



HAL
open science

Mapping Chronicles to a k -dimensional Euclidean Space via Random Projections

Alexandre Sahuguède, Soheib Fergani, Euriell Le Corrnc, Marie-Véronique V
Le Lann

► **To cite this version:**

Alexandre Sahuguède, Soheib Fergani, Euriell Le Corrnc, Marie-Véronique V Le Lann. Mapping Chronicles to a k -dimensional Euclidean Space via Random Projections. 14th annual IEEE International Conference on Automation Science and Engineering (IEEE CASE 2018), Aug 2018, Munich, Germany. 6p. hal-01817539

HAL Id: hal-01817539

<https://laas.hal.science/hal-01817539v1>

Submitted on 18 Jun 2018

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.

Mapping Chronicles to a k -dimensional Euclidean Space via Random Projections

Alexandre Sahuguède*, Soheib Fergani*, Euriell Le Corronc* and Marie-Véronique Le Lann*

*LAAS-CNRS, Université de Toulouse, CNRS, INSA, UPS, Toulouse, France

e-mail: {asahugue, sfergani, elecorro, mvlelann}@laas.fr

Abstract—This paper is concerned with an innovative strategy that maps chronicles, that are timed discrete event models, to a k -dimensional Euclidean space via random projections. The proposed approach is a projection that takes into account both characteristics of events, namely event types, and temporal constraints of chronicles. This will lead to an unbounded convex polytope in the Euclidean space that contains all the possible instances of the corresponding chronicle. It allows to easily and efficiently compare chronicles. Such comparisons are useful in a fault diagnosis purpose to discriminate chronicles representing behaviors of dynamic processes. Examples and preliminary results are provided in this paper to introduce the proposed methodology.

Index Terms—Timed Discrete Event Systems, Chronicles, Random Projection, Euclidean Space.

I. INTRODUCTION

In the recent years, both academic and industrial communities have been interested in the study of the timed discrete event models called chronicles for a fault diagnosis purpose (see [1] for a recent survey, [2], [3]). Indeed, chronicles are temporal patterns where events are partially ordered which are well-suited to capture dynamic processes with an event abstraction of the information of interest. Namely, to encapsulate the nominal behavior, only necessary data are captured in one or several chronicles whereas another set of chronicles encapsulates the faulty behaviors. Then, from on-line timed observations of the system, a pattern recognition of the chronicles expressing a faulty situation can be done efficiently and leads to detect that it is in that faulty situation [4].

Recently, many studies have been achieved on chronicles, more particularly on discovering chronicles in data generated by a specific process. Although diagnosis is more efficient with chronicles, their design is not trivial. Indeed, for complex systems it is necessary to achieve it automatically. An interesting chronicle discovery algorithm that finds chronicles in a temporal sequence on a frequency criterion is proposed in [5]. An extension to multiple temporal sequences is presented in [6]. The chronicle discovery process is discussed in healthcare applications in [7] and in [8] where clinical pathways are analyzed. However, these algorithms provide a large number of chronicles that the expert must then sort, analyze and compare. For instance, some learned chronicles are irrelevant when only few well defined chronicles are enough to obtain an efficient fault diagnosis. Similar chronicles are close representations of the same dynamic behavior of a complex process and should be

reduced to improve the diagnosis. For this sake, solutions have to be provided.

In this paper, we propose to project chronicles in a k -dimensional Euclidean space in order to compare and classify them easily and efficiently. In [9], a method based on random projections is used to map timed event sequence to a point in this space for finding similar situations in large sequence of events. Here, we use this mathematical projection method on a chronicle. The proposed method allows to map chronicles to an unbounded polytope in an Euclidean space. In this space, it is easier and faster to compare values than a direct comparison of chronicles where both event and time must be taken into account.

The rest of the paper is organized as follows. Section II gives the definition of the required chronicle concepts with a first simple mapping of the chronicle's temporal constraints. Section III presents the projection of chronicles into polytopes in a k -dimensional Euclidean space. Section IV explains some properties of the obtained polytopes in the Euclidean space and how they relate to the reference chronicles. Section V analyzes the impact of the value of k . Section VI concludes this work and discusses interesting perspectives.

II. CHRONICLE CONCEPT

A. Background on chronicle

This section explains the chronicle concepts used in this work. Chronicles are a way to express temporal patterns of a specific process [4]. Common temporal properties of chronicles are easily checked by means of graph theory. Chronicles are then represented as Temporal Constraint Networks [10].

Definition 1 (Event): An event is defined by $x = (e, t)$ with an event type $e \in E$, and a time instant $t \in \mathbb{R}$.

Definition 2 (Temporal sequence): A temporal sequence is a time-ordered set of events denoted $\mathcal{S} = \{x_1, \dots, x_n\}$ with n a finite number of events, and $t_j < t_{j+1}$, $j = 1, \dots, n - 1$. The set of all event types occurring in \mathcal{S} is called $E_{\mathcal{S}}$.

