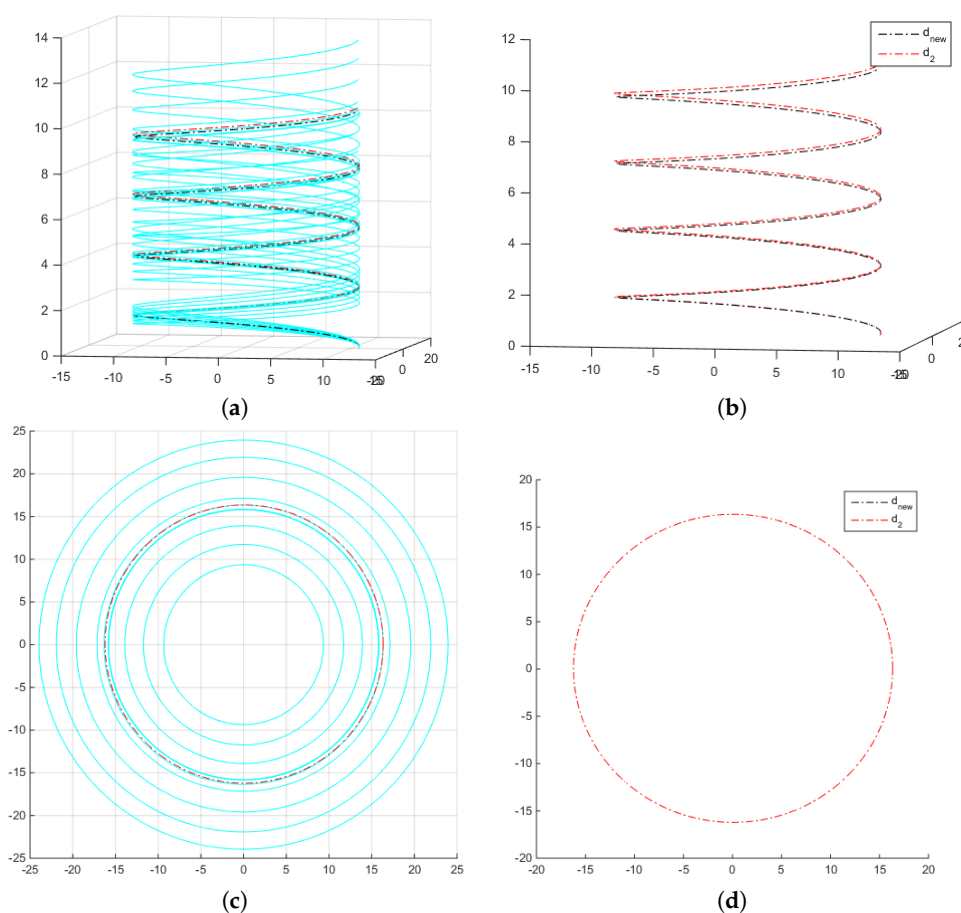
Figure 2. (**a**,**c**) Simulated curves and their corresponding means. (**b**,**d**) Comparison of the Karcher means obtained with the two different distances.

An example comparing the geodesics obtained with d_{new} and d_2, can be seen in Figure 3, without great differences among them.

Finally, the distance matrices D_{new} and D_2 between the twenty curves are computed using both metrics, and in order to compare the performance of d_2 and d_{new}, a multidimensional scaling (MDS) analysis [13] has been carried out. The MDS algorithm is a descriptive data reduction procedure to display the information contained in a $(m \times m)$-distance matrix, D, in a low-dimensional space such that the between-object distances are preserved as well as possible. Then, for each distance matrix D, the method looks for a set of orthogonal variables $\{y_1, \cdots, y_p\}$, $p < m$ such that the Euclidean

