

# Generalized Multiple Correlation Coefficient as a Similarity Measurement between Trajectories

Julen Urain<sup>1</sup>, and Jan Peters<sup>1,2</sup>

**Abstract**—Similarity distance measure between two trajectories is an essential tool to understand patterns in motion, for example, in Human-Robot Interaction or Imitation Learning. The problem has been faced in many fields, from Signal Processing, Probabilistic Theory field, Topology field or Statistics field. Anyway, up to now, none of the trajectory similarity measurement metrics are invariant to all possible linear transformation of the trajectories (rotation, scaling, reflection, shear mapping or squeeze mapping). Also not all of them are robust in front of noisy signals or fast enough for real-time trajectory classification. To overcome this limitation this paper proposes a similarity distance metric that will remain invariant in front of any possible linear transformation. Based on Pearson's Correlation Coefficient and the Coefficient of Determination, our similarity metric, the Generalized Multiple Correlation Coefficient (GMCC) is presented like the natural extension of the Multiple Correlation Coefficient. The motivation of this paper is two-fold: First, to introduce a new correlation metric that presents the best properties to compute similarities between trajectories invariant to linear transformations and compare it with some state of the art similarity distances. Second, to present a natural way of integrating the similarity metric in an Imitation Learning scenario for clustering robot trajectories.

## I. INTRODUCTION

In the industry of tomorrow it is expected the robots to collaborate with humans. The robots should not only imitate the humans [1], [2], but also interact with us. The objective of Human-Robot Interaction (HRI) is to understand intentions [3] and shape the interactions [4] between one or more humans and one or more robots [5].

There is a lot of work done in human intention recognition and generalization. In [6], Amor et al. developed a Human-robot solution based on Probabilistic Principal Component Analysis (PPCA). Their algorithm computes a low-dimensional projection to encapsule human-robot interactions in the training and then in testing, observing human motion, the robot was able to infer his movement. They found out that for certain motions, even if the robot was taught how to interact when human was moving hands up, the robot was able to generalize even when the human hands where moving down. In [7], Interaction Primitives (IP) [8] were used to learn human robot interactions. Probabilistic Movement Primitives (ProMP) [9] were used to model human motions for each skill and then the robot was able to guess the skill human was doing through Bayes and then

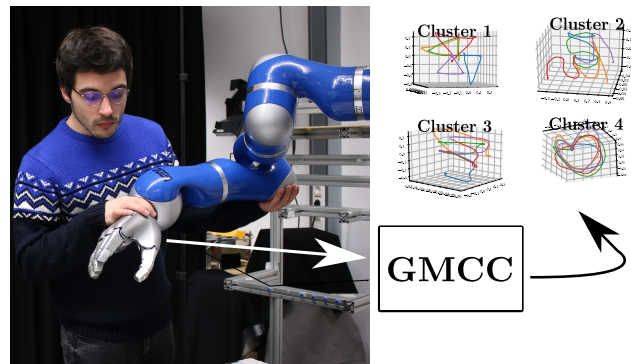


Fig. 1: GMCC applied for clustering taught trajectories in Imitation Learning

generalize on the skill applying conditioning over ProMP. In [10], ProMP's were also used to adapt robot movements to a human-robot table tennis match. In [11], Hidden Markov Model (HMM) were used to recognize human gestures. The recognition was invariant to the starting position of the gestures. The sequences of human motion were segmented into atomic components and clustered by HMM.

Anyway, up to now, the generalization on human gestures recognition has been applied on a few specific tasks and very limited to small variations between the learned human motion and the observed one. The actual algorithms tends to fail recognizing gestures if these are similar shape to the ones already in our gestures library but with different rotations.

This limitation can be overcome developing a good similarity measurement that can extract the relevant features from the learned human motions and then be able to recognize rotated or scaled new observations. In an ideal case of HRI, it should be enough teaching the robot how to interact with the human for some particular cases and then, during the testing, when the human is doing a similar shape motion, but maybe rotated, the robot could find the similarities between the previously learned motions and the new ones.

### A. Problem Statement

The motion similarity has been studied in different fields. For example, In Probabilistic Theory field, the motion has been represented as a probability distributions. This let the similarity distance to be robust with noisy motions. In Signal Processing field the motion are represented as time series. Similarity metrics deals with time morphed time series. None of the commented similarity metrics consider spatial transformations among trajectories and so, these similarity metrics give poor performance measuring the similarity distance among rotated or scaled trajectories.