Example 1: The temporal sequence $\mathcal{S} = \{x_1, x_2, x_3, x_4, x_5\}$ where $x_1 = (a, 1)$, $x_2 = (b, 4)$, $x_3 = (b, 8)$, $x_4 = (a, 16)$, $x_5 = (b, 18)$, is graphically represented in Figure 1. This sequence points out the difference between event and event type that is several events can share the same event type. For instance, events x_1 and x_4 are different occurrences of event type a at different times.

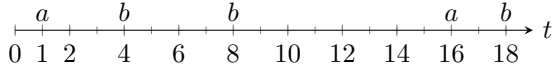


Fig. 1: A temporal sequence \mathcal{S} .

Definition 3 (Temporal constraint): A temporal constraint is a tuple $\tau_{ij} = (x_i, x_j, t^-, t^+)$, also noted $\tau_{ij} = x_i[t^-, t^+]x_j$, where $0 < t^- \leq t^+$. A temporal constraint $\tau_{ij} = x_i[t^-, t^+]x_j$ is said satisfied by a couple of events $x_i = (e_i, t_i)$, $x_j = (e_j, t_j)$ if and only if $(t_j - t_i) \in [t^-, t^+]$.

Definition 4 (Chronicle): A chronicle \mathcal{C} is a pair $(\mathcal{X}, \mathcal{T})$ where $\mathcal{X} = \{x_1, \dots, x_n\}$ is a set of partially ordered events with n a finite number of events, and $\mathcal{T} = \{\tau_{ij}\}_{1 \leq i < j \leq n}$ is a set of temporal constraints on \mathcal{X} . $E_{\mathcal{C}}$ denotes the set of all event types of \mathcal{C} . An n -length chronicle is a chronicle with n events.

A chronicle can be graphically represented by a directed graph where the nodes are the events of \mathcal{X} and the transitions are the temporal constraints of \mathcal{T} . In each temporal constraint τ_{ij} , x_i is the starting node and x_j is the ending node.

Example 2: Figure 2 is the graphical representation of the 3-length chronicle $\mathcal{C} = (\mathcal{X}, \mathcal{T})$ where $\mathcal{X} = \{x_1 = (a, t_1), x_2 = (b, t_2), x_3 = (c, t_3)\}$ and $\mathcal{T} = \{\tau_{12} = x_1[10, 20]x_2, \tau_{13} = x_1[25, 35]x_3, \tau_{23} = x_2[10, 20]x_3\}$. Another 3-length chronicle $\mathcal{C}_2 = (\mathcal{X}_2, \mathcal{T}_2)$ where $\mathcal{X}_2 = \{x_1 = (a, t_1), x_2 = (b, t_2), x_3 = (b, t_3)\}$ and $\mathcal{T}_2 = \{\tau_{12} = x_1[1, 3]x_2, \tau_{13} = x_1[2, 4]x_3\}$ is illustrated Figure 3. This chronicle points out the partial order between events in the chronicle formalism.

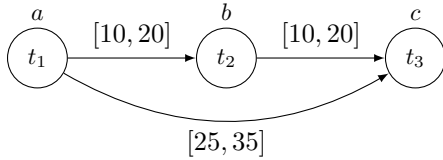


Fig. 2: A 3-length chronicle $\mathcal{C} = (\mathcal{X}, \mathcal{T})$.

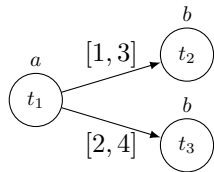


Fig. 3: Another 3-length chronicle $\mathcal{C}_2 = (\mathcal{X}_2, \mathcal{T}_2)$.

Definition 5 (Chronicle instance): Given a chronicle $\mathcal{C} = (\mathcal{X}, \mathcal{T})$, a chronicle instance is a set of events denoted $\mathcal{I}_{\mathcal{C}}$ such that their event types are those of \mathcal{X} and their time occurrences satisfy all the temporal constraints \mathcal{T} of \mathcal{C} .

Example 3: Let us define a chronicle $\mathcal{C} = (\mathcal{X}, \mathcal{T})$ with $\mathcal{X} = \{x_1 = (a, t_1), x_2 = (b, t_2)\}$ and $\mathcal{T} = \{\tau_{12} = x_1[1, 3]x_2\}$. Given the temporal sequence $\mathcal{S} = \{(a, 1), (b, 4), (b, 8), (a, 16), (b, 18)\}$, two instances of the chronicle \mathcal{C} appear: $\mathcal{I}_{\mathcal{C}}^1 = \{x_1, x_2\} = \{(a, 1), (b, 4)\}$, and $\mathcal{I}_{\mathcal{C}}^2 = \{x_1, x_2\} = \{(a, 16), (b, 18)\}$.