distances of the elements with respect to these variables are as close as possible to the distances given in the original matrix D. In Figure 4, MDS has been applied to the distance matrices computed with both metrics. In both graphics (Figure 4a,b), the black points represent the twenty spirals α_{1i} shown in Figure 1a, and the green points represent the twenty circumferences α_{2i} shown in Figure 1b.
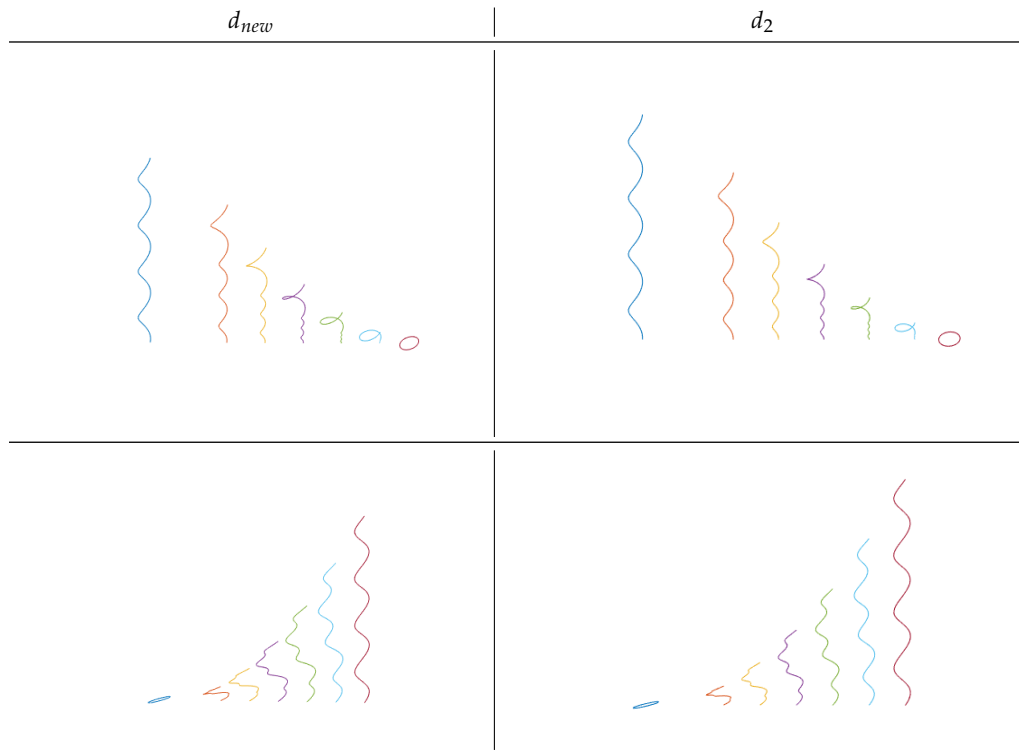

Figure 3. Comparison of the geodesic obtained using the two different distances.
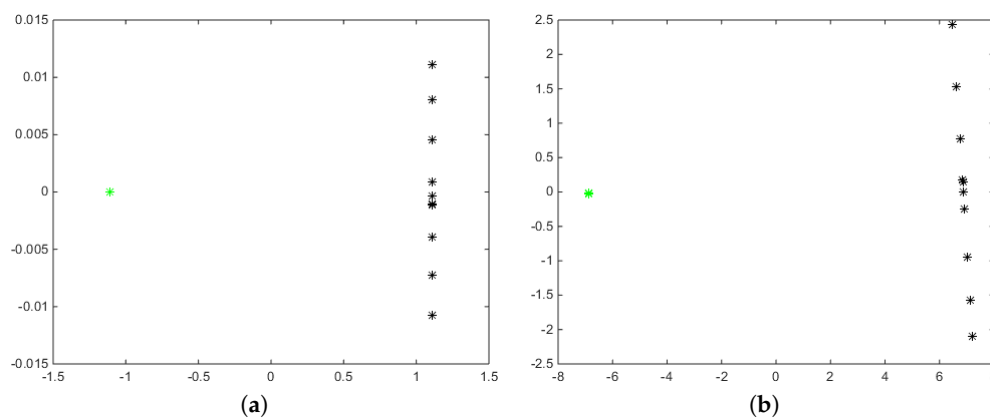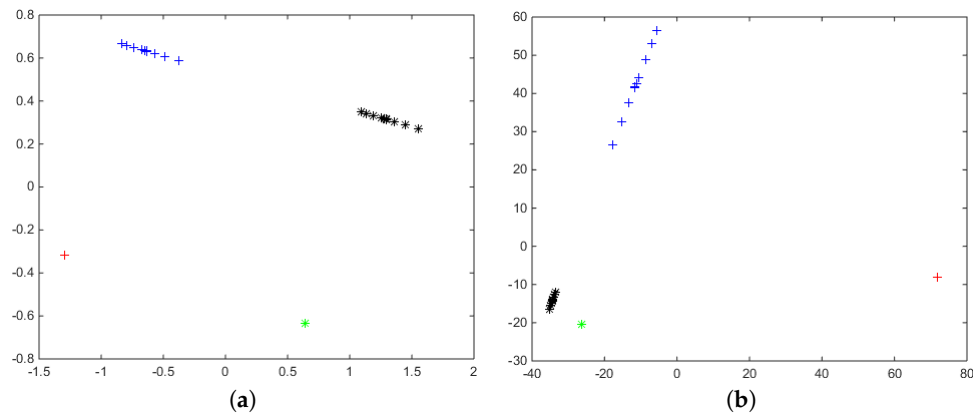

