Finsler Metrics Induced by a Similarity Function

Nisachon Kumankat[1]*, Praiboon Pantaragphong[1] and Sorin V. Sabau[2]

[1] Department of Mathematics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2] School of Biological Science, Department of Biology, Tokai University, Sapporo, Japan

Received: 2 December 2019, Revised: 21 March 2020, Accepted: 2 April 2020

Abstract

In the present paper, the geometrical properties of a topological space endowed with a similarity was studied. Its relation with weighted quasi-metrics and Finsler metrics of Randers type was discussed. Finally, some applications to bioinformatics and computer science by relating similarities to dynamic programming algorithms are considered. In conclusion, the space containing the real-world data is non-symmetric and non-linear.

Keywords: Finsler metrics, similarity function, weighted quasi-metrics
DOI 10.14456/cast.2020.20

1. Introduction

Metric spaces, symmetric distances are used in different fields of pure mathematics like analysis, geometry and so on, as well as in applied mathematics, for instance in computer science, bioinformatics, data analysis, etc. These are the natural generalization of Euclidean and Riemannian spaces. On the other hand, non-symmetric distances, Minkowski norms, and Finsler metrics are also widely used in the analysis, differential geometry, data analysis, etc. We claim that symmetric distances, Euclidean and Riemannian metrics are just convenient approximations (usually obtained by averaging) of the real world. The real world, based on real data measurements is highly non-symmetric and non-linear. Of course, proving such a fact in its most generality is a very complex and difficult task, beyond the purpose of this paper.

However, we will argue that the similarity induced by the dynamic programming algorithm Needleman-Wunsch is actually equivalent in nature to non-symmetric distances (so-called quasi-distances) and Finsler metrics. Our main statement in this paper is that the following motions are equivalent in nature, i.e. symmetric similarity function, weighted quasi-distance and Finsler metrics of Randers type with reversible geodesics.

*Corresponding author: Tel.: +66 91 8517097
E-mail: Nisachon.Kumankat@gmail.com

The topic studied in the present paper is very important for analysis, geometry, computer science, data analysis, bioinformatics and so on because in some sense it shows that at least partially, the reality we are living in, is non-symmetric, non-linear, non-homogeneous and the study of data sets from the real world is actually equivalent to the use of weighted quasi-metrics on topological spaces, or of a Finsler metric of Randers type on smooth manifolds.

2. Similarities and Distances

We start by recalling the following definition [1, 2, 3].

Definition 2.1 Let X be a topological space. If $s : X \times X \to \mathbb{R}$ is a continuous mapping such that

i. $s(x,x) > 0$ for any $x \in X$,
ii. $s(x,x) \geq s(x,y)$ for any $x, y \in X$,
iii. if $s(x,y) = s(x,x)$ and $s(y,x) = s(y,y)$, then $x = y$ for any $x, y \in X$,
iv. $s(x,y) + s(y,z) \leq s(x,z) + s(y,y)$ for any $, y, z \in X$,

Then s is called a *similarity function* on X.

The relation with quasi-metrics is well-known [4, 5].

Definition 2.2 Let X be a non-empty set and d a real-valued function $d : X \times X \to [0,\infty)$ that satisfies:

i. $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$ for any $x, y \in X$,
ii. $d(x,y) \leq d(x,z) + d(z,y)$ for any $x, y, z \in X$,
iii. if $d(x,y) = d(y,x) = 0$ then $x = y$ for any $x, y \in X$.

Then (X,d) is called a *quasi-metric space*.

Definition 2.3 A *weighted quasi-metric space* is a triple (X,d,w), where X is a non-emply set, $d : X \times X \to [0,\infty)$ and $w : X \to [0,\infty)$ that satisfies:

i. $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$ for any $x, y \in X$,
ii. $d(x,y) \leq d(x,z) + d(z,y)$ for any $x, y, z \in X$,
iii. if $d(x,y) = d(y,x) = 0$ then $x = y$ for any $x, y \in X$,
iv. $d(x,y) + w(x) = d(y,x) + w(y)$ for any $x, y, z \in X$.

The function d is called a *quasi-metric*, and w is the *weight function*.