The number of chronicle instances $\mathcal{I}_{\mathcal{C}}$ of a consistent chronicle \mathcal{C} is infinite. A consistent chronicle is a chronicle that have at least one instance [10]. In a chronicle, events are temporally constrained to each other but they are not constrained themselves. The first event could occur anytime between $-\infty$ and $+\infty$.

B. Mapping chronicles with time

This subsection shows that an n -length chronicle can be represented by an unbounded convex polytope¹, a geometrical volume in an n -dimension Euclidean space. Chronicle can be easily mapped to this geometrical volume by a simple orthogonal projection along each corresponding timeline. For instance, [11] uses this orthogonal projection on simpler temporal patterns than chronicles where only time is of interest.

Given an n -length chronicle $\mathcal{C} = (\mathcal{X}, \mathcal{T})$, its temporal constraints \mathcal{T} define a polytope \mathcal{P} in \mathbb{R}^n . Each dimension of \mathbb{R}^n is associated with a time occurrence of the events of \mathcal{X} . The polytope \mathcal{P} is defined by the following formula:

$$\forall \tau_{ij} \in \mathcal{T}, \mathcal{P} = \{-t_i + t_j \geq t_{ij}^-, t_i - t_j \geq -t_{ij}^+\}. \quad (1)$$

\mathcal{P} can be rewritten as the following linear constraint²:

$$At \geq b \quad (2)$$

with $A \in \mathbb{R}^{m \times n}$, $t \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$, where m is the number of inequalities and n the length of the chronicle \mathcal{C} . Vector b contains the bounds of all the temporal constraints, and matrix A defines events on which those constraints take effect. Solutions of Equation (2) are the time occurrences of all the acceptable instances $\mathcal{I}_{\mathcal{C}}$ of \mathcal{C} . As there is an infinite number of chronicle instances $\mathcal{I}_{\mathcal{C}}$, the polytope \mathcal{P} will always be unbounded [13].

Example 4: Let $\mathcal{C} = (\mathcal{X}, \mathcal{T})$ be the chronicle of Example 2 and Figure 2. This chronicle defines the polytope \mathcal{P} shown in Figure 4 such that $A \in \mathbb{R}^{6 \times 3}$, $t \in \mathbb{R}^3$ and $b \in \mathbb{R}^6$ with:

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 10 \\ -20 \\ 25 \\ -35 \\ 10 \\ -20 \end{bmatrix}.$$

One solution of Equation (2), highlighted in Figure 4 by a black square, is:

$$t = [8 \quad 21 \quad 33]^\top$$

and corresponds to the projection of the instance $\mathcal{I}_{\mathcal{C}} = \{(a, 8), (b, 21), (c, 33)\}$.

This representation in \mathbb{R}^n has several limitations: the dimension of the Euclidean space is the length n of the chronicle, so comparisons between chronicles of different lengths is not easily done in this representation; the projection is done only on time occurrences of the events and the

¹In the rest of this paper, convex polytopes will be called polytopes.

²Let us note that \mathcal{P} can also be seen as a Difference Bound Matrix [12].

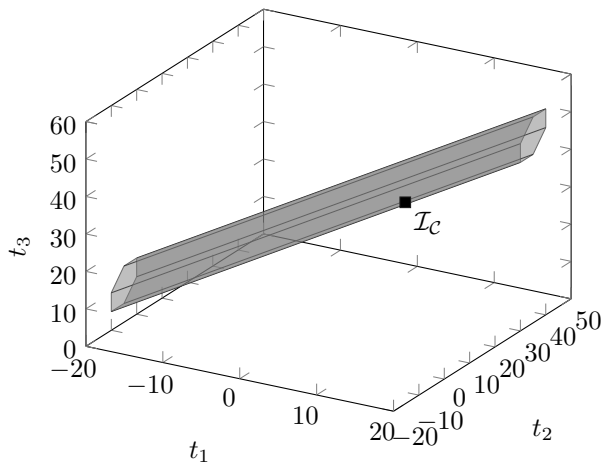


Fig. 4: Polytope \mathcal{P} of Example 4. Each axis corresponds to the timeline of each event. The projection of $\mathcal{I}_C = \{(a, 8), (b, 21), (c, 33)\}$ is highlighted with a black square.

event types are not considered. We aim to try and tackle these limitations in the next section by adapting a technique proposed in [9]. This technique, originally used with timed event sequences, is applied here to project a chronicle in an Euclidean space.

III. MAPPING CHRONICLES TO A k -DIMENSIONAL EUCLIDEAN SPACE

Chronicles can be easily mapped to a Euclidean space by a simple orthogonal projection along each corresponding timeline as seen in Section II-B. However, in such a simple projection, only temporal constraints are taken into account whereas event types are not used. Event types can bring more information when an objective of comparison and classification of projected chronicles is needed. In the work presented here, in addition to temporal constraints, event types are considered in the chronicle projection by means of methods coming from the pattern recognition domain, such as random projection [14]. Let us start by the *fingerprint* of an event type e in the k -dimensional Euclidean space.

