HAL
open science

Discrete Time Elastic Vector Spaces

Pierre-François Marteau

To cite this version:

Pierre-François Marteau. Discrete Time Elastic Vector Spaces. 2011. hal-00556007v3

HAL Id: hal-00556007

https://hal.science/hal-00556007v3

Preprint submitted on 14 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Discrete Time Elastic Vector Spaces

Pierre-Francois Marteau

VALORIA, Université de Bretagne Sud (Université Européenne de Bretagne), Campus de Tohannic, 56000 Vannes, France
pierre-francois.marteau@univ-ubs.fr

Abstract

This paper proposes a framework dedicated to the construction of what we call *time elastic inner products* allowing one to embed sets of non-uniformly sampled multivariate time series of varying lengths into vector space structures. This framework is based on a recursive definition that covers the case of multiple embedded *time elastic* dimensions. We prove that such inner products exist in our framework and show how a simple instance of this inner product class operates on some toy or prospective applications, while generalizing the Euclidean inner product.

Keywords: Vector Space, Discrete Time Series, Sequence mining, Non Uniform Sampling, Elastic Inner Product, Time Warping

1. Introduction

Time series analysis in metric spaces has attracted much attention over numerous decades and in various domains such as biology, statistics, sociology, networking, signal processing, etc, essentially due to the ubiquitous nature of time series, whether they are symbolic or numeric. Among other characterizing tools, time warp distances (see [1], [2], and more recently [3], [4] among other references) have shown some interesting robustness compared to the Euclidean metric especially when similarity searching in time series data bases is an issue. Unfortunately, this kind of elastic distance does not enable direct construction of definite kernels which are useful when addressing regression, classification or clustering of time series. A fortiori, they do not make it possible to directly construct inner products involving some *time elasticity*, which are namely able to cope with some time stretching or some time compression. Recently, [5] have shown that it is quite easy to propose

inner product with time elasticity capability at least for some restricted time series spaces, basically spaces containing uniformly sampled time series, all of which have the same lengths (in such cases, time series can be embedded easily in Euclidean spaces).

The aim of this paper is to derive an extension from this preliminary work for the construction of time elastic inner products, to achieve the construction of a time elastic inner product for a *quasi-unrestricted* set of time series, i.e. sets for which the times series are not uniformly sampled and have any lengths. Section two of the paper, following preliminary results presented in [5], gives the main notations used throughout this paper and presents a recursive construction for inner-like products. It then gives the conditions and the proof of existence of time elastic inner products (and time elastic vector spaces) defined on a *quasi-unrestricted* set of times series while explaining what we mean by *quasi-unrestricted*. The third section succinctly presents some applications, mainly to highlight some of the features of Time Elastic vector Spaces such as orthogonality.

2. Discrete Time Elastic Vector Spaces

2.1. Sequence and sequence element

Definition 2.1. Given a finite sequence A we note $A(i)$ the ith element (symbol or sample) of sequence A. We will consider that $A(i) \in S \times T$ where $(S, \oplus_S, \otimes_S)$ is a vector space that embeds the multidimensional *space* variables (e.g. $S \subset \mathbb{R}^d$, with $d \in \mathbb{N}^+$) and $T \subset \mathbb{R}$ embeds the *timestamps* variable, so that we can write $A(i) = (a(i), t_{a(i)})$ where $a(i) \in S$ and $t_{a(i)} \in T$, with the condition that $t_{a(i)} > t_{a(j)}$ whenever $i > j$ (timestamps strictly increase in the sequence of samples). A_i^j with $i \leq j$ is the subsequence consisting of the i_{th} through the j_{th} element (inclusive) of A. So $A_i^j = A(i)A(i+1)...A(j)$. Λ denotes the null element. By convention A_i^j with $i > j$ is the null time series, e.g. Ω.

2.2. Sequence set

Definition 2.2. The set of all finite discrete time series is thus embedded in a spacetime characterized by a single discrete *temporal* dimension, that encodes the timestamps, and any number of *spatial* dimensions that encode

the value of the time series at a given timestamp. We note $\mathbb{U} = \{A_1^p | p \in \mathbb{N}\}$ the set of all finite discrete time series. A_1^p is a time series with discrete index varying between 1 and p. We note Ω the empty sequence (with null length) and by convention $A_1^0 = \Omega$ so that Ω is a member of set $\mathbb{U}$. $|A|$ denotes the length of the sequence A. Let $\mathbb{U}_p = \{A \in \mathbb{U} \mid |A| \leq p\}$ be the set of sequences whose length is shorter or equal to p. Finally let $\mathbb{U}^*$ be the set of discrete times series defined on $(S - \{0_S\}) \times T$, i.e. the set of time series that do not contain the null *spatial* value. We denote by 0_S the null value in S.

2.3. Scalar multiplication on $\mathbb{U}^*$

Definition 2.3. For all $A \in \mathbb{U}^*$ and all $\lambda \in \mathbb{R}$, $C = \lambda \otimes A \in \mathbb{U}^*$ is such that for all $i \in \mathbb{N}$ such that $0 \leq i \leq |A|$, $C(i) = (\lambda.a(i), t_{a(i)})$ and thus $|C| = |A|$.

2.4. addition on $\mathbb{U}^*$

Definition 2.4. For all $(A, B) \in (\mathbb{U}^*)^2$, the addition of A and B, noted $C = A \oplus B \in \mathbb{U}^*$, is defined in a constructive manner as follows:_ Let i, j and k be in $\mathbb{N}$.