<sup>1</sup>Intelligent Autonomous Systems, TU Darmstadt

<sup>2</sup>MPI for Intelligent Systems, Tuebingen

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No. #820807 (SHAREWORK)

{urain,peters}@ias.tu-darmstadt.de

Statistics field provides the best similarity metrics to compare spatially transformed trajectories. In Statistics, two trajectories similarity depends on how correlated are between each other. A high correlation between trajectories will mean a high similarity. For the case of unidimensional trajectories  $x_t \in \mathbb{R}$  and  $y_t \in \mathbb{R}$  where  $t$  is the time samples, Pearson's correlation coefficient is measured as follows

$$\rho(x_t, y_t) = \frac{\text{cov}(x_t, y_t)}{\sigma_x \sigma_y}. \quad (1)$$

The correlation coefficient will be 1 if the trajectories are totally positive correlated, 0 if there is no any linear correlation and -1 if there is a total negative linear correlation. When it comes multidimensional variables, several correlation coefficients exist. The first multidimensional coefficient was proposed by Harold Hotelling in 1936 [12]. The Canonical Correlation Coefficient (CAC), measures the correlation between a linear combination of the first variable  $\mathbf{X}_t \in \mathbb{R}^n$  and a linear combination of second variable  $\mathbf{Y}_t \in \mathbb{R}^n$ . During this paper bold symbols will be used to represent vectors and matrices in order to differentiate from scalars in italic.

$$\begin{aligned} (\mathbf{a}^*, \mathbf{b}^*) &= \arg \max_{\mathbf{a}, \mathbf{b}} \rho(\mathbf{a}^\top \mathbf{X}_t, \mathbf{b}^\top \mathbf{Y}_t), \\ \text{CAC}(\mathbf{X}_t, \mathbf{Y}_t) &= \rho(\mathbf{a}^{*\top} \mathbf{X}_t, \mathbf{b}^{*\top} \mathbf{Y}_t). \end{aligned} \quad (2)$$

Nowadays, the most famous multidimensional correlation coefficients for considering spatial transformations are two:  $R_v$  Coefficient ( $R_v$ ) and Distance Correlation (dCor).  $R_v$  [13] is computed as follows

$$R_v(\mathbf{X}_t, \mathbf{Y}_t) = \frac{\text{tr}(\mathbf{X}_t \mathbf{X}_t^\top \mathbf{Y}_t \mathbf{Y}_t^\top)}{\sqrt{\text{tr}(\mathbf{X}_t \mathbf{X}_t^\top)^2 \text{tr}(\mathbf{Y}_t \mathbf{Y}_t^\top)^2}}. \quad (3)$$

As it can be seen  $R_v$  is computed with  $\mathbf{X}_t \mathbf{X}_t^\top$  and  $\mathbf{Y}_t \mathbf{Y}_t^\top$ . If we consider  $\mathbf{X}_2 = \mathbf{X} \mathbf{Q}$  where  $\mathbf{Q}$  is any orthogonal matrix, then,  $\mathbf{X}_2 \mathbf{X}_2^\top = \mathbf{X} \mathbf{Q} \mathbf{Q}^\top \mathbf{X}^\top = \mathbf{X} \mathbf{X}^\top$ . From here, it is concluded that  $R_v$  will be invariant to any orthogonal transformation, so invariant to rotations.

On the other hand, dCor [14] is not only invariant to linear transformations but also to some non linear transformations also. Having two multidimensional vectors  $\mathbf{X}_t \in \mathbb{R}^n$  and  $\mathbf{Y}_t \in \mathbb{R}^n$ ; first, the euclidean distance is computed between different time samples

$$\begin{aligned} a_{k,j} &= \|\mathbf{X}_j - \mathbf{X}_k\|, \quad k, j = 0, \dots, T \\ b_{k,j} &= \|\mathbf{Y}_j - \mathbf{Y}_k\|, \quad k, j = 0, \dots, T \end{aligned}$$

where  $\|\cdot\|$  is the euclidean distance. Once computed,  $a_{k,j}$  and  $b_{k,j}$  are double normalized

$$\begin{aligned} A_{j,k} &= a_{k,j} - \bar{a}_{\cdot,k} - \bar{a}_{j,\cdot} + \bar{a}_{\cdot,\cdot}, \\ B_{j,k} &= b_{k,j} - \bar{b}_{\cdot,k} - \bar{b}_{j,\cdot} + \bar{b}_{\cdot,\cdot}, \end{aligned}$$

and the obtained values used for computing the distance covariance (dCov) and the distance correlation (dCor)

$$\begin{aligned} \text{dCov}^2(\mathbf{X}, \mathbf{Y}) &= \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n A_{j,k} B_{j,k}, \\ \text{dCor}(\mathbf{X}, \mathbf{Y}) &= \frac{\text{dCov}(\mathbf{X}, \mathbf{Y})}{\sqrt{\text{dVar}(\mathbf{X}) \text{dVar}(\mathbf{Y})}} \end{aligned} \quad (4)$$

where,  $\text{dVar}^2(\mathbf{X}) = \text{dCov}^2(\mathbf{X}, \mathbf{X})$ . dCor will only consider the sign of the distances computed between different time samples. If the normalized distances from  $k$  to  $j$  have the same sign in both  $A_{k,j}$  and  $B_{k,j}$ , the correlation will increase and if the signs are different then decrease. So, dCor does not care if the relation between the variables is linear or nonlinear.

In this paper, we propose a new correlation measurement for multidimensional variables. As far as we know, there is no any trajectory similarity measurement that remains invariant to any linear transformations and only linear transformations. Our algorithm is computed as an extension of the Multiple Correlation Coefficient for multidimensional variables and it will be demonstrated that the Multiple Correlation Coefficient can be considered as a particular case of the the Generalized Multiple Correlation Coefficient (GMCC) for unidimensional variables.