Definition 6 (Fingerprint of an event type): Let e be an event type of E . Let $\rho(e)$ be a k -dimensional vector filled with k normally distributed random variables with mean 0 and variance 1. The random k -dimensional vector $\rho(e)$ is the fingerprint of the event type e in the k -dimensional Euclidean space.

Since a chronicle has several events containing several event types, each of them will have its own fingerprint. Those fingerprints are compiled in the *fingerprint matrix*.

Definition 7 (Fingerprint matrix of a chronicle): Let \mathcal{C} be a n -length chronicle, $\rho \in \mathbb{R}^{n \times k}$ is the matrix containing the fingerprints of all the event types e_1, \dots, e_n of the set $\mathcal{X} = \{x_1, \dots, x_n\}$ of \mathcal{C} :

$$\rho = [\rho(e_1) \quad \rho(e_2) \quad \dots \quad \rho(e_n)]. \quad (3)$$

Thanks to the use of the chronicle fingerprint matrix in addition to the simple orthogonal projection, both event and

time information contained in a chronicle instance \mathcal{I}_C can be considered in the projection proposed here. Indeed, time occurrences are easily taken into account by an orthogonal projection whereas event types are considered by means of their fingerprints. So, the random projection of a chronicle instance \mathcal{I}_C in the Euclidean space is the fingerprint matrix ρ of \mathcal{C} weighted by its time occurrences t .

Definition 8 (Random projection of a chronicle instance): The random projection $x' \in \mathbb{R}^k$ of a chronicle instance \mathcal{I}_C to the k -dimensional Euclidean space is the product between the fingerprint matrix $\rho \in \mathbb{R}^{n \times k}$ of \mathcal{C} and the time occurrences $t \in \mathbb{R}^n$ of \mathcal{I}_C :

$$x' = \rho t. \quad (4)$$

Example 5: Let $\mathcal{I}_C = \{(a, 8), (b, 21), (c, 33)\}$ be the instance detailed in Example 4 and $\rho = [\rho(a) \quad \rho(b) \quad \rho(c)]$ be the fingerprint matrix of \mathcal{C} randomly obtained such that:

$$\begin{aligned} \rho(a) &= [-0.1220 \quad -1.0868 \quad 0.6843]^\top, \\ \rho(b) &= [-1.0752 \quad 0.0333 \quad 0.7448]^\top, \\ \rho(c) &= [0.0336 \quad -0.5266 \quad 0.4625]^\top. \end{aligned}$$

For the sake of simplicity in this example, we have set k to 3 even though it is clear that some problems occur when k is set too low as discussed in Section V. The random projection x' of \mathcal{I}_C in this 3-dimension space is highlighted in Figure 5 by a black square. It is obtained by means of the linear projection seen in Equation (4):

$$x' = [-22.4457 \quad -25.3749 \quad 36.3794]^\top.$$

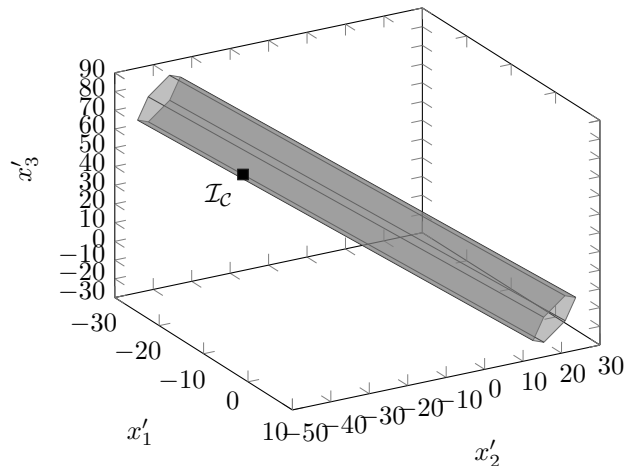


Fig. 5: Polytope \mathcal{P}' where the projection of $\mathcal{I}_C = \{(a, 8), (b, 21), (c, 33)\}$ is highlighted with a black square.