$k = i = j = 1$,
As far as $1 \leq i \leq |A|$ and $1 \leq j \leq |B|$,
if $t_{a_i} < t_{b_j}$, $C(k) = (a(i), t_{a_i})$ and $i \leftarrow i + 1, k \leftarrow k + 1$
else if $t_{a_i} > t_{b_j}$, $C(k) = (b(j), t_{b_j})$ and $j \leftarrow j + 1, k \leftarrow k + 1$
else if $a_i + b_j \neq 0$, $C(k) = (a(i) + b(j), t_{a_i})$ and $i \leftarrow i+1, j \leftarrow j+1, k \leftarrow k+1$
else $i \leftarrow i + 1, j \leftarrow j + 1$

Three comments need to be made at this level to clarify the semantic of the operator $\oplus$:

i) Note that the $\oplus$ addition of two time series of equal lengths and uniformly sampled coincides with the classical addition in vector spaces. Fig. 1 gives an example of the addition of two time series that are not uniformly sampled and that have different lengths.
ii) Implicitly (in light of the last case described in Def. 2.4), any sequence element of the sort $(0_S, t)$, where 0_S is the null value in S and $t \in T$ must be assimilated to the null sequence element Λ. For instance, the addition of $A = (1, 1)(1, 2)$ with $B = (-1, 1)(1, 2)$ is $C = A \oplus B = (2, 2)$: the addition of the two first sequence elements is $(0, 1)$ that is assimilated to Λ and as such suppressed in C.

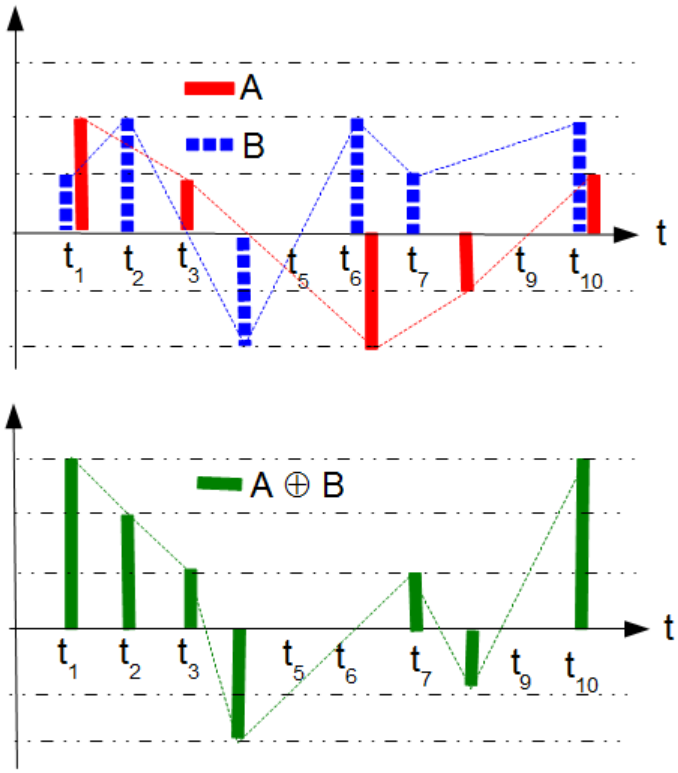

Figure 1: The $\oplus$ binary operator when applied to two discrete time series of variable lengths and not uniformly sampled. Co-occurring events have been slightly separated at the top of the figure for readability purposes.

iii) The $\oplus$ operator, when restricted to the set $\mathbb{U}^*$ is reversible in that if $C = A \oplus B$ then $A = C \oplus ((-1) \otimes B)$ or $B = C \oplus ((-1) \otimes A)$. This is not the case if we consider the entire set $\mathbb{U}$.

2.5. Time elastic product (TEP)

Definition 2.5. A function $< .,. >: \mathbb{U}^* \times \mathbb{U}^* \to \mathbb{R}$ is called a Time Elastic Product if, for any pair of sequences A_1^p, B_1^q, there exists a function $f : S^2 \to \mathbb{R}$, a non negative symmetric function $g : T^2 \to \mathbb{R}^+$ and three constants α, β and ξ in $\mathbb{R}$ such that the following recursive equation holds:

$$
\begin{aligned}
&< A_1^p, B_1^q >_{tep}= \\
&\sum \begin{cases} \alpha\cdot < A_1^{p-1}, B_1^q >_{tep} \\ \beta\cdot < A_1^{p-1}, B_1^{q-1} >_{tep} + f(a(p), b(q)) \cdot g(t_{a(p)}, t_{b(q)}) \\ \alpha\cdot < A_1^p, B_1^{q-1} >_{tep} \end{cases}
\end{aligned} \tag{1}
$$

This recursive definition requires defining an initialization. To that end we set, $\forall A \in \mathbb{U}^*$, $< A, \Omega >_{tep}=< \Omega, A >_{tep}=< \Omega, \Omega >_{tep}= \xi$, where ξ is a real constant (typically we set $\xi = 0$), and Ω is the null sequence, with the convention that ${A_i}^j = \Omega$ whenever $i > j$.

It has been shown in [5] that time elastic inner products can easily be constructed from Def. 2.5 using the $\oplus$ and $\otimes$ operations when we restrict the set of time series to some subset containing uniformly sampled time series of equal lengths (in that case, the $\oplus$ coincides with the classical addition on S). For instance, definitions 2.6 and 2.7 recursively define two TEP that are inner products on such restrictions.