The rest of the paper is organized as follows: In Section I-B, similarity measures from other fields are presented. In Section II, first the mathematical basis of our algorithm is briefly explained, to continue with the presentation of the novel similarity metric, GMCC. Section III is dedicated to the experiments developed on the robot. First, GMCC's characteristics are studied and compared to  $R_v$  and dCor. Later, the GMCC is applied in a clustering task of robot trajectories Fig. 1. Finally, the paper concludes in Section IV discussing ideas for future work.

## B. Related Work

The similarity measurements presented in Section I-A are the best similarity measurements when it comes to measure spatial invariant similarities. Nevertheless, there are also relevant similarity measurement in other fields: in 1) similarity distances from Signal Processing field are presented, in 2) similarity distances from Topology field and in 3) similarity distances from Probabilistic Theory field.

- 1) In **Time Series** field, the similarity has been faced from time shifting perspective. Dynamic Time Warping (DTW) [15] is a well-known algorithm. Considering two similar trajectories, DTW finds the best matching between them applying time shifting, anyway it doesn't give any metric about how similar two time series are. A similarity measure considering the time shifting is the Edit Distance With Real Penalty (EDR) [16]. On it, Chen et al. propose a similarity measurement based on string edit distance. The similarity distance between two trajectories will be related with the number of elements we should add or delete to make them equal. None of these similarity metrics are invariant to spatial transformations.
- 2) On **Topology** field, the similarity between trajectories is studied without considering the time information and so, only spatial information is used. The two most well-known similarity measurements are Frechet distance [17] and Hausdorff distance [18]. Frechet distance is presented as the bottleneck maximum distance between two trajectories. From a different approach,

Hausdorff distance is the greatest of all the distances from a point in one set to the closest point in the other set. These distances doesn't consider spatial transformations of the trajectories.

- 3) On **Probabilistic Theory** field, Maximum Likelihood Estimation (MLE) is the most used algorithm for measuring the similarity between two trajectories. In [3], MLE was used to compute the similarity between learned gestures and the observations. The benefits of using probabilistic approaches are several. In contrast with similarity metrics from other fields, MLE can not only give the probability of how similar two trajectories are, but also, predict the trajectory from partial observations of the trajectory. So, the algorithms can be used for recognizing and predicting the human motions. The drawback is that no spatial transformations are considered between learned motions and the observed ones.

## II. GENERALIZED MULTIPLE CORRELATION COEFFICIENT

We introduce a new correlation coefficient. This coefficient can be used as a similarity measure between multidimensional time series. Having two multidimensional trajectories  $\mathbf{X}_t \in \mathbb{R}^n$  and  $\mathbf{Y}_t \in \mathbb{R}^n$  where  $t$  is the time samples and  $n$  the dimensions, the GMCC is a measure of how well can be mapped a linear transformation between the trajectories  $\mathbf{X}_t$  and  $\mathbf{Y}_t$ . As lot of correlation measurements, the GMCC takes values between 0 and 1. A coefficient of 1 means that there is a complete correlation between  $\mathbf{X}_t$  and  $\mathbf{Y}_t$ , and a value of 0 means that there is no any possible linear relation between the vectors  $\mathbf{X}_t$  and  $\mathbf{Y}_t$ .

The invariability on linear transformation is the biggest contributions of this new similarity metric. The coefficient will remain invariant in front of any possible linear transformation and only linear transformations, so the correlation coefficient will decrease sharply when no linear correlations appear. This contribution is particularly interesting in comparison with other correlation coefficients such as  $R_v$ , that remains invariant with only with orthogonal transformations and dCor that remains invariant with not only linear but also some nonlinear transformations.

As additional contributions, and similarly to other correlation coefficients, GMCC is robust with noisy trajectories and the coefficient is obtained instantly in comparison with other Signal Processing field similarity metrics such as EDR [16] or Time Warp Edit Distance (TWED) [19] which computation time is of  $\mathcal{O}(t^2)$  where  $t$  is the time samples of the compared trajectories  $\mathbf{X}_t$  and  $\mathbf{Y}_t$ .

GMCC is an extension of the Multiple Correlation Coefficient. While the Multiple Correlation Coefficient computes the correlation between an univariate dependant variable  $y$  and multiple independent variables  $\mathbf{X}$ , the GMCC extends it to multivariate dependant variables  $\mathbf{Y}$ . Moreover, in Section II-B it will be demonstrated that Multiple Correlation Coefficient can be computed as a particular case of the GMCC.

The Section II continues as follows: In Section II-A a brief introduction to the mathematical basis for developing our coefficient is presented. Once the Coefficient of Determination and the Multiple Correlation Coefficients are introduced, in Section II-B our novel coefficient, GMCC is presented and computed.