Since the fingerprint matrix of a chronicle is randomly generated, the random projection of instances of different chronicles must be done in the same k -dimensional Euclidean space. This means that fingerprint matrices of different chronicles should be computed coherently. Indeed, when chronicles share event types, the fingerprints are also

shared. For instance, event types a and b are common to both chronicles \mathcal{C} and \mathcal{C}_2 of Example 2. When event types or use of event types are not seen in fingerprint matrices previously processed, a new fingerprint is generated. For instance, this is the case for the second use of the event type b in chronicle \mathcal{C}_2 . Thus, the fingerprint matrix ρ_2 of \mathcal{C}_2 is obtained in the same k -dimensional Euclidean space than \mathcal{C} ($k = 3$):

$$\begin{aligned} \rho_2 &= [\rho(a) \quad \rho(b_1) \quad \rho(b_2)], \\ &= \begin{bmatrix} -0.1220 & -1.0752 & 0.5377 \\ 1.0868 & 0.0333 & 1.8339 \\ 0.6843 & 0.7448 & -2.2588 \end{bmatrix}. \end{aligned}$$

$\rho(a)$ is the same for both chronicles \mathcal{C} and \mathcal{C}_2 . $\rho(b)$ of \mathcal{C} is the same than $\rho(b_1)$ of \mathcal{C}_2 whereas $\rho(b_2)$ of \mathcal{C}_2 is new.

By means of the fingerprint matrix ρ , x' is a representation in the k -dimensional Euclidean space of an instance \mathcal{I}_C . Applying this fingerprint matrix directly on the polytope \mathcal{P} of a chronicle \mathcal{C} allows to project all its acceptable chronicle instances \mathcal{I}_C . This leads to a new polytope \mathcal{P}' in \mathbb{R}^k . However, the size k of the Euclidean space should be higher than the size n of the chronicle to obtain ρ left invertible. The solutions of the linear constraint defined by the polytope \mathcal{P}' are all the instances \mathcal{I}_C of the chronicle \mathcal{C} projected in the Euclidean space.

Proposition 1: Let $At \geq b$ be the linear constraint defined by a chronicle \mathcal{C} , and let $x' = \rho t$ be the projection of an instance \mathcal{I}_C in the k -dimensional Euclidean space. There exists a matrix $A' \in \mathbb{R}^{m \times k}$, iff $k \geq n$, such that:

$$A'x' \geq b, \quad (5)$$

Proof: As ρ is filled with random variables that are independent and identically distributed, each row and each column are linearly independent. The pseudo-inverse $\rho^+ \in \mathbb{R}^{k \times n}$ can be computed [15]. When $k \geq n$, ρ^+ is a left inverse and is given by:

$$\rho^+ = (\rho^\top \rho)^{-1} \rho^\top.$$

Equation (4) becomes:

$$\rho^+ x' = t.$$

With Equation (2):

$$A\rho^+ x' \geq b.$$

Finally, with $A' = A\rho^+$,

$$A'x' \geq b.$$

Example 6: Let $At \geq b$ be the projection in \mathbb{R}^3 of the 3-length chronicle \mathcal{C} seen in Example 4 ($n = 3$). Let $x' = \rho t$ be the projection of an instance \mathcal{I}_C in the k -dimensional space with $k = 3$ detailed in Example 5. First, the pseudo-inverse $\rho^+ \in \mathbb{R}^{3 \times 3}$ is computed:

$$\rho^+ = \begin{bmatrix} -1.7670 & -2.2641 & -2.4495 \\ -0.6169 & 0.3442 & 0.4367 \\ 3.6075 & 2.7953 & 5.0827 \end{bmatrix}.$$

Then, the matrix $A' \in \mathbb{R}^{6 \times 3}$ is computed with $A' = A\rho^+$:

$$A' = \begin{bmatrix} 1.1501 & 2.6083 & 2.8863 \\ -1.1501 & -2.6083 & -2.8863 \\ 5.3745 & 5.0594 & 7.5322 \\ -5.3745 & -5.0594 & -7.5322 \\ 4.2244 & 2.4511 & 4.6459 \\ -4.2244 & -2.4511 & -4.6459 \end{bmatrix}.$$

The polytope \mathcal{P}' defined by $A'x' \geq b$ is shown in Figure 5. The projection of the instance \mathcal{I}_C detailed in Example 5 is a solution of $A'x' \geq b$. So:

$$\begin{bmatrix} 1.1501 & 2.6083 & 2.8863 \\ -1.1501 & -2.6083 & -2.8863 \\ 5.3745 & 5.0594 & 7.5322 \\ -5.3745 & -5.0594 & -7.5322 \\ 4.2244 & 2.4511 & 4.6459 \\ -4.2244 & -2.4511 & -4.6459 \end{bmatrix} \begin{bmatrix} -22.4457 \\ -25.3749 \\ 36.3794 \end{bmatrix} \geq \begin{bmatrix} 10 \\ -20 \\ 25 \\ -35 \\ 10 \\ -20 \end{bmatrix}.$$

Remark 1: Let us remark that the chronicle projection depends on the partial order of the events x_i in the set \mathcal{X} of the chronicle. Thus, similar chronicles for which the only difference is the arbitrary choice of the numbering of the events may have different projections. For instance, the following chronicles have a different mapping whether their events are numbered as in Figure 6a or as in Figure 6b. Indeed, their fingerprint matrices will be identical but the difference in the temporal constraints will have a different result on the projection. More work needs to be done to take into account this phenomenon due to the partial order of the events.