Definition 2.6.

$$
\begin{aligned}
&< A_1^p, B_1^q >_{twip_1}= \tfrac{1}{3}\cdot \\
&\sum \begin{cases} < A_1^{p-1}, B_1^q >_{twip_1} \\ < A_1^{p-1}, B_1^{q-1} >_{twip_1} + e^{-\nu.d(t_{a(p)}, t_{b(q)})}(a(p) \cdot b(q)) \\ < A_1^p, B_1^{q-1} >_{twip_1} \end{cases}
\end{aligned} \tag{2}
$$

where d is a distance, and ν a *time stiffness* parameter.

Definition 2.7.

$$
\begin{aligned}
&< A_1^p, B_1^q >_{twip_2}= \tfrac{1}{1+2\cdot e^{-\nu}}\cdot \\
&\sum \begin{cases} e^{-\nu}\cdot < A_1^{p-1}, B_1^q >_{twip_2} \\ < A_1^{p-1}, B_1^{q-1} >_{twip_2} + e^{-\nu.d(t_{a(p)}, t_{b(q)})}(a(p) \cdot b(q)) \\ e^{-\nu}\cdot < A_1^p, B_1^{q-1} >_{twip_2} \end{cases}
\end{aligned} \tag{3}
$$

where d is a distance, and ν a *time stiffness* parameter.

It can be shown that $< ., . >_{twip_2}$ coincides with the Euclidean inner product on the considered restrictions of $\mathbb{U}$ when $\nu \to \infty$.

This paper addresses the more interesting question of the existence of similar elastic inner products on the set $\mathbb{U}^*$ itself, i.e. without any restriction on the lengths of the considered time series nor the way they are sampled. If the choice of functions f and g, although constrained, is potentially large, we show hereinafter that the choice for constants α, β and ξ is unique.

2.6. Existence of TEP inner products defined on $\mathbb{U}^*$

Theorem 2.1. *$< .,. >_{tep}$ is an inner product on $(\mathbb{U}^*, \oplus, \otimes)$ iff:*

i) $\xi = 0$.

ii) *$h : (S \times T) \to \mathbb{R}$ defined as $h((a, t_a)) = f(a, a) \cdot g(t_a, t_a)$ is strictly positive on $((S - \{0_S\}) \times T)$,*

iii) *f is an inner product on $(S, \oplus_S, \otimes_S)$, if we extend the domain of f on S while setting $f(0_S, 0_S) = 0$.*

iv) *$\alpha = 1$ and $\beta = -1$,*

2.6.1. proof of theorem 2.1

Proof of the direct implication

Let us suppose first that $< .,. >_{tep}$ is an inner product defined on $\mathbb{U}^*$. Then $< .,. >_{tep}$ is positive-definite, and thus $< \Omega, \Omega >_{tep} = \xi = 0$. Furthermore, for any $A = (a, t_a) \in \mathbb{U}^*$, $< A, A >_{tep} = h(a, t_a)) > 0$. Thus i) and ii) are satisfied. As g is non-negative, if we set $f(0_S, 0_S) = 0$, f is positive-definite on S.

It is also straightforward to show that f is symmetric if g and $< .,. >_{tep}$ are symmetric.

Since $\xi = 0$, for any A, B, and $C \in \mathbb{U}^*$ such that $A = (a, t)$, $B(b, t)$ and $C = (c, t_c)$, we have:
$< A \oplus B, C >_{tep} = h((a \oplus_S b, t), (c, t_c)) = f(a \oplus_S b, c).g(t, t_c)$.
As $< A \oplus B, C >_{tep} = < A, C >_{tep} + < B, C >_{tep}$
$= h((a, t), (c, t_c)) + h((b, t), (c, t_c))$
$= f(a, c).g(t, t_c) + f(a, c).g(t, t_c) = (f(a, c) + f(b, c)).g(t, t_c)$,
As g is non negative, we get that $f(a \oplus_S b, c) = (f(a, c) + f(b, c))$.
Furthermore, $< \lambda \otimes A, C >_{tep} = h((\lambda \otimes_S a, t), (c, t_c)) = f(\lambda \otimes_S a, c).g(t, t_c)$.
As $< \lambda \otimes A, C >_{tep} = \lambda. < A, C >_{tep} = \lambda.f(a, c).g(t, t_c)$ and g is non negative, we get that $f(\lambda \otimes_S a, c) = \lambda.f(a, c)$.
This shows that f is linear, symmetric and positive-definite. Hence it is an inner product on $(S, \oplus_S, \otimes_S)$ and iii) is satisfied.

Let us show that necessarily $\alpha = 1$ and $\beta = -1$. To that end, let us consider any A_1^p, B_1^q and C_1^r in $\mathbb{U}^*$, such that $p > 1, q > 1, r > 1$ and such that $t_{a_p} < t_{b_q}$, i.e. if $X_1^s = A_1^p \oplus B_1^q$, then $X_1^{s-1} = A_1^p \oplus B_1^{q-1}$.
Since by hypothesis $< .,. >_{tep}$ is an inner product $(\mathbb{U}^*, \oplus, \otimes)$, it is linear and thus we can write:
$< A_1^p \oplus B_1^q, C_1^r >_{tep} = < A_1^p, C_1^r >_{tep} + < B_1^q, C_1^r >_{tep}$.