Figure 4. Multidimensional scaling (MDS) applied to the distance matrices: (**a**) using d_{new}, (**b**) using d_2. The ten spirals are marked in green and the ten circumferences are represented with black asterisks.

As can be seen, there are slights differences among the MDS representations of both metrics. If we perform a *k*-means cluster with $k = 2$ from D_{new} and D_2, in both cases the two groups are perfectly recovered. We also recover the two groups if we apply DBSCAN [14,15].

If we re-scale the twenty figures, multiplying them by 50, and we consider the twenty resulting curves jointly to the twenty original ones, we can compute again the distance matrices D_{new} and D_2 between the 40 curves. The MDS scaling representation of these distance matrices can be found in Figure 5. In both graphics, the initial spirals $\{\beta_{i1}\}_{i=1,\cdots,10}$ are plotted in green; the circumferences $\{\beta_{i2}\}_{i=1,\cdots,10}$ are plotted in black, and their scaled versions, $\{50\beta_{i1}\}_{i=1,\cdots,10}$ and $\{50\beta_{i2}\}_{i=1,\cdots,10}$ are plotted in red and blue color, respectively.

Figure 5 shows one important difference between the performance of both metrics. By definition, d_{new} is invariant to changes of scale i.e., $d_{new}(\beta_{il}, \beta_{jm}) = d_{new}(k\beta_{il}, k\beta_{jm})$, $\forall k \in \mathbb{R}$ and this equality does not hold for d_2, where the distance among curves increases with the scaling factor.

Figure 5. MDS applied to the distance matrices: (**a**) using d_{new}, (**b**) using d_2. The ten initial spirals are marked in green and the ten initial circumferences are represented with black asterisk, the spirals re-scaled by 50 are the red asterisks and the circumferences are plotted in blue.