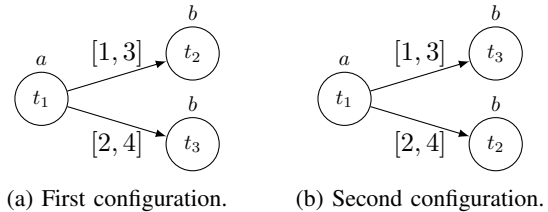


Fig. 6: Two configurations of the numbering of the events of the same chronicle that can be mapped differently.

Remark 2: When mapping several chronicles of different lengths in the same k -dimensional Euclidean space, k should be greater than the length of the longest chronicle.

IV. ANALYSIS OF THE PROJECTED CHRONICLE

Properties of the random mapping of a chronicle in the k -dimensional Euclidean space are closely related to the studied chronicle model. In this section, some of these relations are explained.

A. Geometry of a projected chronicle

The projected chronicle is a polytope where all the edges are parallel and share the same direction vector. This direction vector \vec{v} could be easily computed in the polytope \mathcal{P} defined in Equation (2). For all $n \in \mathbb{N}$, consider the n -length

chronicle \mathcal{C} defined by $At \geq b$. The kernel of the linear map A is the set of equations $At = 0$:

$$\ker(A) = \{t \in K^n | At = 0\}.$$

As all time occurrences t_i of an event are constrained to a time occurrence t_j of another event, $At = 0$ can be rewritten as follows:

$$\forall i, j, \quad -x_i + x_j = 0.$$

So, the rank of the kernel of a n -length chronicle \mathcal{C} is always 1. For all $n \in \mathbb{N}$, the direction vector $\vec{v} \in \mathbb{R}^n$ of a n -length chronicle \mathcal{C} is unique and defined by:

$$\vec{v} = [1 \quad 1 \quad \dots \quad 1]^\top. \quad (6)$$

In the k -dimensional Euclidean space, the direction vector \vec{v}' of the polytope \mathcal{P}' can be deduced from \vec{v} by means of the fingerprint matrix ρ :

$$\vec{v}' = \rho \vec{v}. \quad (7)$$

Therefore, the direction vector \vec{v}' is a representation of the set of event types present in the chronicle \mathcal{C} .

Another relation between temporal constraints of the chronicle and the projected chronicle exists. Let us define \mathcal{V} as the volume of the section between the polytope \mathcal{P}' of a projected chronicle and the hyper-plan orthogonal to its direction vector \vec{v}' . The bounds of the temporal constraints define the vector b and is associated with \mathcal{V} . With narrow temporal constraints, a smaller set of acceptable time occurrences leads to a smaller \mathcal{V} . While with wide temporal constraints, much more time occurrences are acceptable for the instances and lead to a larger \mathcal{V} . Furthermore, low values of lower and upper bounds of the temporal constraints generate a small offset for the edges, the projected chronicle is then close to the origin. High values for the bounds of the temporal constraints will have the opposite effect.

B. Intersection of two projected chronicles

A common way to compare chronicles is through their instances. For example, two chronicles are said *equivalent* when they share the exact same set of instances [10]. In the Euclidean space, instances that are common to two chronicles means an intersection of two polytopes.

Let $\mathcal{C}_1 = \{\mathcal{X}_1, \mathcal{T}_1\}$ with $\mathcal{X}_1 = \{x_1 = (a, t_1), x_2 = (b, t_2)\}$, $\mathcal{T}_1 = \{\tau_{12} = x_1[2, 6]x_2\}$, and $\mathcal{C}_2 = \{\mathcal{X}_2, \mathcal{T}_2\}$ with $\mathcal{X}_2 = \{x_3 = (a, t_3), x_4 = (b, t_4)\}$, $\mathcal{T}_2 = \{\tau_{34} = x_3[5, 9]x_4\}$ be two 2-length chronicles. Some instances of these chronicles are identical, e.g. $\mathcal{I}_{\mathcal{C}_1} = \mathcal{I}_{\mathcal{C}_2} = \{(a, 3), (b, 8)\}$ is an instance of both \mathcal{C}_1 and \mathcal{C}_2 . In the Euclidean space, the random projections of $\mathcal{I}_{\mathcal{C}_1}$ and $\mathcal{I}_{\mathcal{C}_2}$ will be equal and both polytopes \mathcal{P}'_1 and \mathcal{P}'_2 will contain this projection. The intersection of these polytopes, shown in Figure 7, are all the equal projected instances of \mathcal{C}_1 and \mathcal{C}_2 .