Decomposing $< A_1^p \oplus B_1^q, C_1^r >_{tep}$, we obtain:
$< A_1^p \oplus B_1^q, C_1^r >_{tep} = \alpha. < A_1^p \oplus B_1^{q-1}, C_1^r >_{tep} +$
$\beta. < A_1^p \oplus B_1^{q-1}, C_1^{r-1} >_{tep} + f(b_q, c_r).g(t_{b_q}, t_{c_r}) + \alpha. < A_1^p \oplus B_1^q, C_1^{r-1} >_{tep}$
As $< .,. >_{tep}$ is linear we get:
$< A_1^p \oplus B_1^q, C_1^r >_{tep} = \alpha. < A_1^p, C_1^r >_{tep} + \alpha. < B_1^{q-1}, C_1^r >_{tep} +$
$\beta. < A_1^p, C_1^{r-1} >_{tep} + \beta. < B_1^{q-1}, C_1^{r-1} >_{tep} + f(b_q, c_r).g(t_{b_q}, t_{c_r}) +$
$\alpha. < A_1^p, C_1^{r-1} >_{tep} + \alpha. < B_1^q, C_1^{r-1} >_{tep}$
Hence,
$< A_1^p \oplus B_1^q, C_1^r >_{tep} = \alpha. < A_1^p, C_1^r >_{tep} + \beta. < A_1^p, C_1^{r-1} >_{tep} +$
$\alpha. < A_1^p, C_1^{r-1} >_{tep} + < B_1^q, C_1^r >_{tep}$

If we decompose $< A_1^p, C_1^r >_{tep}$, we get:
$< A_1^p \oplus B_1^q, C_1^r >_{tep} = (\alpha^2 + \beta + \alpha) < A_1^p, C_1^{r-1} >_{tep} + \alpha.\beta. < A_1^{p-1}, C_1^{r-1} >_{tep}$
$+ \alpha.f(a_p, c_r).g(t_{a_p}, t_{c_r}) + \alpha^2. < A_1^{p-1}, C_1^r >_{tep} + < B_1^q, C_1^r >_{tep}$

Thus we have to identify $< A_1^p, C_1^r >_{tep} = \alpha. < A_1^p, C_1^{r-1} >_{tep} + \beta. < A_1^{p-1}, C_1^{r-1} >_{tep} + f(a_p, c_r).g(t_{a_p}, t_{c_r}) + \alpha. < A_1^{p-1}, C_1^r >_{tep}$
with $(\alpha^2 + \beta + \alpha) < A_1^p, C_1^{r-1} >_{tep} + \alpha.\beta. < A_1^{p-1}, C_1^{r-1} >_{tep} + \alpha.f(a_p, c_r).g(t_{a_p}, t_{c_r}) +$
$\alpha^2. < A_1^{p-1}, C_1^r >_{tep}$.

The unique solution is $\alpha = 1$ and $\beta = -1$. That is if $< .,. >_{tep}$ is an existing inner product, then necessarily $\alpha = 1$ and $\beta = -1$, establishing iv).

Proof of the converse implication
Let us suppose that i), ii), iii) and iv) are satisfied and show that $< .,. >_{tep}$ is an inner product on $\mathbb{U}^*$.

First, by construction, since f and g are symmetric, so is $< .,. >_{tep}$.

It is easy to show by induction that $< .,. >_{tep}$ is non-decreasing with the

length of its arguments, namely, $\forall A_1^p$ and B_1^q in $\mathbb{U}^*$,
$< A_1^p, B_1^q >_{tep} - < A_1^p, B_1^{q-1} >_{tep} \geq 0$. Let $n = p + q$. The proposition is true at rank $n = 0$. It is also true if $A_1^p = \Omega$, whatever B_1^q is, or $B_1^q = \Omega$, whatever $< A_1^p$ is. Suppose it is true at a rank $n \geq 0$, and consider $A_1^p \neq \Omega$ and $B_1^q \neq \Omega$ such that $p + q = n$.
By decomposing $< A_1^p, B_1^q >_{tep}$ we get:
$< A_1^p, B_1^q >_{tep} - < A_1^p, B_1^{q-1} >_{tep} = - < A_1^{p-1}, B_1^{q-1} >_{tep} + f(a_p, b_q).g(t_{a_p}, t_{b_q}) + < A_1^{p-1}, B_1^q >_{tep}$
Since $f(a_p, b_q).g(t_{a_p}, t_{b_q}) > 0$ and the proposition is true by inductive hypothesis at rank n, we get that $< A_1^p, B_1^q >_{tep} - < A_1^p, B_1^{q-1} >_{tep}) > 0$. By induction the proposition is proved.

Let us show by induction on the length of the times series the positive definiteness of $< ., . >_{tep}$.
At rank 0 we have $< \Omega, \Omega >_{tep} = \xi = 0$. At rank 1, let us consider any time series of length 1, A_1^1. $< A_1^1, A_1^1 >_{tep} = f(a_1, a_1).g(t_{a_1}, t_{a_1}) > 0$ by hypothesis on f and g. Let us suppose that the proposition is true at rank $n > 1$ and let consider any time series of length $n+1$, A_1^{n+1}. Then, since $\alpha = 1$ and $\beta = -1$,
$< A_1^{n+1}, A_1^{n+1} >_{tep} = 2. < A_1^{n+1}, A_1^n >_{tep} - < A_1^n, A_1^n >_{tep} + f(a_{n+1}, a_{n+1}).g(t_{a_{n+1}}, t_{a_{n+1}})$.
Since $< A_1^{n+1}, A_1^n >_{tep} - < A_1^n, A_1^n >_{tep} \geq 0$, and $h(A(n+1), A(n+1) > 0$, $< A_1^{n+1}, A_1^{n+1} >_{tep} > 0$, showing that the proposition is true at rank $n+1$. By induction, the proposition is proved, which establishes the positive-definiteness of $< ., . >_{tep}$ since $< A_1^p, A_1^p >_{tep} = 0$ only if $A_1^p = \Omega$.