### A. Coefficient of Determination

Considering an univariate dependant variable  $y_t \in \mathbb{R}$  and multiple independent variables  $\mathbf{X}_t \in \mathbb{R}^n$ , Coefficient of Determination ( $R^2$ ) is the fraction of the variance in the dependent variable  $y$  that is predictable from the independent variables  $\mathbf{X}$  [20]

$$R^2 = \frac{\sum_{t=0}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=0}^T (y_t - \bar{y})^2} \quad (5)$$

where  $\hat{y}_t = \mathbf{A}^\top \mathbf{X}_t = a_1 x_{1t} + a_2 x_{2t} + \dots + a_n x_{nt}$  with  $\mathbf{A} \in \mathbb{R}^n$  is the prediction model.  $y_t$  can be expressed in terms of predicted values  $\hat{y}_t$  and residual errors  $\epsilon_t$

$$y_t = \hat{y}_t + \epsilon_t$$

then, the parameters  $\mathbf{A}$  that minimize the squared sum of the residual error  $\epsilon_t$  can be obtained through an optimization function

$$\mathbf{A}^* = \arg \min_{\mathbf{A}} \left( \sum_{t=1}^T \epsilon_t^2 \right) \quad (6)$$

The solution of the optimization function from Eq. 6,  $\mathbf{A}^*$ , is computed by Least-Squared regression and so, the optimal prediction model is  $\hat{y}_t^* = \mathbf{A}^{*\top} \mathbf{X}_t$ . In the particular case of having this optimal model,  $R^2$  can be expressed in terms of the Pearson's Correlation coefficient [20]

$$R^2 = \mathbf{c}^\top \mathbf{R}_{xx}^{-1} \mathbf{c} \quad (7)$$

where  $\mathbf{c} = (\rho_{yx_1}, \dots, \rho_{yx_n})$  is a vector with the correlation coefficient between each individual independent variables  $x_i$  and the dependent variable  $y$ .  $\mathbf{R}_{xx}$  is a correlation matrix measuring the correlation between the independent variables

$$\mathbf{R}_{xx} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} & \dots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \rho_{23} & \dots & \rho_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \rho_{n3} & \dots & \rho_{nn} \end{bmatrix}.$$

It is important to remark that in the case of having the independent variables completely uncorrelated between each other,  $\mathbf{R}_{xx}$  becomes into the identity matrix

$$\mathbf{R}_{xx} = \mathbb{I}_{n \times n} \quad (8)$$

and so, the coefficient of determination can be expressed as a sum of squared Pearson's Correlation coefficients. Mixing Eq. 7 and Eq. 8

$$R^2 = \sum_{i=1}^n \rho_{yx_i}^2.$$

Multiple Correlation Coefficient is the square root of Eq. 7 and it represents how well we can predict the parameters

of the dependant variable  $y$  with the multiple independent variables  $\mathbf{X}$  using the optimal linear model.

$$R = \sqrt{\mathbf{c}^T \mathbf{R}_{xx}^{-1} \mathbf{c}^T} \quad (9)$$

As remarked in [20], Multiple Correlation Coefficient will increase when the number of the independent variables increase and so it will give a high correlation coefficient between any two variables. From regression perspective is logical. As we are applying Least-Square regression to fit  $\mathbf{X}$  with  $y$ , when the independent variable increase, the noise can be used to fit in  $y$  and so, we could have over-fitting problems. Anyway, it is expected the noise not to affect a lot in low-dimensional cases such as three dimensional euclidean motions.

### B. Generalized Multiple Correlation Coefficient

We propose a new correlation coefficient for the case of multiple dependant variables  $\mathbf{Y}_t \in \mathbb{R}^n$  and multiple independent variables  $\mathbf{X}_t \in \mathbb{R}^n$  where  $t$  is the time sample and  $n$  is the dimension. The given name is Generalized Multiple Correlation Coefficient (GMCC), as it can be introduced as the extension of the Multiple Correlation Coefficient for multivariate dependent variables. The most relevant property from GMCC in comparison with the state of the art multivariate correlation coefficients such as  $R_v$  and dCor is that it will remain invariant to any possible linear transformations and only linear transformations. For any nonsingular matrix  $\mathbf{H}$

$$\text{GMCC}(\mathbf{X}, \mathbf{H}\mathbf{X}) = 1. \quad (10)$$

For HRI or Imitation Learning framework, a similarity measurement that remains invariant between linearly transformed trajectories is a useful tool. This is the case of different scale, rotation, mirror, squeeze mapped or shear mapped trajectories. This similarity measurement can be applied from classification of human gestures in HRI environment to clustering of taught skills in Imitation Learning as we will show in Section III.

GMCC lays between  $R_v$  that remains invariant in front of rotations and scaling and the dCor that remains invariant for not only linear transformations but also some nonlinear transformations. In order to obtain a scalar correlation coefficient between  $\mathbf{Y}$  and  $\mathbf{X}$  multidimensional trajectories, we propose a modified Coefficient of Determination. Based on Eq. 5, we define  $\text{GMCC}^2$  as

$$\text{GMCC}^2(\mathbf{X}, \mathbf{Y}) = \frac{\sum_{t=0}^T \|\hat{\mathbf{Y}}_t - \bar{\mathbf{Y}}\|^2}{\sum_{t=0}^T \|\mathbf{Y}_t - \bar{\mathbf{Y}}\|^2} \quad (11)$$

where  $\|\cdot\|$  is the euclidean norm and  $\hat{\mathbf{Y}}_t = \mathbf{H}\mathbf{X}$  where  $\mathbf{H}$  is any nonsingular matrix. The  $\text{GMCC}^2$  coefficient from Eq. 11 is presented as the generalization of the Coefficient of Determination. Moreover,  $R^2$  from Eq. 5, can be computed as a particular case of Eq. 11 for unidimensional variable  $y$ .