Proposition 2.4 *If* $s : X \times X \to \mathbb{R}$ *be a similarity function on* X, *then* $d : X \times X \to \mathbb{R}$ *defined by*

$$d(x,y) := s(y,y) - s(y,x), \qquad \textit{for all } x, y \in X,$$

is a quasi-metric on X.

Proof Let $x, y, z \in X$. We verify the conditions in the definition of the quasi-metric.

i. Positiveness: $d(x,y) = s(y,y) - s(y,x)$,
Since $s(y,y) - s(y,x) \geq 0$, it is clear that $d(x,y) \geq 0$.
And $d(y,y) = s(y,y) - s(y,y) = 0$, then $x = y$.

ii. Triangle inequality:

$$\begin{aligned} d(x,y) &= s(y,y) - s(y,x) \\ &= s(y,y) - s(y,z) + s(y,z) - s(y,x) \\ &\leq s(y,y) - s(y,z) + s(z,z) - s(z,x) \\ &= d(z,y) + d(x,z) \\ &= d(x,z) + d(z,y). \end{aligned}$$

iii. Separation axiom:

$$d(x,y) = 0 \ \wedge \ d(y,x) = 0 \ \Longrightarrow x = y$$
$$s(y,y) - s(y,x) = 0 \ \wedge \ s(x,x) - s(x,y) = 0 \ \Longrightarrow x = y$$
$$s(y,y) = s(y,x) \ \wedge \ s(x,x) = s(x,y) \ \Longrightarrow x = y.$$

Therefore, (X,d) is a quasi-metric on X.

Proposition 2.5 *Let* $s : X \times X \to \mathbb{R}$ *is a similarity function on* X. *If* s *is a symmetric, i.e.* $s(x,y) = s(y,x)$ *for all* $x,y \in X$, *then* (X,d) *is a weighted quasi-metric space with weight function* $w : X \to \mathbb{R}$, $w(x) = s(x,x)$.

<u>Proof</u> By proposition 2.4, we have $d(x,y) = s(y,y) - s(y,x)$.
Let $x,y,z \in X$. We verify the conditions in the definition of a weighted quasi-metric.

i. Positiveness: $d(x,y) = s(y,y) - s(y,x)$,
since $s(y,y) - s(y,x) \geq 0$, it is clear that $d(x,y) \geq 0$.
And $d(y,y) = s(y,y) - s(y,y) = 0$, then $x = y$.

ii. Triangle inequality:

$$\begin{aligned} d(x,y) &= s(y,y) - s(y,x) \\ &= s(y,y) - s(y,z) + s(y,z) - s(y,x) \\ &\leq s(y,y) - s(y,z) + s(z,z) - s(z,x) \\ &= d(x,z) + d(z,y). \end{aligned}$$

iii. Separation axiom:

$$d(x,y) = 0 \ \wedge \ d(y,x) = 0 \ \Longrightarrow x = y$$
$$s(y,y) - s(y,x) = 0 \ \wedge \ s(x,x) - s(x,y) = 0 \ \Longrightarrow x = y$$
$$s(y,y) = s(y,x) \ \wedge \ s(x,x) = s(x,y) \ \Longrightarrow x = y.$$

iv. Let $w : X \to \mathbb{R}$, $w(x) = s(x,x)$, we have

$$\begin{aligned} d(x,y) + w(x) &= s(y,y) - s(y,x) + s(x,x) \\ &= s(x,x) - s(x,y) + s(y,y) \\ &= d(y,x) + w(y). \end{aligned}$$

Therefore, (X,d) is a weighted quasi-metric space.

We will consider in the following only symmetric similarity function.

Remark 2.6 :

1) Observe that the quasi-distance d and the weight function w are determined only by the similarity function s.

2) Let we assume the quasi-distance d is actually a distance function, i.e. $d(x,y) = d(y,x)$ for all $x, y \in X$, it follows

$$s(y,y) - s(y,x) = s(x,x) - s(x,y),$$

and if we take into account that s is symmetric, then we obtain

$$s(x,x) = s(y,y), \qquad \text{for all } x, y \in X.$$

In other words, the quasi-distance induced by a similarity function is a distance if and only if the similarity is the same on the diagonal.