Let us consider any $\lambda \in \mathbb{R}$, and any A_1^p, B_1^q in $\mathbb{U}^*$ and show by induction on $n = p + q$ that $< \lambda \otimes A_1^p, B_1^q >_{tep} = \lambda. < A_1^p, B_1^q >_{tep}$:
The proposition is true at rank $n = 0$. Let us suppose that the proposition is true at rank $n \geq 0$, i.e. for all $r \leq n$, and consider any pair A_1^p, B_1^q of time series such that $p + q = n + 1$.
We have: $< \lambda \otimes A_1^p, B_1^q >_{tep} = \alpha. < \lambda \otimes A_1^p, B_1^{q-1} >_{tep} + \beta. < \lambda \otimes A_1^{p-1}, B_1^{q-1} >_{tep}$
$+ f(\lambda \otimes_S a_p, b_q).g(t_{a_p}, t_{b_q}) + \alpha. < \lambda \otimes A_1^{p-1}, B_1^q >_{tep}$
Since f is linear on $(S, \oplus_S, \otimes_S)$, and since the proposition is true by hypothesis at rank n, we get that $< \lambda \otimes A_1^p, B_1^q >_{tep} = \lambda.\alpha < A_1^p, B_1^{q-1} >_{tep}$
$+ \lambda.\beta. < A_1^{p-1}, B_1^{q-1} >_{tep} + \lambda.f(a_p, b_q).g(t_{a_p}, t_{b_q}) + \lambda.\alpha. < A_1^{p-1}, B_1^q >_{tep} =$
$\lambda. < A_1^p, B_1^q >_{tep}$.
By induction, the proposition is true for any n, and we have proved this proposition.

Furthermore, for any A_1^p, B_1^q and C_1^r in $\mathbb{U}^*$, let us show by induction on $n = p + q + r$ that $< A_1^p \oplus B_1^q, C_1^r >_{tep} = < A_1^p, C_1^r >_{tep} + < B_1^q, C_1^r >_{tep}$. Let X_1^s be equal to $A_1^p \oplus B_1^q$. The proposition is obviously true at rank $n = 0$. Let us suppose that it is true up to rank $n \geq 0$, and consider any A_1^p, B_1^q and C_1^r such that $p + q + r = n + 1$.

Three cases need then to be considered:

1) if $X_1^{s-1} = A_1^{p-1} \oplus B_1^{q-1}$, then $t_{a_p} = t_{b_q} = t$ and $< A_1^p \oplus B_1^q, C_1^r >_{tep} = \alpha. < A_1^p \oplus B_1^q, C_1^{r-1} >_{tep} + \beta. < A_1^{p-1} \oplus B_1^{q-1}, C_1^{r-1} >_{tep} + f((a_p + b_q), c_r).g(t, t_{c_r}) + \alpha. < A_1^{p-1} \oplus B_1^{q-1}, C_1^r >_{tep}$. Since f is linear on $(S, \oplus_S, \otimes_S)$, and the proposition true at rank n, we get the result.
2) if $X_1^{s-1} = A_1^p \oplus B_1^{q-1}$, then $t_{a_p} < t_{b_q} = t$ and $< A_1^p \oplus B_1^q, C_1^r >_{tep} = \alpha. < A_1^p \oplus B_1^q, C_1^{r-1} >_{tep} + \beta. < A_1^p \oplus B_1^{q-1}, C_1^{r-1} >_{tep} + f(b_q, c_r).g(t, t_{c_r}) + \alpha. < A_1^p \oplus B_1^{q-1}, C_1^r >_{tep}$. Having $\alpha = 1$ and $\beta = -1$ with the proposition supposed to be true at rank n we get the result.
3) if $X_1^{s-1} = A_1^{p-1} \oplus B_1^{q-1}$, we proceed similarly to case 2).

Thus the proposition is true at rank $n + 1$, and by induction the proposition is true for all n. This establishes the linearity of $< ., . >_{tep}$.
This ends the proof of the converse implication and theorem 2.1 is therefore established $\square$

.

The existence of functions f and g entering into the definition of $< ., . >_{tep}$ and satisfying the conditions allowing for the construction of an inner product on $(\mathbb{U}^*, \oplus, \otimes)$ is ensured by the following proposition:

Proposition 2.2. *The functions $f : S^2 \to \mathbb{R}$ defined as $f(a, b) = < a, b >_S$ where $< ., . >_S$ is an inner product on $(S, \oplus_S, \otimes_S)$ and $g : T^2 \to \mathbb{R}$ defined as $f(t_a, t_b)) = e^{-d(t_a, t_b)}$, where d is a distance defined on T^2 and $\nu \in \mathbb{R}^+$, satisfy the conditions required to construct an elastic inner product on $(\mathbb{U}^*, \oplus, \otimes)$.*

The proof of Prop.2.2 is obvious. This proposition establishes the existence of TEP inner products, that we will denote $TEIP$ (Time Elastic Inner Product). Note that $< ., . >_S$ can be chosen to be a $TEIP$ as well, in the case where a second *time elastic* dimension is required. This leads naturally to recursive definitions for TEP and $TEIP$.