If we perform a *k*-means cluster analysis with *k* = 4 from the distance matrices, the four groups are recovered from D_{new}, but for D_2, the distance among the scaled circumferences increases regarding to the distance among the initial circumferences, so the group of large circumferences is splitted into two clusters while the initial (short) curves (spirals and circumferences) are joined in a unique cluster (Figure 6). However, the algorithm DBSCAN applied on both distance matrices, allow us in both cases recover again the four initial groups.
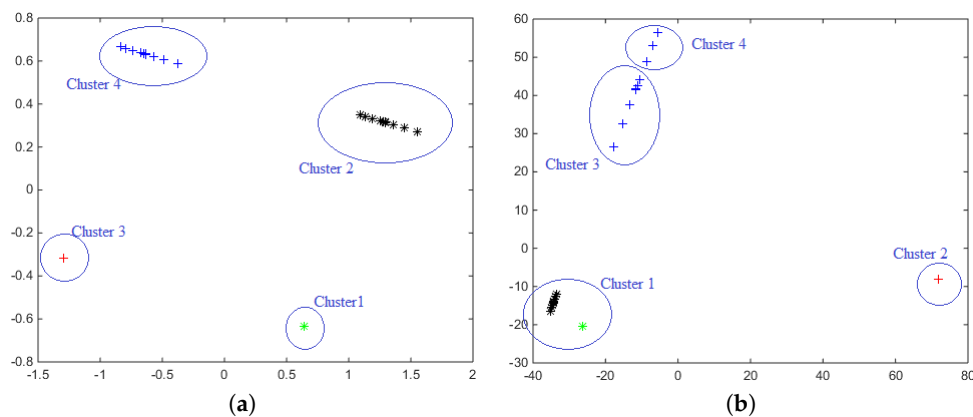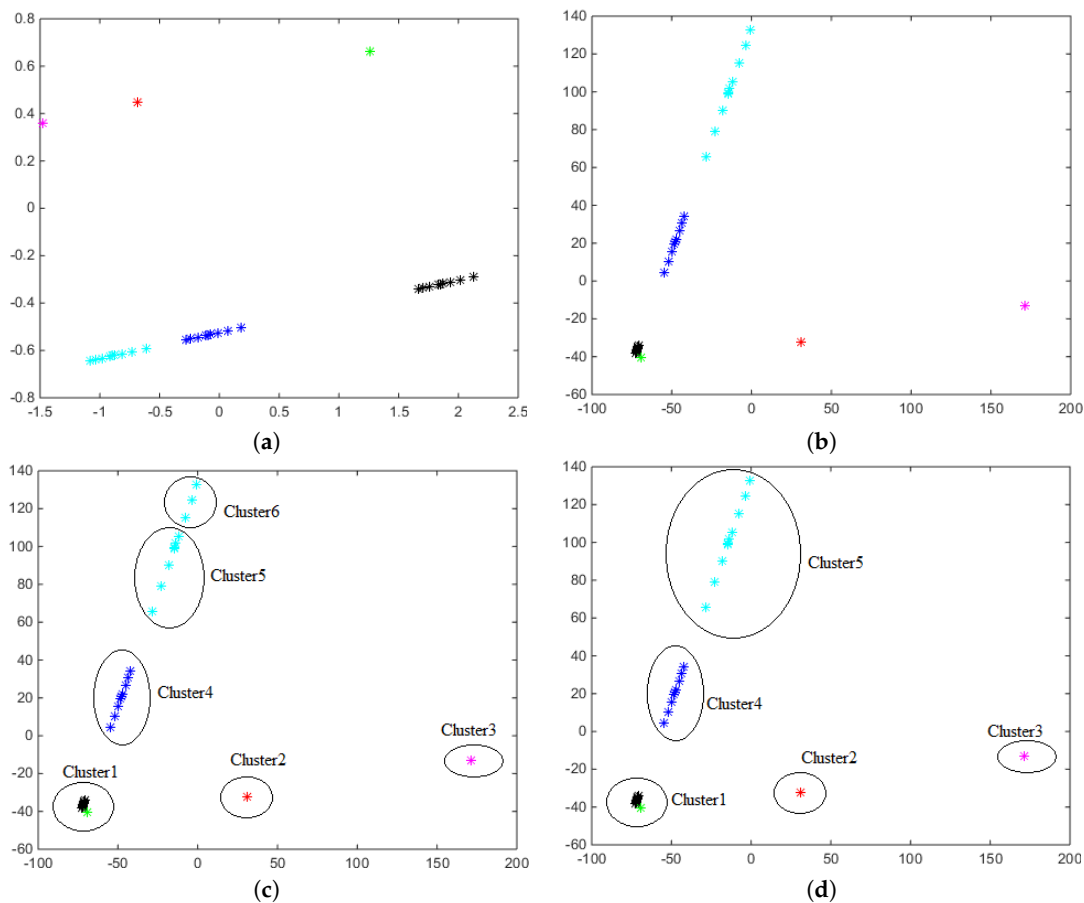

Figure 6. MDS applied to the distance matrices: (**a**) using d_{new}, (**b**) using d_2. The ten initial spirals are marked with green asterisks and the ten initial circumferences are represented with black asterisks, the enlarged spirals are plotted in red and the enlarged circumferences are plotted in blue.