If Eq. 11 is extended

$$\text{GMCC}^2 = \frac{\sum_{t=0}^T \sqrt{(\hat{y}_{1t} - \bar{y}_1)^2 + \dots + (\hat{y}_{nt} - \bar{y}_n)^2}}{\sum_{t=0}^T \sqrt{(y_{1t} - \bar{y}_1)^2 + \dots + (y_{nt} - \bar{y}_n)^2}}$$

and now, if square root is eliminated by the power in both the numerator and denominator and the numerator split

$$\text{GMCC}^2 = \frac{\sum_{t=0}^T (\hat{y}_{1t} - \bar{y}_1)^2}{\sigma_Y^2} + \dots + \frac{\sum_{t=0}^T (\hat{y}_{nt} - \bar{y}_n)^2}{\sigma_Y^2}$$

where  $\sigma_Y^2 = \sum_{t=0}^T (y_{1t} - \bar{y}_1)^2 + \dots + \sum_{t=0}^T (y_{nt} - \bar{y}_n)^2 = \sigma_{y_1}^2 + \dots + \sigma_{y_n}^2$ . If now, each fraction is multiplied by their respective variance of the dependant variable  $\sigma_{y_i}^2$ , the obtained equation is

$$\text{GMCC}^2 = \frac{\sum_{t=0}^T (\hat{y}_{1t} - \bar{y}_1)^2 \sigma_{y_1}^2}{\sigma_{y_1}^2 \sigma_Y^2} + \dots + \frac{\sum_{t=0}^T (\hat{y}_{nt} - \bar{y}_n)^2 \sigma_{y_n}^2}{\sigma_{y_n}^2 \sigma_Y^2}.$$

For the particular case of the predicted linear models  $\hat{y}_{it}$  been optimal, based on Eq. 7, it can be rewritten as a weighted sum of the square of the Multiple Correlation Coefficient of each variable of  $\mathbf{Y}$

$$\text{GMCC}^2 = \sum_{i=1}^n R_{y_i}^2 \frac{\sigma_{y_i}^2}{\sigma_Y^2} \quad (12)$$

GMCC is the square root of the general case of Coefficient of Determination from Eq. 11 under the particular case of optimal linear prediction model computed in Eq. 12

$$\text{GMCC} = \sqrt{\sum_{i=1}^n R_{y_i}^2 \frac{\sigma_{y_i}^2}{\sigma_Y^2}}. \quad (13)$$

It can be observe, that for the particular case of  $\mathbf{Y}$  being unidimensional, The General Multiple Correlation Coefficient becomes into the Multiple Correlation Coefficient.

As Eq. 12 represents GMCC as a weighted sum of  $R^2$ , we can demonstrate Eq. 10 considering each dimension of  $\mathbf{X}$  individually. Given the sequence of a certain dimension  $X_i$ , if there is linear combination of  $\mathbf{H}\mathbf{X}$  that gives  $X_i$ , then  $R^2$  is going to be 1 for that dimension. If we can find a linear combination for each dimension  $i$ , then all  $R^2$  would be 1 and GMCC also.

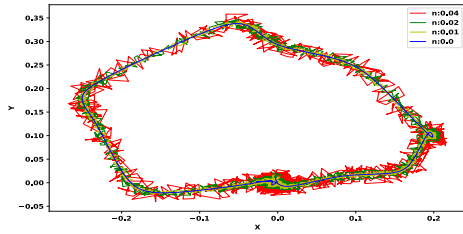
GMCC represents how well a set of independent multidimensional variables can predict a set of dependent multidimensional variables by a linear model. GMCC will take values between 0 and 1, where 1 means a complete prediction of  $\mathbf{Y}_t$  with  $\mathbf{X}_t$ . One of the drawbacks of GMCC in comparison with  $R_v$  and dCor is that it is not symmetric so,  $\text{GMCC}(\mathbf{X}, \mathbf{Y}) \neq \text{GMCC}(\mathbf{Y}, \mathbf{X})$ . In order to solve it, we can compute the symmetric GMCC

$$\text{GMCC}_{sym}(X, Y) = \frac{\text{GMCC}(X, Y) + \text{GMCC}(Y, X)}{2} \quad (14)$$

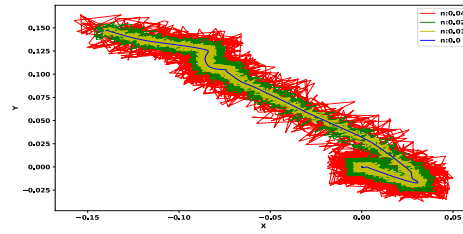
This symmetric correlation coefficient from Eq. 14 is used to compute a similarity distance measure

$$d_{X,Y} = 1 - \text{GMCC}_{sym}(X, Y). \quad (15)$$

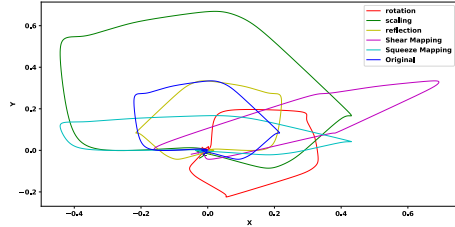
A similarity distance of 1 means that  $\mathbf{X}$  and  $\mathbf{Y}$  are completely uncorrelated and the closer the distance is to 0, the higher the linear correlation between  $\mathbf{X}$  and  $\mathbf{Y}$ .