Proposition 2.3. *For any $n \in \mathbb{N}$, and any discrete subset $T = \{t_1, t_2, \cdots, t_n\} \subset \mathbb{R}$, let $\mathbb{U}_{n,\mathbb{R},T}$ be the set of all time series defined on $\mathbb{R} \times T$ whose lengths are n (the time series in $\mathbb{U}_{n,\mathbb{R},T}$ are considered to be uniformly sampled). Then, the*

$TEIP$ on $\mathbb{U}_{n,\mathbb{R}}$ constructed from the functions f and g defined in Prop. 2.2 tends towards the Euclidean inner product when $\nu \to \infty$ if S is an Euclidean space and $< a, b >_S$ is the Euclidean inner product defined on S.

The proof of Prop.2.3 is straightforward and is omitted. Prop.2.3 shows that a $TEIP$ generalizes the classical Euclidean inner product.

3. Some applications

We present in the following sections some applications to highlight the properties of Time Elastic Vector Spaces ($TEVS$).

3.1. Distance in $TEVS$

The following proposition provides $\mathbb{U}^*$ with a norm and a distance, both induced by a $TEIP$.

Proposition 3.1. *For all $A_1^p \in \mathbb{U}^*$, and any $< .,. >$ $TEIP$ defined on $(\mathbb{U}^*, \oplus, \otimes)$ $\sqrt{< A_1^p, A_1^p >}$ is a norm on $\mathbb{U}^*$.*
For all pair $(A_1^p, B_1^q) \in (\mathbb{U}^*)^2$, and any $TEIP$ defined on $(\mathbb{U}^*, \oplus, \otimes)$, $\delta(A_1^p, B_1^q) = \sqrt{< A_1^p \oplus (-1. \otimes B_1^q), A_1^p \oplus (-1. \otimes B_1^q) >}$ defines a distance metric on $\mathbb{U}^*$.

The proof of Prop. 3.1 is straightforward and is omitted.

3.2. Orthogonalization in $TEVS$

To exemplify the effect of elasticity in $TEVS$, we give below the result of the Gram-Schmidt orthogonalization algorithm for two families of independent time series. The first family is composed of uniformly sampled time series having increasing lengths. The second family (a sine-cosine basis) is composed of uniformly sampled time series, all of which have the same length.

The tests which are described in the next sections were performed on a set $\mathbb{U}^*$ of discrete time series whose elements are defined on $(\mathbb{R} - \{0\} \times [0; 1])^2$ using the following $TEIP$:

$$
\begin{aligned}
&< A_1^p, B_1^q >_{teip} = \\
&\sum \begin{cases} < A_1^p, B_1^{q-1} >_{teip} \\ - < A_1^{p-1}, B_1^{q-1} >_{teip} + a(p)b(q) \cdot e^{-\nu.|t_{a_p} - t_{b_q}|} \\ < A_1^{p-1}, B_1^q >_{teip} \end{cases}
\end{aligned} \tag{4}
$$

3.2.1. Orthogonalization of an independent family of time series with increasing lengths

The family of time series we are considering is composed of 11 time series uniformly sampled, whose lengths are 11 samples:

$$
\begin{aligned}
&(1,0) \\
&(\epsilon,0)(1,1/10) \\
&(\epsilon,0)(\epsilon,0)(1,1/10) \\
&\cdots \\
&(\epsilon,0)(\epsilon,1/10)(\epsilon,2/10)\cdots(1,1)
\end{aligned} \tag{5}
$$

Since, the zero value cannot be used for the space dimension, we replaced it by ϵ, which is the smallest non zero positive real for our test machine (i.e. 2^{-1074}). The result of the Gram-Schmidt orthogonalization process using $\nu = .01$ on this basis is given in Fig.2.

3.2.2. Orthogonalization of a sine-cosine basis

An orthonormal family of discrete sine-cosine functions is not anymore orthogonal in a $TEVS$. The result of the Gram-Schmidt orthogonalization process using $\nu = .01$ when applied on a discrete sine-cosine basis is given in Fig.3, in which only the 8 first components are displayed. The lengths of the waves are 128 samples.

3.3. Kernel methods in $TEVS$

A wide range of literature exists on kernels, among which [6], [7] and [8] present some large syntheses of major results.

Definition 3.1. A kernel on a non empty set U refers to a complex (or real) valued symmetric function $\varphi(x,y) : U \times U \to \mathbb{C}$ (or $\mathbb{R}$).