As a third step of the simulation study, let us consider a broader data set with the initial spirals $\{\beta_{i1}\}_{i=1,\cdots,10}$, the circumferences $\{\beta_{i2}\}_{i=1,\cdots,10}$, their scaled versions $\{50\beta_{i1}\}_{i=1,\cdots,10}$ and $\{50\beta_{i2}\}_{i=1,\cdots,10}$, jointly with two new re-scaled sets $\{250\beta_{i1}\}_{i=1,\cdots,10}$ and $\{250\beta_{i2}\}_{i=1,\cdots,10}$. The distance matrices D_{new} and D_2 between the 60 curves are computed and the MDS scaling representation of these distance matrices can be found in Figure 7.

A *k*-means cluster analysis with *k* = 6 so as the DBSCAN algorithm applied on D_{new} recovers the 6 simulated groups. However, the DBSCAN algorithm applied on D_2 provides 5 clusters on this data set joining in a single cluster the initial (short) curves (spirals and circumferences) and distinguishing the other groups (Figure 7d). The *k*-means algorithm with *k* = 5 provides the same result, but if the

k-means algorithm is applied with $k = 6$ clusters, the set of the largest circumferences is split into two groups (Figure 7c). Once again, it can be clearly seen that the distance d_2 among shapes increases with the scaling factor.

Figure 7. MDS applied to the distance matrices: (**a**) using d_{new}, (**b**) using d_2. The ten initial spirals are marked in green and the ten initial circumferences are represented with black asterisks. The enlarged spirals are plotted in red (factor 50) and magenta (factor 250) and the enlarged circumferences are plotted in blue (factor 50) and cyan (factor 250). The clusters obtained on d_2 are plotted: (**c**) using k-means algorithm with $k = 6$, (**d**) using DBSCAN and k-means, $k = 5$.

7. Detection of Outliers

Although there are a variety of techniques for outlier detection for different types of data in any metric space based on nearest-neighbor techniques (see [16] for a detailed explanation), they have not been fully exploited in the shape and size space of curves. Some of the main references are based on box-plots of the distances to the median to detect outliers, such as [12,17], and more recently the method based on elastic depths proposed by [18].

We propose a technique for outlier detection based on the proposed distance. Nearest-neighbor techniques are very popular due to their good results, conceptual simplicity and interpretability in the classic multivariate case [19]. We consider this idea for the shape and size space of curves. The k-NN Anomaly Detection algorithm searches for the nearest k-neighbors, i.e., the k closest curves, for every element in the database, and calculates the average distance of the k-neighbors. In the multivariate case, the Euclidean distance is used, but here we use the proposed distance to find the neighbors. This procedure returns outlier scores; as usual, the highest score denotes the highest degree of outlierness. A way to establish a binary decision about whether or not to label a point as an outlier is

to use a box-plot with the outlier scores and to consider the points detected as outliers by the box-plot as anomalies.

We compare our procedure with that introduced in [12,17,18] using the data sets of open curves used in [12,17], which are available from [20]. For the Example 1 considered in [17] formed by 70 spirals, Ref. [12] found 6 outliers, Ref. [17] also found 6 outliers (2 scale outliers and 4 mild shape outliers) and [18] found 4 outliers (3 due to amplitude and 1 due to phase) with the recommended value of $k = 2$, which is the boxplot multiplier, while 9 outliers are found with the classical $k = 1.5$. However, with our methodology, we detect 8 outliers (the results are stable, we obtain the same outliers with $k = 5$, 10 or 15). We have also computed the square of the distance in Equation (17) $(d(p,q)^2)$ and we have computed the contribution in percentage due to shape $d_{\mathcal{C}0}^2(\pi(p),\pi(q))/d(p,q)^2$ and due to size $d_{\mathbb{R}}^2(\ln R(p), \ln R(q))/d(p,q)^2$, for each outlier. The percentages of contribution due to shape for the 8 outliers are: 21%, 29%, 34%, 35%, 41%, 50%, 50% and 74%. For the Example 3 in [17] formed by 176 fiber tracts in the human brain extracted from a diffusion tensor magnetic resonance image (DT-MRI), we detect the same 11 outliers also detected by [12,17], and all are due to shape, with percentages of contribution due to shape of 90%. However, [18] with $k = 2$ returned 62 outliers (62 due to amplitude, 23 of them are also outliers due to phase), i.e., 35% of points of the sample are considered outliers.