3) The symmetrized distance induced by a quasi-distance d is

$$\rho(x,y) := \frac{1}{2}[d(x,y) + d(y,x)] = \frac{1}{2}[s(x,x) + s(y,y)] - s(x,y).$$

Conversely, a weighted quasi-metric space induces a symmetric similarity function. Indeed, we have the following proposition hold

Proposition 2.7 *Let (X,d) is a weighted quasi-metric space with the weight function $w : X \to \mathbb{R}$. Then the mapping $s : X \times X \to \mathbb{R}$,*

$$s(x,y) := w(x) - d(y,x), \qquad \text{for all } x, y \in X,$$

is a symmetric similarity function on X.

Proof Let $x, y, z \in X$. We verify the conditions in definition 2.1, we have:

i. Since $w : X \to \mathbb{R}$ is weight function, then $w(x) > 0$ and $d(x,x) = 0$. It is clear that $s(x,x) = w(x) - d(x,x) = w(x) > 0$.

ii. We will show that $s(x,x) - s(x,y) \geq 0$, we have

$$\begin{aligned} s(x,x) - s(x,y) &= w(x) - d(x,x) - w(x) + d(y,x) \\ &= -d(x,x) + d(y,x) \\ &= d(y,x). \end{aligned}$$

Since $d(y,x) \geq 0$, thus $s(x,x) - s(x,y) \geq 0$, i.e. $s(x,x) \geq s(x,y)$.

iii. Suppose that $s(x,y) = s(x,x)$ and $s(y,x) = s(y,y)$, we have

$$\begin{aligned} s(x,y) = s(x,x) \ &\wedge\ s(y,x) = s(y,y) \implies x = y \\ w(x) - d(y,x) = w(x) - d(x,x) \ &\wedge\ w(y) - d(x,y) = w(y) - d(y,y) \implies x = y \\ d(x,x) - d(y,x) = w(x) - w(x) \ &\wedge\ d(y,y) - d(x,y) = w(y) - w(y) \implies x = y \\ d(x,x) - d(y,x) = 0 \ &\wedge\ d(y,y) - d(x,y) = 0 \implies x = y \\ d(x,x) = d(y,x) \ &\wedge\ d(y,y) = d(x,y) \implies x = y. \end{aligned}$$

iv. We will show that $s(x,y) + s(y,z) \leq s(x,z) + s(y,y)$, we have

$$\begin{aligned} s(x,y) + s(y,z) &= w(x) - d(y,x) + w(y) - d(z,y) \\ &= w(x) + w(y) - d(z,y) - d(y,x) \\ &\leq w(x) + w(y) - d(z,x) \\ &= w(x) - d(z,x) + w(y) - 0 \\ &= w(x) - d(z,x) + w(y) - d(y,y) \\ &= s(x,z) + s(y,y). \end{aligned}$$

Therefore, (X,s) is a symmetric similarity function on X.

Example 2.8 Let us consider the metric space (S,ρ) and the interval $I := (0,\infty)$. It is known that the product space $G := S \times I$ inherits a natural structure of generalized weighted quasi-metric structure (G,Q,W), where

$$Q : G \times G \to I, \quad Q(u,v) := \rho(x,y) + \eta - \xi,$$

$$W : G \to I, \qquad W(u) := 2\xi,$$

for any $u = (x, \xi),\ v = (y, \eta)$ on $G = S \times I$.

The similarity function $\varphi : G \times G \to \mathbb{R}$ induced by the weighted quasi-metric structure (G, Q, W) is given by

$$\varphi(u, v) := -\rho(x, y) + \xi + \eta, \qquad \text{for any } u = (x, \xi),\ v = (y, \eta) \in G.$$

Clearly this is a symmetric similarity function on G.