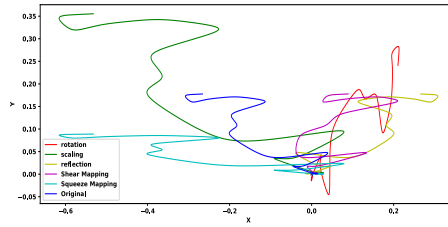
(a) Circular trajectory with artificial gaussian noise



(b) Linear trajectory with artificial gaussian noise



(c) Circular trajectories artificially linearly transformed



(d) Linear trajectories artificially linearly transformed

Fig. 2: Artificially generated trajectories for testing the experiments in Section III-A

In the experimental evaluation, this distance measure is going to be compared with the two most relevant multivariate correlation coefficients metrics  $R_v$  and dCor.

### III. EXPERIMENTAL EVALUATION

For the experimental evaluation of GMCC, the experiments have been split in two sections. In Section III-A, the GMCC characteristics will be studied in some predefined trajectories. This Section will be useful to facilitate the reader the understanding of the capacities and limitations of the novel metric.

In Section III-B, GMCC is applied for a trajectory clustering problem and the obtained solution compared with  $R_v$  and dCor solutions. Six different type of trajectories have been recorded with a KUKA LWR arm. The recorded trajectories have been clustered through a hierarchical clustering method.

#### A. GMCC characteristics

For this first evaluation, two simple 2D signals are going to be used: a linear and a circular trajectory. To observe the robustness of GMCC in front of noisy signals, different variance gaussian noises were added in both the linear and circular trajectories. The obtained noisy trajectories are presented in Fig. 2a and Fig. 2b. The similarity distances

Traj. type	dCor	$R_v$	GMCC
line 0.01	0.0006	0.0	0.001
line 0.02	0.02	0.0	0.005
line 0.04	0.01	0.0	0.02
circle 0.01	0.0	0.0	0.001
circle 0.02	0.0	0.0	0.002
circle 0.04	0.0035	0.0	0.0039

TABLE I: Similarity distance measurements for line and circular trajectories with different noises for dCor,  $R_v$  and GMCC

between the original trajectories and the noisy signals were computed for  $R_v$ , dCor and GMCC and the obtained values are presented in Table I.

From Table I, it can be observed that in terms of noise the three similarity distances are pretty robust. GMCC is giving the worst results in comparison with dCor and  $R_v$ , but from a general perspective, three of them remains robust in front of noisy signals as the obtained values are close to 0.

In a second experiment the selected two trajectories, a line and a circle, were transformed by rotation, scaling, reflection, shear mapping and squeezing. The obtained trajectories are presented in Fig. 2c and Fig. 2d. As the previous case, the similarity distances were computed between the original trajectory and the transformed ones for dCor,  $R_v$  and GMCC and obtained results organized in Tables II and III.

These tables are representative of the real power of GMCC. While  $R_v$  is measuring big distances between the original line and the one transformed by shear mapping, rotation or squeezing, the GMCC remains robust in null distance. dCor's distance is also pretty close to 0, but in shear mapping and squeezing gives a little bit worse results. GMCC is the coefficient that works better for similarity invariant to linear transformations and will remain always in lowest distance for any linear transformations.

GMCC is limited to linear transformations. GMCC dis-

Transf. type	dCor	$R_v$	GMCC
rotation	0.0	0.01	0.0
scale	0.0	0.0	0.0
reflection	0.0	0.009	0.0
shear	0.12	0.18	0.0
squeeze	0.11	0.14	0.0

TABLE II: Similarity distance measurements for circular trajectories with different transformations for dCor,  $R_v$  and GMCC

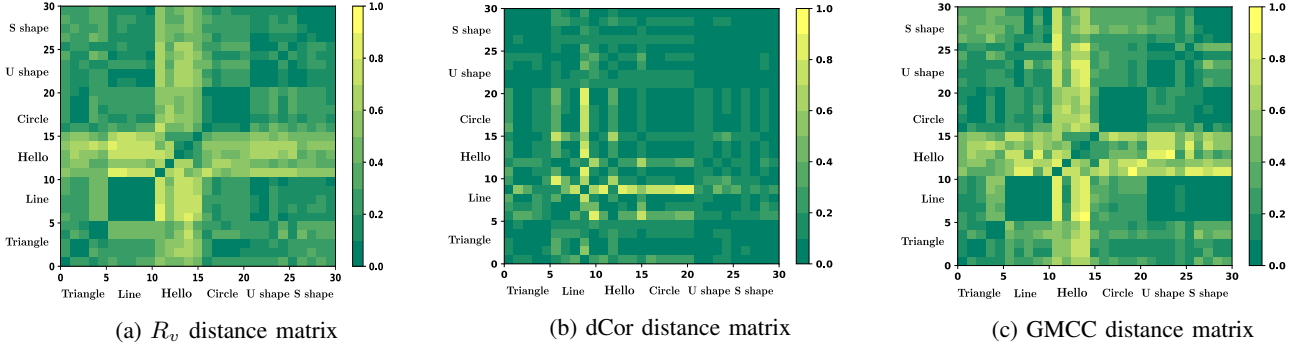


Fig. 3: Distance Matrices computed for  $R_v$ , dCor and GMCC

tance starts increasing when no linear mapping can be set between the compared trajectories. Meanwhile, dCor is capable of finding also nonlinear mappings between trajectories. The distance measured by dCor will be smaller than the distance with GMCC for nonlinear correlations. In order to prove this, the similarity distance has been measured between the original line and the original circle and the obtained distances presented in Table IV.