As expected, direction vectors v'_1 and v'_2 of the polytopes \mathcal{P}'_1 and \mathcal{P}'_2 are similar. Indeed, both \mathcal{C}_1 and \mathcal{C}_2 have the same set of events $\mathcal{X}_1 = \mathcal{X}_2$ with the same set of event types $E_{\mathcal{C}_1} = E_{\mathcal{C}_2}$. The value of \mathcal{V} of these polytopes are similar as the temporal constraints τ_{12} of \mathcal{C}_1 and τ_{34} of \mathcal{C}_2 have the

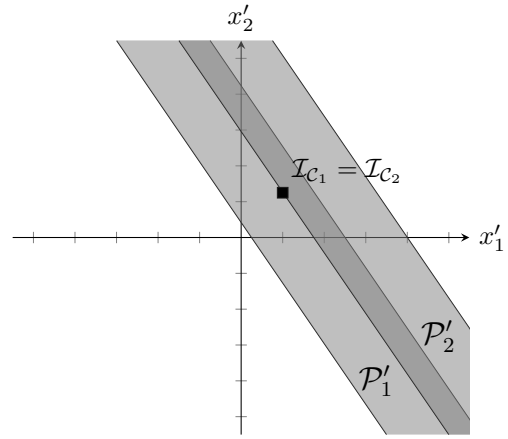


Fig. 7: Polytopes \mathcal{P}'_1 and \mathcal{P}'_2 in \mathbb{R}^k with $k = 2$. Projection of $\mathcal{I}_{\mathcal{C}_1} = \mathcal{I}_{\mathcal{C}_2}$ is highlighted with a black square.

same width. \mathcal{P}'_2 is farther from the origin than \mathcal{P}'_1 since the lower bound of temporal constraint of \mathcal{C}_2 is higher than \mathcal{C}_1 .

V. IMPACT OF k

The choice of the value of k is very important and must get a particular attention as described in this section. An approximation of the information contained in the event types is expected due to the use of random projection. In the Johnson-Lindenstrauss Theorem [16], this approximation can be quantified by a value ϵ in the interval $[0, 1]$. A high value of ϵ , obtained by setting k to a high value, will reduce the approximation achieved. A value of k set too low leads to an increased approximation. This can be seen in the following example.

Let \mathcal{C}_3 and \mathcal{C}_4 be two 2-length chronicles with $E_{\mathcal{C}_3} = \{a, b\}$ and $E_{\mathcal{C}_4} = \{c, d\}$. Let us project both chronicles in \mathbb{R}^k , with $k = 2$. This projection gives the two polytopes \mathcal{P}'_3 and \mathcal{P}'_4 shown in Figure 8. In this plot, it can be seen that \mathcal{P}'_3 and \mathcal{P}'_4 are intersecting. The intersecting part means that the projection of some instances $\mathcal{I}_{\mathcal{C}_3}$ of \mathcal{C}_3 and some instances $\mathcal{I}_{\mathcal{C}_4}$ of \mathcal{C}_4 are equal. However, \mathcal{C}_3 and \mathcal{C}_4 can not have common instances as they have different sets of event types. This intersection can happen because of the random projection approximation. A k set too low will increase the probability of this undesirable effect.

Moreover, with a statistical analysis, it can be shown that the value of k will have an impact on how far from the origin the chronicles will be in the Euclidean space. To do so, let us consider the squared distance from the origin to a vector obtained by mapping an instance of a n -length chronicle in the k -dimensional Euclidean space, that is:

$$d^2 = \sum_{j=1}^k x'(j)^2.$$

Equation (4) defines the vector x' and $x'(j)^2$ is the square of each element of x' :

$$x'(j)^2 = \left(\sum_{i=1}^n \rho(e_i, j) t_i \right)^2.$$

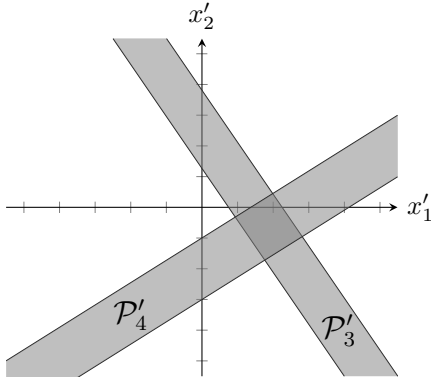


Fig. 8: An undesirable effect when k is set too low.