Indeed, let $u = (x, \xi),\ v = (y, \eta),\ l = (z, \zeta) \in G$, then

i. It is clear that

$$\varphi(u, u) = \varphi\big((x, \xi), (x, \xi)\big) = -\rho(x, x) + \xi + \xi = 2\xi > 0.$$

ii. We will show that $\varphi(u, u) - \varphi(u, v) \geq 0$, we have

$$\begin{aligned}\varphi(u, u) - \varphi(u, v) &= -\rho(x, x) + \xi + \xi + \rho(x, y) - \xi - \eta \\ &= -\rho(x, x) + \rho(x, y) + \xi - \eta \\ &= \rho(x, y) + \xi - \eta \\ &= \rho(y, x) + \xi - \eta \\ &= Q\big((y, \eta), (x, \xi)\big) = Q(v, u) \quad \geq 0.\end{aligned}$$

So, $\varphi(u, u) - \varphi(u, v) \geq 0$.

iii. Suppose that $\varphi(u, v) = \varphi(u, u)$ and $\varphi(v, u) = \varphi(v, v)$, that is

$$\begin{aligned}\varphi\big((x, \xi), (y, \eta)\big) = \varphi\big((x, \xi), (x, \xi)\big) \ &\wedge\ \varphi\big((y, \eta), (x, \xi)\big) = \varphi\big((y, \eta), (y, \eta)\big) \\ -\rho(x, y) + \xi + \eta = -\rho(x, x) + \xi + \xi \ &\wedge\ -\rho(y, x) + \eta + \xi = -\rho(y, y) + \eta + \eta \\ \rho(x, x) - \rho(x, y) + \eta - \xi = 0 \ &\wedge\ \rho(y, y) - \rho(y, x) + \xi - \eta = 0,\end{aligned}$$

by subtracting these two equalities we get, $\rho(x, x) - 2\xi = \rho(y, y) - 2\eta$, whence $x = y$, so that $\xi = \eta$. Hence $u = v$.

iv. We will show that $\varphi(u, v) + \varphi(v, l) \leq \varphi(u, l) + \varphi(v, v)$, we have

$$\begin{aligned}\varphi(u, v) + \varphi(v, l) &= \varphi\big((x, \xi), (y, \eta)\big) + \varphi\big((y, \eta), (z, \zeta)\big) \\ &= -\rho(x, y) + \xi + \eta - \rho(y, z) + \eta + \zeta \\ &= -\rho(x, y) - \rho(y, z) + \xi + \zeta + \eta + \eta \\ &\leq -\rho(x, z) + \xi + \zeta + \eta + \eta \\ &= -\rho(x, z) + \xi + \zeta - \rho(y, y) + \eta + \eta \\ &= \varphi\big((x, \xi), (z, \zeta)\big) + \varphi\big((y, \eta), (y, \eta)\big) \\ &= \varphi(u, l) + \varphi(v, v).\end{aligned}$$

Therefore, φ is a symmetric similarity function on G.

Example 2.9 Let (S, ρ) be a metric space and $f : S \to (0, \infty)$ a Lipschitz function with respect to ρ. Then it is known that the graph of f, i.e. $G_f = \{(x, f(x)) : x \in S\}$ has a weighted quasi-metric space structure (G_f, Q, W) given by

$$Q : G_f \times G_f \to (0, \infty), \quad Q(u, v) := \rho(x, y) + f(y) - f(x),$$
$$W : G_f \to (0, \infty), \qquad W(u) := 2f(x),$$

for any $u = (x, f(x)),\ v = (y, f(y))$ on G_f.

It follows that the function $\varphi_f : G_f \times G_f \to \mathbb{R}$ given by

$$\varphi_f(u, v) := -\rho(x, y) + f(x) + f(y),$$

for any $u = (x, f(x)),\ v = (y, f(y)) \in G_f$, is a symmetric similarity function on G_f.

We can conclude that a metric space (X, ρ) with a Lipschitz function $f : X \to \mathbb{R}$ induces a similarity function on X.

The similarity space $\left(G_f, \varphi_f\right)$ constructed here is called the bundle over the metric space (S, ρ).

3. Embeddings and relation to Finsler space

Let (X, σ) and (Y, τ) be two topoltgical spaces with similarities functions σ and τ, respectively. A continuous mapping $\varphi : X \to Y$ is called a *similarity embedding* if

$$\tau\big(\varphi(x), \varphi(y)\big) = \sigma(x, y),$$

for all $x, y \in X$.