Definition 3.2. Let U be a non empty set. A function $\varphi : U \times U \to \mathbb{C}$ is called a positive (resp. negative) definite kernel if and only if it is Hermitian (i.e. $\varphi(x,y) = \overline{\varphi(y,x)}$ where the *overline* stands for the conjugate number)

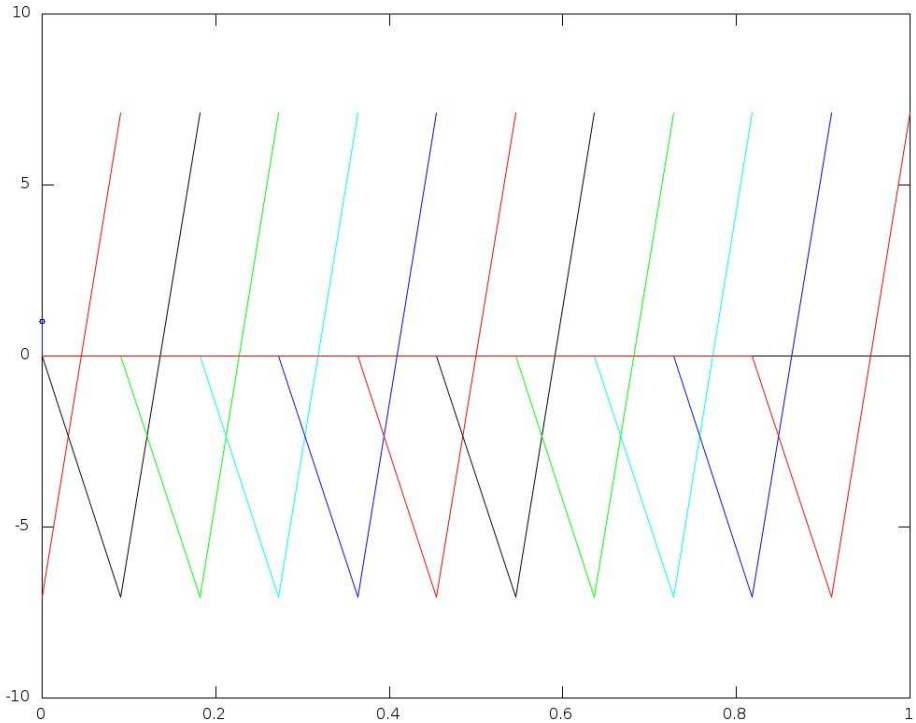

Figure 2: Result of the orthogonalization of the family of length time series defined in Eq.5 using $\nu = .01$: except for the first *spike* located at *time* 0, each original *spike* is replaced by two *spikes*, one negative the other positive.

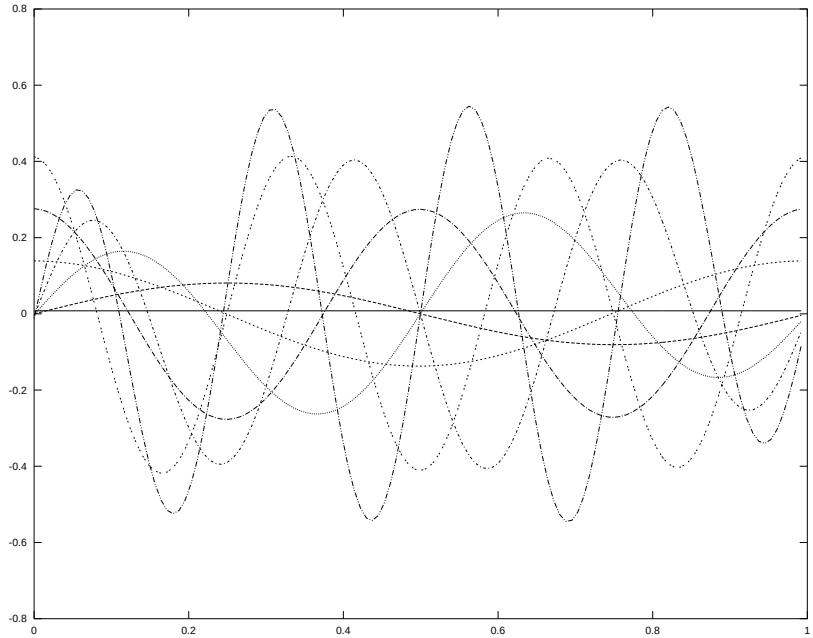

Figure 3: Orthogonalization of the *sine-cosine* basis using $\nu = .01$: the waves are slightly deformed jointly in amplitude and in frequency. For readability of the figure, we have presented the 8 first components

for all x and y in U and $\sum_{i,j=1}^{n} c_i \bar{c}_j \varphi(x_i, x_j) \geq 0$ (resp. $\sum_{i,j=1}^{n} c_i \bar{c}_j \varphi(x_i, x_j) \leq 0$), for all n in $\mathbb{N}$, $(x_1, x_2, ..., x_n) \in U^n$ and $(c_1, c_2, ..., c_n) \in \mathbb{C}^n$.

Definition 3.3. Let U be a non empty set. A function $\varphi : U \times U \to \mathbb{C}$ is called a conditionally positive (resp. conditionally negative) definite kernel if and only if it is Hermitian (i.e. $\varphi(x, y) = \overline{\varphi(y, x)}$ for all x and y in U) and $\sum_{i,j=1}^{n} c_i \bar{c}_j \varphi(x_i, x_j) \geq 0$ (resp. $\sum_{i,j=1}^{n} c_i \bar{c}_j \varphi(x_i, x_j) \leq 0$), for all $n \geq 2$ in $\mathbb{N}$, $(x_1, x_2, ..., x_n) \in U^n$ and $(c_1, c_2, ..., c_n) \in \mathbb{C}^n$ with $\sum_{i=1}^{n} c_i = 0$.

In the last two above definitions, it is easy to show that it is sufficient to consider mutually different elements in U, i.e. collections of distinct elements $x_1, x_2, ..., x_n$.

Definition 3.4. A positive (resp. negative) definite kernel defined on a finite set U is also called a positive (resp. negative) semidefinite matrix. Similarly, a positive (resp. negative) conditionally definite kernel defined on a finite set is also called a positive (resp. negative) conditionally semidefinite matrix.