8. Application to a Real Data Set

Footwear design relies greatly on knowledge of foot size and shape. Proper fit is an essential condition for potential shoe buyers, besides the fact that poorly fitting footwear can cause foot pain and deformity, especially in women. Although people with extreme feet (very different from the rest) may be the most likely customers with poor fit, in anthropometric studies they are not usually searched. However, outliers report very valuable information for the footwear design process, since they can help shoe designers adjust their designs to a larger part of the population and can increase their awareness of customers characteristics that will make them uncomfortable to wear, whether when considering a range of special sizes or modifying any shoe feature to fit more users.

The aim of this section was to detect the outliers in an anthropometric foot database. We carry out a separate analysis for men and women, since gender foot shape differences are well-known [21,22]. Furthermore, footwear designers usually propose different types of shoes for women and men.

8.1. Foot Database

A total of 770 3D right foot scans were carried out. A total of 389 men and 381 women representing the Spanish adult female and male population were measured. The data were collected in different regions across Spain at shoe shops and workplaces using an INFOOT laser scanner [23]. The scanning process is carried out while the participant stands upright placing equal weight on each foot, in a specific position and orientation (see Figure 8). The result is a 3D point cloud representing the complete outer surface of the foot, including the sole of the foot.

3D foot shapes were registered using the method described by [24] with a template made up of 5000 vertices, with five foot landmarks (i.e., 1st and 2nd toe tips; 1st and 5th metatarsal heads; and pternion; see Figure 9). This set of landmarks is automatically located on 3D foot scans and allows the extraction of foot measurements and contours, according to the definitions used by the Human Shape Lab of the Biomechanics Institute of Valencia (IBV), which comply with standards and are compatible with the accepted definitions found in the literature [25–28]. In particular, we consider the longitudinal contour passing through the Ball Position. The mean shapes for men and women in $\mathcal{S}_2$ are displayed in Figure 10.
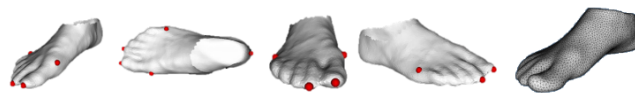
Figure 8. Infoot® scanner.

Figure 9. Foot landmarks used for registration in the database and foot template topology (the last image).

Figure 10. Mean shapes of contours for men (**left**) and women (**right**).

8.2. Detection of Foot Outliers

We have applied our outlier procedure to the curves of men and women with $k = 10$. A total of 24 and 18 outliers are detected for men and women, respectively. In order to briefly describe the outlier curves detected, we show the percentiles of each outlier for the four variables that could most influence shoe fit according to shoe design experts. Specifically, these variables are: Foot Length, FL (distance between the rear and foremost point the foot axis); Ball Girth, BG (perimeter of the ball section); Ball Width, BW (maximal distance between the extreme points of the ball section projected onto the ground plane); and Instep Height, IH (maximal height of the instep section, located at 50% of the foot length). Tables 1 and 2 show the percentile profiles of the outliers found for women and men, respectively. Note that for some of the outliers some of the variables show extreme percentiles, i.e., very high or very low percentiles. However, in many other cases, outliers do not show extreme values in these variables. Therefore, outliers can be due to the particular combination of the variables or due to the particular configuration of the curve that cannot be summarized by these four variables. In summary, with the proposed procedure, we can detect feet that are "not normal", which may not be detected with a classic multivariate analysis. Figure 11 shows the most outlier feet for men and women. For men, the outlier feet are the 14th and 9th, while for the women the outliers are the 7th and 17th. Note that those feet do not have really extreme percentiles.

We have also computed the contributions of shape and size. The main contribution is due to shape for both men and women. The mean is 96% for men and 97% for women.

Table 1. Percentile profiles of outliers of foot shape variables for women.

FL	BG	BW	IH
64	44	70	84
36	90	61	65
25	39	17	24
61	95	100	100
66	57	83	70
28	35	73	53
70	58	100	99
66	49	76	46
63	47	79	88
69	88	67	94
48	50	84	64
79	66	23	22
42	54	45	11
73	40	68	35
100	99	68	98
43	4	41	23
30	31	32	13
57	17	35	77

Table 2. Percentile profiles of outliers of foot shape variables for men.