Proposition 3.1 *Let (X, q, w) and (Y, p, u) be two weighted quasi-metric spaces with the associated similarities σ and τ, respectively. The continuous function $\varphi : X \to Y$ is a similarity embedding if and only if it is an embedding of weighted quasi-metric spaces.*

Proof We assume that $\varphi : (X, \sigma) \to (Y, \tau)$ is a similarity embedding, i.e.

$$\tau\big(\varphi(x), \varphi(y)\big) = \sigma(x, y), \qquad \forall x, y \in X.$$

The weighted quasi-metric (d, w) associated with the similarity function σ on X is given by

$$\begin{aligned} d(x, y) &= \sigma(y, y) - \sigma(y, x), && \forall x, y \in X, \\ w(x) &= \sigma(x, x), && \forall x \in X. \end{aligned}$$

The weighted quasi-metric $\big(\hat{d}, \widehat{w}\big)$ associated with the similarity function τ on Y is given by

$$\begin{aligned} \hat{d}(x, y) &= \tau(y, y) - \tau(y, x), && \forall x, y \in Y, \\ \widehat{w}(x) &= \tau(x, x), && \forall x \in Y. \end{aligned}$$

We compute

$$\begin{aligned} \hat{d}\big(\varphi(x), \varphi(y)\big) &= \tau\big(\varphi(y), \varphi(y)\big) - \tau\big(\varphi(y), \varphi(x)\big) \\ &= \sigma(y, y) - \sigma(y, x) \\ &= d(x, y), \end{aligned}$$

for all $x, y \in X$. Likewise,

$$\widehat{w}\big(\varphi(x)\big) = \tau\big(\varphi(x), \varphi(x)\big) = \sigma(x, x) = w(x),$$

and hence it results that $\varphi : (X, d, w) \to \big(Y, \hat{d}, \widehat{w}\big)$ is an embedding of weighted quasi-metrics spaces.

Conversely, we assume that $\varphi : (X, d, w) \to \big(Y, \hat{d}, \widehat{w}\big)$ is an embedding of weighted quasi-metrics spaces, i.e.

$$\begin{aligned} \hat{d}\big(\varphi(x), \varphi(y)\big) &= d(x, y), \\ \widehat{w}\big(\varphi(x)\big) &= w(x), \end{aligned}$$

for all $x, y \in X$. Using now relations

$$\begin{aligned} \sigma(x, y) &= w(x) - d(y, x), && \forall x, y \in X, \\ \tau(x, y) &= \widehat{w}(x) - \hat{d}(y, x), && \forall x, y \in Y. \end{aligned}$$

We have

$$\begin{aligned} \tau\big(\varphi(x), \varphi(y)\big) &= \widehat{w}\big(\varphi(x)\big) - \hat{d}\big(\varphi(y), \varphi(x)\big) \\ &= w(x) - d(y, x) \\ &= \sigma(x, y), \end{aligned}$$

for all $x, y \in X$, therefore $\varphi : (X, \sigma) \to (Y, \tau)$ is a similarity embedding.

Theorem 3.2 *Every space with a symmetry function (X, s) is embeddable in a bundle over a suitable metric space (S, ρ).*
Proof The proof is quite straightforward by taking into account the construction in Example 2.9.
This result can also be proved directly, using the fact that any weighted quasi-metric space is embeddable in a bundle over a suitable metric space [6, 7].

Theorem 3.3 *1. Let (S, ρ) be a metric space and $f : S \to [0, \infty)$ a Lipschitz function on this metric space. Then the graph of f admits a similarity function $\varphi : G_f \times G_f \to \mathbb{R}$ that depends on ρ and f only.*

2. Conversely, every similarity space (X, s) can be constructed in this way.

Proof (1) Statement 1 follows immediately from Example 2.9.

(2) Conversely, if we start with a similarity space (X, s), then we can consider :

(a) the associated weighted quasi-metric space (X, d, w), where d and w are given in Proposition 2.4 and 2.5.

(b) the symmetrized associated distance s has given in Remark 2.6.

By putting $f := \frac{1}{2}w$, using (X, s) and f, the construction from Statement 1 given that (G_f, φ) is a similarity space.
Moreover, (X, s) can be embedded in (G_f, φ) and the conclusion follows.