The distance measured by GMCC is 0.87 while the distance measured by dCor is 0.57. dCor is less sensible to nonlinear transformations than GMCC. Our correlation coefficient is a worse tool for measuring nonlinear correlation but at the same time is a better metric for measuring only linear correlations.

### B. Clustering robot trajectories

Six different type of motions were recorded in a KUKA LWR arm. The end effector's cartesian position was recorded for  $x$ ,  $y$  and  $z$ . For each motion type five different demonstrations were recorded with different scales, different starting points and different orientations. The selected motions are lines, U-shape, S-shape, circular, triangle and hello trajectories. All the demonstrations were fixed to the same time length and same amount of time samples. Then, the similarity distance matrix was computed among the demonstrations and this distance matrix was used as the input for a hierarchical clustering.

Transf. type	dCor	$R_v$	GMCC
rotation	0.0	0.82	0.0
scale	0.0	0.0	0.0
reflection	0.0	0.819	0.0
shear	0.007	0.88	0.0
squeeze	0.0029	0.16	0.0

TABLE III: Similarity distance measurements for linear trajectories with different transformations for dCor,  $R_v$  and GMCC

Transf. type	dCor	$R_v$	GMCC
line Vs circle	0.5258	0.27378	0.871

TABLE IV: Similarity distance measurements for line Vs circle

The similarity matrix of each 30 demonstrations ( six gestures by five demonstrations) was computed for the three correlation distances dCor,  $R_v$  and GMCC. The obtained similarity matrices are presented in the Fig. 3.

The first conclusion is extracted from Fig. 3b. dCor distance is measuring small distances not only between the demonstrations of the same motion but also between the demonstrations of different motions. From here, we can predict that at least with the recorded motions, dCor will fail classifying or clustering them correctly.  $R_v$  and GMCC are obtaining similar results. For some gestures GMCC differentiates better. For example, between circular trajectories and S-shape trajectories. Anyway, GMCC is getting worse results discriminating linear movements and S-shape movements. It is also relevant to remark how badly the hello trajectories are correlated. None  $R_v$  or GMCC is giving good similarity distance between most of the hello trajectories.

The recorded demonstrations were clustered by hierarchical clustering to observe the application of our algorithm in a real case scenario. During first evaluation, we considered to cluster the demonstrations in six clusters. However, the hello trajectories were very dissimilar not only in comparison with other trajectories but also between each other as it can be observed in the dendrograms from Fig. 4. This was making the clustering to fail as the hello trajectories were not clustered in the same cluster, but each one was a unique cluster and then in a unique cluster we could find all the other gestures. In order to solve this, the hierarchical clustering tries to cluster the no-hello motions in five different clusters.

The applied linkage criterion was the average distance between the clusters. The average linkage criterion will use the average distance between all the elements in both cluster to decide if they should merge or not

$$d(u, v) = \sum_{i=1}^n \sum_{j=1}^m \frac{d(u[i], v[j])}{nm}.$$

From the obtained dendrograms several conclusions can be extracted. Expected from the similarity matrix in Fig. 3b, dCor solution is not useful for clustering the recorded human gestures. In Fig. 4b, it can be observed how the clusters has been built with dCor. There are two main clusters. The one in yellow, clustering in same cluster some triangle trajectories,



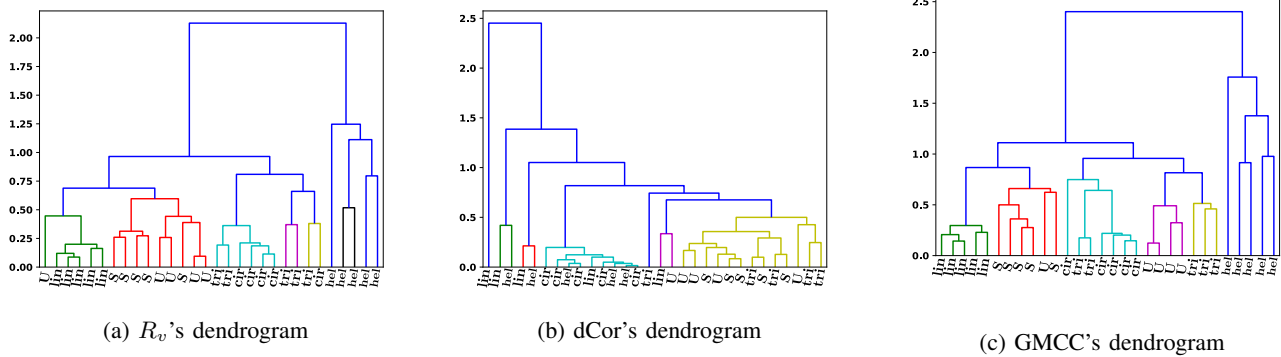


Fig. 4: Dendrograms computed for  $R_v$ , dCor and GMCC

U-shape trajectories and S-shape trajectories. The other big clustering some line trajectories, hello trajectories and circular trajectories. This poor results can be understood as dCor finds not only linear but also non-linear correlations between trajectories and so, it will find a small similarity distance between most of the trajectories as they are nonlinearly correlated.