FL	BG	BW	IH
69	51	45	53
65	49	38	77
25	39	17	24
75	17	24	14
37	25	60	37
86	74	2	13
47	85	15	21
58	32	3	11
43	59	11	1
72	39	31	48
10	2	8	4
99	78	43	43
7	30	30	82
56	30	12	2
82	39	40	28
20	7	11	54
96	58	53	28
42	54	44	11
53	80	91	100
53	58	28	49
88	72	38	84
3	3	86	79
19	54	84	97
13	37	99	91
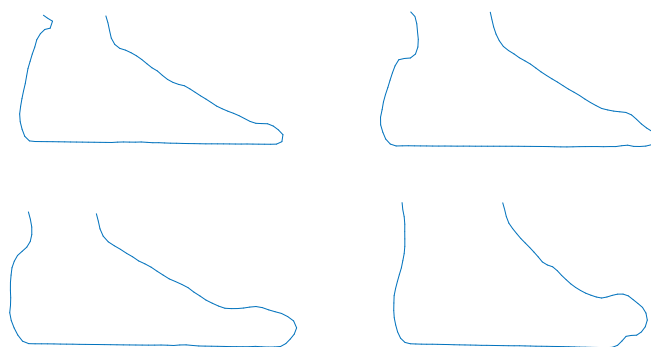
Figure 11. The most outlier feet for men (**first row**) and women (**second row**).

9. Conclusions

We have proposed a new metric in the shape and size space $\mathcal{S}_2$ that, unlike the previous proposals, allows us to distinguish whether the distance between two shapes $[q_1]$ and $[q_2]$ is due to the difference in shape or to the difference in size between the corresponding curves β_1 and β_2. It has been compared with the metric proposed by [5] in a simulation study, where our proposal is shown to perform better. Furthermore, we also show the advantages of the new metric, such as its invariance to changes of scale.

For the first time, we have also proposed a procedure based on the distances and NN techniques in $\mathcal{S}_2$ for finding outlier curves in $\mathcal{S}_2$. We have applied it to a novel industrial data set. The foot outliers found by considering their contour can help shoe designers improve their designs in order to provide customers with a better fit.

In future work, in regards to the theory, closed curves could be considered and an the appropriate metric defined. Furthermore, the new metric could be used in other types of statistical problems besides outlier detection, such as classification, clustering, or new ones, where curves in $\mathcal{S}_2$ have never been used before, such as archetype analysis [29] or archetypoid analysis [30]. Finally, in regards to the footwear application, the outlier procedure with the new metric could be applied to other kinds of foot contours, such as the Ball Girth, and of course, scopes for other fields of application.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dryden, I.; Mardia, K. Size and shape analysis of landmark data. *Biometrika* **1992**, *79*, 57–68. [CrossRef]
2. Klassen, E.; Srivastava, A.; Mio, M.; Joshi, S.H. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 372–383. [CrossRef] [PubMed]
3. Younes, L.; Michor, P.; Shah, J.; Mumford, D. A Metric on Shape Space With Explicit Geodesics. *Rend. Lincei Mat. Appl.* **2008**, *19*, 25–57. [CrossRef]
4. Srivastava, A.; Klassen, E.; Joshi, S.H.; Jermyn, I.H. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1415–1428. [CrossRef] [PubMed]
5. Kurtek, S.; Srivastava, A.; Klassen, E.; Ding, Z. Statistical Modeling of Curves Using Shapes and Related Features. *J. Am. Stat. Assoc.* **2012**, *107*, 1152–1165. [CrossRef]
6. Sundaramoorthi, G.; Mennucci, A.; Soatto, S.; Yezzi, A. A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering. *SIAM J. Imaging Sci.* **2011**, *4*, 109–145.

7. Kouchi, M. 3—Anthropometric methods for apparel design: Body measurement devices and techniques. In *Anthropometry, Apparel Sizing and Design*; Gupta, D., Zakaria, N., Eds.; Woodhead Publishing: Cambridge, UK, 2014; pp. 67–94.