Lemma 3.4 *Let M be a compact smooth manifold. If (M, ρ) is an upper and lower curvature bounded metric space, then there exists a Riemannian metric g on M whose distance function coincides with p.*
Proof The proof is quite obvious. Every an upper and lower curvature bounded metric space (M, ρ) constructed on a n -dimensional compact smooth manifold M can be embedded isometrically in the Euclidean space $\mathbb{R}^{2n+1}$ with the canonical metric. On the other hand, the manifold M as a submanifold in $\mathbb{R}^{2n+1}$ inherits a canonical Riemannian metric [8, 9] from the embedding in the Euclidean space whose distance function obviously coincides with ρ.

Theorem 3.5 *1. If $(M, F = \alpha + \beta)$ is a simply connected Randers defined by a Riemannian metric $\alpha = \sqrt{a_{ij}(x)y^i y^j}$ and a closed 1-form β, then M is endowed with a naturally induced similarity function.*

2. Conversely, let s be a symmetric, similarity function defined on a compact differentiable manifold M whose associated distance p is upper and lower curvature bounded. If s is differentiable, then there exists a naturally constructed Randers metric on M that depends on s only.
Proof 1. It is clear since a Randers metric with β closed induces a weighted quasi metric on M.

2.Conversely, the similarity metric induces a weighted quasi metric (M, d, w). From Lemma 3.4 we can see that there exists a Riemannian metric on M whose distance function is exactly ρ, and since s was assumed smooth we can define $\beta := dw$, where w is the weight induced by s.

4. Relation with Bioinformatics and Computer Science

In order to assess the application of this theory, we start by recalling that Dynamic Programming is, at the same time, a mathematical optimization method as well as an algorithmic method in computer science. Dynamic Programming originates in the research of R. Bellman in the 1950s and it was applied eventually in many fields of science like engineering, economics and others. In the majority cases, this method works by simplifying a much more complicated problem by dividing it into much easier small problems using a recursive way. It is known that if a problem in computer science can be solved optimally by dividing it into smaller problems and then recursively determine the optimal solutions to these small problems, then the original problem has an optimal substructure. The algorithms involving Dynamic Programming are popular in the field of bioinformatics being extremely useful for some specific problems as DNA or amino acids sequences alignment, RNA structure prediction, protein structure research, and others [4, 5].

In the case of sequence comparison analysis in Bioinformatics, a similarity measure on Σ together with a gap penalties function can be used to define the global similarity between two sequences in Σ^*. The computation is handled using the Needleman-Wunsch dynamic programming algorithm which is quite similar to the W-S-B algorithm for computation of distances. It is possible to define global similarity using a dynamic programming matrix.

To be more precise, let Σ be a non-empty set. Then a *free monoid* Σ^*on Σ is the monoid whose elements are all *finite* sequences of zero or more elements, from Σ with the operation of *concatenation*. The set $\Sigma = \{A, B, C, \dots, Z\}$ is called *alphabet*, and its elements $A, B, C, \dots, Z$ are called **letters** of the alphabet, or *generators*. The elements $u \in \Sigma^*$ are called *words* or *strings*. The unique sequence of zero letters (the empty word) denoted by e is the *identity element* in Σ^*.

The *free semigroup* Σ^+on Σ is defined as $\Sigma^+ := \Sigma^* \setminus \{e\}$.

Remark 4.1 Biological motivation

The macromolecule that contains the essential information of living cells can be represented as a family of words over a finite alphabet. Consequently, DNA (or RNA) molecules can be seen as long words in the free semigroup with the generators $\Sigma = \{A, C, T, G\}$ (nucleotide alphabet). Proteins molecules can be regarded as words in the free semigroup whose generators are the 20 amino acids which compose the proteins in living cells (aminoacids alphabet) Σ_{AA}.

As an example, we mention here the insulin, whose intensive research, starting around 1950, has facilitated the development of the theory of molecular evolution of living organisms. Insulin is present in almost all living organisms on the Earth, hence by comparing the insulin sequences found in different species and computing their similarity, one can get a very detailed insight into the evolution of life on Earth. Sequence comparison, similarity estimation and so on, is one of the most fundamental research topics in bioinformatics [6].

We define global similarity using a dynamic programming matrix.