3.3.1. Definiteness of $TEIP$ based kernel

Proposition 3.2. *A $TEIP$ is a positive definite kernel.*

The proof of Prop. 3.2 is straightforward and is omitted.

3.3.2. SVM classification using a TEP based kernel

In [5], $< .,. >_{twip_2}$ (Eq.2.7) have been experimented on a classification task using a SVM classifier on 20 datasets containing times series uniformly sampled and having the same lengths inside each dataset. On the same data, we get similar results for $< .,. >_{teip}$ (Eq.4) and do not report them in this paper. The benefit of introducing some time elasticity, controlled using the parameter ν is quite clear when comparing the classification error rates obtained using a Gaussian kernel exploiting the distance derived from $< .,. >_{teip}$ (Prop. 3.1) with the classification error rates obtained using a Gaussian kernel exploiting the Euclidean distance.

3.4. Elastic Cosine similarity in $TEVS$, with application to symbolic (e.g. textual) information retrieval

Similarly to the definition of the cosine of two vectors in Euclidean space, we define the elastic cosine of two sequences by using any TEP that satisfies the conditions of theorem 2.1.

Definition 3.5. Given two sequences, A and B, the elastic cosine similarity of these two sequences is given using a time elastic inner product $< X, Y >_e$ and the induced norm $\|X\|_e = \sqrt{< X, X >_e}$ as
$similarity = \cos_e(\theta) = \frac{<A \cdot B>_e}{\|A\|_e \|B\|_e}$

In the case of textual information retrieval, namely text matching, the timestamps variable coincides with the index of words into the text, and the spatial dimensions encode the words into a given dictionary. For instance, each word can be represented using a vector whose dimension is the size of the set of concepts (or senses) that cover the conceptual model associated

to the dictionary and each coordinate selected into $[0;1]$ encodes the degree of presence of the concept or senses into the considered word. In that case, the elastic cosine similarity measure takes value into $[0;1]$, 0 indicating the lowest possible similarity value between two texts and 1 the greatest possible similarity value between two texts. The elastic cosine similarity takes into account the order of occurrence of the words into a text which could be an advantage compared to the Euclidean cosine measure that does not cope with the words ordering.

Let us consider the following elastic inner product dedicated to text matching. In the following definition, A_1^p and B_1^q are sequences of words that represent textual content.

Definition 3.6.

$$< A_1^p, B_1^q >_{teip_{tm}} = \sum \begin{cases} < A_1^{p-1}, B_1^q >_{teip_{tm}} \\ - < A_1^{p-1}, B_1^{q-1} >_{teip_{tm}} + e^{-\nu.|t_{a(p)} - t_{b(q)}|} \delta(a(p), b(q)) \\ < A_1^p, B_1^{q-1} >_{teip_{tm}} \end{cases} \tag{6}$$

where $\delta(x, y) = 1$ if $x = y$ (x and y identify the same word), 0 otherwise, and ν a *time stiffness* parameter.

Proposition 3.3. *For $\nu = 0$, the elastic inner product defined in Eq.3.6 coincides with the euclidean inner product between two vectors whose coordinates correspond to term frequencies observed into the A_1^p and B_1^q text sequences. If, we change the definition of δ by the $\delta(x, y) = IDF(x)$ if $x = y$, 0 otherwise, where $IDF(x)$ is the inverse document frequency of term x into the considered collection, then for $\nu = 0$, $< A_1^p, B_1^q >_{teip_{tm}}$ coincides with the euclidean inner product between two vectors whose coordinates correspond to the TF-IDF (term frequency times the inverse document frequency) of terms occurring into the A_1^p and B_1^q text sequences.*

The proof of proposition 3.3 is straightforward an is omitted.

Thus, the elastic cosine measure derived from the elastic inner product defined by Eq.3.6 generalizes somehow the cosine measure implemented in the vector model [9] and commonly used in the text information retrieval community.

4. Conclusion

This paper proposed what we call a family of *time elastic inner products* able to cope with non-uniformly sampled time series of various lengths, as far as they do not contain the *zero* value. These constructions allow one to embed any such time series in a single vector space, that some how generalizes the notion of Euclidean vector space. The recursive structure of the construction offers the possibility to manage several *time elastic* dimensions. Some applicative benefits could be expected in time series analysis when time *elasticity* is an issue, for instance in the field of numeric or symbolic sequence data mining.

[1] V. M. Velichko, N. G. Zagoruyko, International Journal of Man-Machine Studies 2 (1970) 223–234.

[2] H. Sakoe, S. Chiba, in: Proceedings of the 7th International Congress of Acoustic, pp. 65–68.

[3] L. Chen, R. Ng, in: Proceedings of the 30th International Conference on Very Large Data Bases, pp. 792–801.

[4] P. F. Marteau, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 306–318.

[5] P.-F. Marteau, S. Gibet, CoRR abs/1005.5141 (2010).

[6] C. Berg, J. P. R. Christensen, P. Ressel, Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions, volume 100 of *Graduate Texts in Mathematics*, Springer-Verlag, New York, 1984.

[7] B. Scholkopf, A. J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, Cambridge, MA, USA, 2001.

[8] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, New York, NY, USA, 2004.

[9] G. Salton, M. McGill, Introduction to Modern Information Retrieval, McGraw-Hill Book Company, 1984.