8. Kuehnapfel, A.; Ahnert, P.; Loeffler, M.; Broda, A.; Scholz, M. Reliability of 3D laser-based anthropometry and comparison with classical anthropometry. *Sci. Rep.* **2016**, *6*, 26672. [CrossRef] [PubMed]

9. Srivastava, A.; Klassen, E.P. *Functional and Shape Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016.

10. Pennec, X. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *J. Math. Imaging Vis.* **2006**, *25*, 127–154. [CrossRef]

11. Le, H. Locating Fréchet means with application to shape spaces. *Adv. Appl. Probab.* **2001**, *33*, 324–338. [CrossRef]

12. Kurtek, S.; Su, J.; Grimm, C.; Vaughan, M.; Sowell, R.; Srivastava, A. Statistical analysis of manual segmentations of structures in medical images. *Comput. Vis. Image Underst.* **2013**, *117*, 1036–1050. [CrossRef]

13. Cox, T.F.; Cox, M.A. *Multidimensional Scaling*; CRC Press: Boca Raton, FL, USA, 2000.

14. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Kdd, Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.

15. Hahsler, M.; Piekenbrock, M.; Doran, D. dbscan: Fast density-based clustering with R. *J. Stat. Softw.* **2019**, *91*, 1–30. [CrossRef]

16. Aggarwal, C.C. *Outlier Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2017.

17. Xie, W.; Chkrebtii, O.; Kurtek, S. Visualization and Outlier Detection for Multivariate Elastic Curve Data. *IEEE Trans. Vis. Comput. Graph.* **2019**. [CrossRef]

18. Harris, T.; Tucker, J.D.; Li, B.; Shand, L. Elastic depths for detecting shape anomalies in functional data. *Technometrics* **2020**, 1–25. [CrossRef]

19. Goldstein, M.; Uchida, S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLoS ONE* **2016**, *11*, e0152173. [CrossRef] [PubMed]

20. Cho, M.H.; Asiaee, A.; Kurtek, S. Elastic Statistical Shape Analysis of Biological Structures with Case Studies: A Tutorial. *Bull. Math. Biol.* **2019**, *81*, 2052–2073. [CrossRef] [PubMed]

21. Krauss, I.; Langbein, C.; Horstmann, T.; Grau, S. Sex-related differences in foot shape of adult Caucasians—A follow-up study focusing on long and short feet. *Ergonomics* **2011**, *54*, 294–300. [CrossRef] [PubMed]

22. Saghazadeh, M.; Kitano, N.; Okura, T. Gender differences of foot characteristics in older Japanese adults using a 3D foot scanner. *J. Foot Ankle Res.* **2015**, *8*, 29. [CrossRef] [PubMed]

23. I-Ware Laboratory. Available online: http://www.i-ware.co.jp/ (accessed on 24 August 2020).

24. Allen, B.; Curless, B.; Popović, Z. The space of human body shapes: Reconstruction and parameterization from range scans. *ACM Trans. Graph. (TOG)* **2003**, *22*, 587–594. [CrossRef]

25. Rossi, W.A.; Tennant, R. *Professional Shoe Fitting*; National Shoe Retailers Association: Bicester, UK, 2013.

26. Ramiro, J.; Alcántara, E.; Forner, A.; Ferrandis, R.; García-Belenguer, A.; Durá, J.; Vera, P.; Brizuela, G.; Llana, S. *Guía de Recomendaciones para el Diseño de Calzado*; Instituto de Biomecánica de Valencia: Valencia, Spain, 1995; pp. 135–151.

27. Goonetilleke, R.S. *The Science of Footwear*; CRC Press: Boca Raton, FL, USA, 2012.

28. Luximon, A. *Handbook of Footwear Design and Manufacture*; Elsevier: Amsterdam, The Netherlands, 2013.

29. Cutler, A.; Breiman, L. Archetypal Analysis. *Technometrics* **1994**, *36*, 338–347. [CrossRef]

30. Vinué, G.; Epifanio, I.; Alemany, S. Archetypoids: A new approach to define representative archetypal data. *Comput. Stat. Data Anal.* **2015**, *87*, 102–115.