Definition 4.2 Let Σ be a set, $x, y \in \Sigma^*$, $s : \Sigma \times \Sigma \to \mathbb{R}$ and $g, h : \mathbb{N}^+ \to \mathbb{R}^+$. Let $x, y \in \Sigma^*$ and let $m = |x|$ and $n = |y|$. The *Needleman-Wunsch* dynamic programming matrix denoted $\mathrm{NW}(x, y, s, g, h)$, is an $(m+1) \times (n+1)$ matrix S with rows and columns indexed from 0 such that $S_{0,0} = 0$, $S_{i,0} = \max\limits_{1 \le k \le i} \{S_{i-k,0} - h(k)\}$, $S_{0,j} = \max\limits_{1 \le k \le j} \{S_{0,j-k} - g(k)\}$ and for all $i = 1,2, \dots, m$ and $j = 1,2, \dots, n$

$$S_{i,j} = \max\left\{S_{i-1,j-1} + s(x_i, y_j),\ \max_{1\le k\le i}\{S_{i-k,j} - h(k)\},\ \max_{1\le k\le j}\{S_{i,j-k} - g(k)\}\right\}.$$

We define the *global similarity* between the sequences x and y (given s, g and h), denoted $S(x, y)$, to be the value $S_{m,n}$.

The global similarity $S(x, y)$ between sequences x and y defined above satisfies all conditions in the definition of similarity [4].

Theorem 4.3 *Let* Π *be a finite set of biological sequences and let* $S(x, y)$ *be the global similarity function given by the Needleman-Wunsch dynamic programming algorithm. If the associated distance p is upper and lower curvature bounded, then there exists a metric of Randers type whose distance function coincides with the weighted quasi-metric induced by* $S(x, y)$.

Proof Based on our Theorem 3.5 we can explain as follows. Let us consider a finite set Π of biological sequences like, for instance, the insulin sequences in all species one can find in the NCBI database. Clearly, this is a finite set of sequences, a finite set of data that can be considered as a compact set in Σ^*. By using the dynamic programming, we can endow this compact set with a symmetric similarity function $S(x, y)$. Theorem 3.5 implies that there is always a Randers type metric whose associated distance function coincides with the weighted quasi distance function obtained from the similarity function.

5. Conclusions

In this paper, we introduce a definition of the similarity function, quasi-metric space and weighted quasi-metric space. We also study the geometrical properties of a topological space endowed with a similarity. The relation with embeddings and bundle and Finsler metrics of Randers type has been explained. Moreover, we present the relation of the mathematical concepts with computer science and bioinformatics. In conclusion, there is always a Randers type metric whose associated distance function coincides with the weighted quasi-metric induced by a similarity function.

References

[1] Sabau, S.V., Shibuya, K. and Shimada, H., 2014. Metric structures associated to Finsler metrics. *Publications Mathematicae Debrecen*, 84(1-2), 89-103.

[2] Shanker, G. and Rani, S., 2018. Weighted quasi-metrics associated with Finsler metrics. [online] Available at : https://www.researchgate.net/publication/322567912_Weighted_quasi-metrics_associated_with_Finsler_metrics

[3] Vitolo, P., 1999. The representation of weighted quasi-metric spaces. *Rendiconti dell'Istituto di Matematica dell'Università di Trieste*, 31, 95-100.

[4] Stojmirovic, A., 2008. *Quasi-metrics, Similarities and Searches: Aspects of Geometry of Protein Datasets*. Ph.D. Victoria University of Wellington.

[5] Stojmirovic, A., 2004. Quasi-metric space with measure. *Topology Proceedings*, 28 (2), 655-671.

[6] Pevsner, J., 2017. *Bioinformatics and Functional Genomics*. 3rd ed. New Delhi: Wiley India.

[7] Vitolo, P., 1995. A representation theorem for quasi-metric spaces. *Topology and Its Applications*, 65, 101-104.

[8] Bao, D., Chern, S.S. and Shen, Z., 2000. *An Introduction to Riemann-Finsler Geometry*. New York: Springer-Verlag.
[9] Chern, S.S. and Shen, Z., 2005. *Riemann-Finsler Geometry*. 6th ed. Singapore: World Scientific.