Since vector x is filled with normally distributed random variables with a mean value equal to 0 and variance 1, $E(\rho(e_i, j)^2) = 1$ and $E(\rho(e_i, j)) = 0$. The expected value of $x'(j)$ is:

$$E(x'(j)^2) = \sum_{i=1}^n t_i^2.$$

So the expected value of d^2 is given by:

$$E(d^2) = \sum_{j=1}^k E(x'(j)^2) = k \sum_{i=1}^n t_i^2.$$

This means that a high value of k will have an important effect on the dispersion of chronicles mapped in the Euclidean space. In other words, an n -length chronicle will be farther from the origin with a high value of k than with a low value of k . For the variance, it is given by:

$$\begin{aligned} VAR(d^2) &= E(d^4) - E(d^2)^2 \\ &= k \left(E(x'(j)^4) - E(x'(j)^2)^2 \right) \\ &= k \left(3 \left(\sum_{i=1}^n t_i^2 \right)^2 - \left(\sum_{i=1}^n t_i^2 \right)^2 \right) \\ &= 2k \left(\sum_{i=1}^n t_i^2 \right)^2. \end{aligned}$$

VI. CONCLUSION AND DISCUSSION

This paper provides a new method based on random projections to map chronicles in a k -dimensional Euclidean space. This approach leads to an unbounded polytope that contains all the possible instances of the corresponding chronicle. Some simple relations between properties of polytopes in the Euclidean space and the chronicle model have been established. The impact of the value of k has been discussed and it was shown that this parameter should be chosen carefully.

The new chronicle representation will improve the chronicle discovery algorithm results, especially for fault diagnosis objectives. Indeed, comparison and analysis of the chronicles obtained by such algorithms can be easily achieved when they are projected in an Euclidean space using directly

distances between polytopes. Computations of distances between polytopes is commonly used in geo-spatial applications [17]. However, a bounded polytope is required for these distances. As seen in Section III, a chronicle will always be mapped to an unbounded polytope. A possible solution to try and tackle this problem is to define the distance between chronicles as the distance between the dual of their mapped polytopes. The dual of a polytope always exists and is bounded [18].

A well defined distance is crucial for defining a notion of similarity of chronicles. Similar chronicles are close representations of the same dynamic behavior of a process. Similarity is useful for fault diagnosis as the less similar chronicles are, the more accurate the diagnosis will be.

REFERENCES

- [1] A. Subias, "Diagnosis from chronicles: an overview of related challenges," in *Congreso Internacional de Ingenieria Mecatronica y Automatizacion*, 2013, pp. 1–11.
- [2] B. Morin and H. Debar, "Correlation of intrusion symptoms: An application of chronicles," in *Recent Advances in Intrusion Detection: 6th International Symposium, RAID 2003*, 2003, pp. 94–112.
- [3] Y. Pencolé and A. Subias, "A chronicle-based diagnosability approach for discrete timed-event systems: Application to web-services," *Journal of Universal Computer Science*, vol. 15, no. 17, pp. 3246–3272, nov 2009.
- [4] C. Dousson and P. Le Maigat, "Chronicle recognition improvement using temporal focusing and hierarchization," in *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007, pp. 324–329.
- [5] D. Cram, B. Mathern, and A. Mille, "A complete chronicle discovery approach: application to activity analysis," *Expert Systems*, vol. 29, no. 4, pp. 321–346, 2012.
- [6] A. Subias, L. Travé-Massuyès, and E. Le Corrionc, "Learning chronicles signing multiple scenario instances," in *19th World Congress of The International Federation of Automatic Control*, 2014, pp. 10397–10402.
- [7] Z. Huang, X. Lu, and H. Duan, "On mining clinical pathway patterns from medical behaviors," *Artificial intelligence in medicine*, vol. 56, no. 1, pp. 35–50, 2012.
- [8] Y. Dauxais, T. Guyet, D. Gross-Amblard, and A. Happe, "Discriminant chronicles mining - application to care pathways analytics," in *16th Conference on Artificial Intelligence in Medicine*, 2017, pp. 234–244.
- [9] H. Mannila and J. K. Seppänen, "Finding similar situations in sequences of events via random projections," in *Proceedings of the First SIAM International Conference on Data Mining*, 2001, pp. 1–16.
- [10] R. Dechter, I. Meiri, and J. Pearl, "Temporal constraint networks," *Artificial Intelligence*, vol. 49, no. 1, pp. 61–95, 1991.
- [11] T. Guyet and R. Quiniou, "Mining temporal patterns with quantitative intervals," in *2008 IEEE International Conference on Data Mining Workshops*, Dec 2008, pp. 218–227.
- [12] D. L. Dill, "Timing assumptions and verification of finite-state concurrent systems," in *International Conference on Computer Aided Verification*, 1989, pp. 197–212.
- [13] C. D. Toth, J. O'Rourke, and J. E. Goodman, *Handbook of Discrete and Computational Geometry*. CRC Press, 2017.
- [14] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: applications to image and text data," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 245–250.
- [15] R. Penrose, "A generalized inverse for matrices," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, no. 3, pp. 406–413, 1955.
- [16] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz maps into a Hilbert space," *Contemporary Mathematics*, vol. 26, pp. 189–206, 01 1984.
- [17] M. J. Atallah, C. C. Ribeiro, and S. Lifschitz, "Computing some distance functions between polygons," *Pattern Recognition*, vol. 24, no. 8, pp. 775 – 781, 1991.
- [18] H. S. M. Coxeter, *Regular polytopes*. Courier Corporation, 1973.