The clustering obtained in GMCC's dendrogram is better than the clustering obtained in  $R_v$ 's dendrogram. There are less missclustered elements in GMCC than in  $R_v$ . GMCC missclusters two triangle trajectories with the circular trajectories and one U-shape trajectory with the S-shape trajectories. Meanwhile,  $R_v$  dendrogram is also creating a unique cluster between all U-shape and S-shape trajectories. Also, one of the found clusters by  $R_v$  is a cluster with only two elements containing one triangle trajectory and one circular trajectory. We can conclude that the best tool for clustering three-dimensional motions is the GMCC.

#### IV. DISCUSSION AND FUTURE WORK

We presented a multidimensional correlation coefficient that remains invariant for any linear transformation and only linear transformations. This novel multidimensional has been proved to be a multidimensional extension of the Multiple Correlation Coefficient. The validity of GMCC has been proved in a real robot learning scenario and have showed better clustering capacities than state of art multivariate correlation coefficients such as  $R_v$  and dCor.

While our similarity distance is invariant to any linear transformation, it cannot deal with time morphing and time shifting. Our similarity metric has poor results measuring the correlation between two similar in shape but different in velocity trajectories. A promising improvement of our algorithm will come from the mixture between DTW and GMCC. Another interesting improvement comes from extending GMCC to probabilistic motions. ProMP are a promising representation of motions in Imitation Learning and we expect to merge ProMPs with GMCC.

#### REFERENCES

- [1] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, J. Peters, *et al.*, "An algorithmic perspective on imitation learning," *Foundations and Trends® in Robotics*, vol. 7, no. 1-2, pp. 1–179, 2018.
- [2] S. Schaal, "Learning from demonstration," in *Advances in neural information processing systems*, pp. 1040–1046, 1997.
- [3] O. Dermy, A. Paraschos, M. Ewerton, J. Peters, F. Chappillet, and S. Ivaldi, "Prediction of intention during interaction with icub with probabilistic movement primitives," *Frontiers in Robotics and AI*, vol. 4, p. 45, 2017.
- [4] H. B. Amor, O. Kroemer, U. Hillenbrand, G. Neumann, and J. Peters, "Generalization of human grasping for multi-fingered robot hands," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2043–2050, IEEE, 2012.
- [5] M. A. Goodrich, A. C. Schultz, *et al.*, "Human-robot interaction: a survey," *Foundations and Trends® in Human-Computer Interaction*, vol. 1, no. 3, pp. 203–275, 2008.
- [6] H. B. Amor, D. Vogt, M. Ewerton, E. Berger, B. Jung, and J. Peters, "Learning responsive robot behavior by imitation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3257–3264, IEEE, 2013.
- [7] G. Maeda, M. Ewerton, R. Lioutikov, H. B. Amor, J. Peters, and G. Neumann, "Learning interaction for collaborative tasks with probabilistic movement primitives," in *2014 IEEE-RAS International Conference on Humanoid Robots*, pp. 527–534, IEEE, 2014.
- [8] H. B. Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *2014 IEEE international conference on robotics and automation (ICRA)*, pp. 2831–2837, IEEE, 2014.
- [9] A. Paraschos, C. Daniel, J. R. Peters, and G. Neumann, "Probabilistic movement primitives," in *Advances in neural information processing systems*, pp. 2616–2624, 2013.
- [10] S. Gomez-Gonzalez, G. Neumann, B. Schölkopf, and J. Peters, "Adaptation and robust learning of probabilistic movement primitives," *arXiv preprint arXiv:1808.10648*, 2018.
- [11] X. Wang, M. Xia, H. Cai, Y. Gao, and C. Cattani, "Hidden-markov-models-based dynamic hand gesture recognition," *Mathematical Problems in Engineering*, vol. 2012, 2012.
- [12] H. Hotelling, "Relations between two sets of variates," in *Breakthroughs in statistics*, pp. 162–190, Springer, 1992.
- [13] Y. Escoufier, "Le traitement des variables vectorielles," *Biometrics*, pp. 751–760, 1973.
- [14] G. J. Székely, M. L. Rizzo, *et al.*, "Brownian distance covariance," *The annals of applied statistics*, vol. 3, no. 4, pp. 1236–1265, 2009.
- [15] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.
- [16] L. Chen, M. T. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pp. 491–502, ACM, 2005.
- [17] T. Eiter and H. Mannila, "Computing discrete fréchet distance," tech. rep., Citeseer, 1994.
- [18] J. Henrikson, "Completeness and total boundedness of the hausdorff metric,"
- [19] P.-F. Marteau, "Time warp edit distance," *arXiv preprint arXiv:0802.3522*, 2008.
- [20] M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied Linear Statistical Models Fifth Edition*.