Arc-Length-Based Warping for Robot Skill Synthesis from Multiple Demonstrations: Metrics Appendix

Giovanni Braglia, Davide Tebaldi,
André E. Lazzaretti, Luigi Biagiotti

1 Introduction

This report comes as additional material to the paper entitled "Arc-Length-Based Warping for Robot Skill Synthesis from Multiple Demonstrations" by G.Braglia, D.Tebaldi, A.Lazzaretti and L.Biagiotti. The aim is to provide a detailed explanation of the metrics used in Section IV *Experiments and Evaluation*, which was not included in the paper because of page limit constraints. In the explanation, we will always consider two generic series $\mathbf{x} \in \mathbf{R}^{d\times N}$ and $\mathbf{y} \in \mathbf{R}^{d\times M}$, with d being the dimension, N and M being respectively the number of samples of $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N]^\top$ and $\mathbf{y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_N]^\top$. While the explanation is given for the dyadic case, i.e. considering two series, the metrics can be easily extended to the multivariate case as explained in [1].

2 Alignment Algorithms Comparisons Based on LASA Handwriting Dataset

2.1 Rho (ρ)

For the definition of ρ we used the following as reference [2]. The metrics ρ was introduced to assess group level synchrony of time-series by analyzing their phase. It does so by calculating an aggregate relative phase across multiple signals to measure how closely each signal's phase aligns with the overall phase. Note that the ρ metric was introduced for uni-dimensional series; as we are working with d-dimensional series, without loss of consistency we can calculate ρ for each individual dimension and then average its value. We will thus indicate with $\boldsymbol{x} = [x_1, x_2, ..., x_N]^\top$ ($\boldsymbol{y} = [y_1, y_2, ..., y_N]^\top$) a generic one-dimensional time-serie selected by its d-dimensional counterpart $\mathbf{x}$ ($\mathbf{y}$). Moreover, this metric is proposed for time series of the same length, i.e. $N = M$. Some re-sampling should be introduced in case this condition is not met.

To compute ρ we start first by compute the phase time-series $\theta_{\boldsymbol{x}}, \theta_{\boldsymbol{y}} \in [-\pi, +\pi]$ from, respectively, $\boldsymbol{x}$ and $\boldsymbol{y}$. Then the group phase time-serie q_t at instant t, or cluster serie, is calculated as:

$$q_t = \operatorname{atan2}(\hat{q}_t), \quad \text{with} \ \ \hat{q}_t = \frac{1}{2}\left[\exp(i\theta_{\boldsymbol{x},t}) + \exp(i\theta_{\boldsymbol{y},t})\right], \tag{1}$$

where i is the imaginary number and atan2 is the 2-argument arc-tangent; note that q_t is still in radian, i.e. $q_t \in [-\pi, +\pi]$.
Finally, we proceed by compute the phase cluster-displacement ϕ for each component, that is:

$$\phi_{\boldsymbol{x},t} = \theta\boldsymbol{x}, t - q_t \ \ \text{and} \ \ \phi_{\boldsymbol{y},t} = \theta\boldsymbol{y}, t - q_t, \tag{2}$$

then compute its mean value $\bar{\phi}$ as

$$\begin{aligned} \bar{\phi}_{\boldsymbol{x},t} &= \operatorname{atan2}(\hat{\phi}_{\boldsymbol{x},t}), \quad \text{with} \ \ \hat{\phi}_{\boldsymbol{x},t} = \frac{1}{N}\sum_{t=1}^{N} \exp(i\phi_{\boldsymbol{x},t}), \quad \text{and} \\ \bar{\phi}_{\boldsymbol{y},t} &= \operatorname{atan2}(\hat{\phi}_{\boldsymbol{y},t}), \quad \text{with} \ \ \hat{\phi}_{\boldsymbol{y},t} = \frac{1}{N}\sum_{t=1}^{N} \exp(i\phi_{\boldsymbol{y},t}). \end{aligned} \tag{3}$$

Finally, one can obtain calculate the degree of synchronization ρ of the $(\boldsymbol{x}, \boldsymbol{y})$ dyad by first calculating its instantaneous value ρ_t as:

$$\rho_t = \left|\frac{1}{2}\left\{\exp[i(\phi_{\boldsymbol{x},t} - \bar{\phi}_{\boldsymbol{x}})] + \exp[i(\phi_{\boldsymbol{y},t} - \bar{\phi}_{\boldsymbol{y}})]\right\}\right|, \tag{4}$$

then calculating ρ as the average $\rho = 1/N \sum_{t=1}^{N}$. Note that $\rho \in [0, 1]$ assumes higher values for a larger degree of group synchronization.

2.2 Entropy (H)

For the definition of H we used the following reference [3]. Entropy represents the average level of unpredictability of outcomes drawn from a probability distribution. A distribution with low entropy indicates that outcomes are more predictable, while higher entropy reflects greater uncertainty and unpredictability.

Given a discrete random variable X, which takes values in the set $\mathcal{X}$ and is distributed according to $p : \mathcal{X} \to [0, 1]$ such that $p(x) := \boldsymbol{P}[X = x]$, the entropy H of X can be computed as:

$$H(X) = \boldsymbol{E}[I(X)] = \boldsymbol{E}[-\log p(X)], \tag{5}$$

where $\boldsymbol{E}$ is the expected value operator, and $I(X)$ describes the information content of X and is itself a random variable. One can write H explicitly as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x), \tag{6}$$

denoting with b the base of the used logarithm (in the experimental tests we used $b = 10$). Without loss of consistency we can use as random variable X

a sample $\mathbf{x}_t$ of the serie $\mathbf{x}$, the same for $\mathbf{y}$. Moreover, as we are considering a dyad of series, one can define the conditional entropy of two variables $\mathbf{x}_t$ and $\mathbf{y}_t$ taking values from $\mathbf{x}$ and $\mathbf{y}$ respectively, as:

$$H(\mathbf{x}|\mathbf{y}) = -\sum_{i=1}^{N}\sum_{j=1}^{M} p_{\mathbf{x},\mathbf{y}}(\mathbf{x}_i, \mathbf{y}_j) \log \frac{p_{\mathbf{x},\mathbf{y}}(\mathbf{x}_i, \mathbf{y}_j)}{p_{\mathbf{y}}(\mathbf{y}_j)}, \tag{7}$$

where $p_{\mathbf{x},\mathbf{y}}(\mathbf{x}_i, \mathbf{y}_j) := \boldsymbol{P}[\mathbf{x} = \mathbf{x}_i, \mathbf{y} = \mathbf{y}_j]$ and $p_{\mathbf{y}}(\mathbf{y}_j) = \boldsymbol{P}[\mathbf{y} = \mathbf{y}_j]$.

2.3 Sum-Normalized Cross-Spectral Density (snCSD)

For the definition of snCSD we used the following reference [1]. This metric was proposed by the authors as an alternative to the averaged coherence metric [4] when dealing with noisy signals. An important consideration for the averaged coherence metric is that it averages coherence values across frequencies, ignoring the relative amplitude at different frequencies. This can be problematic for signals with Gaussian noise, which affects all frequencies, while the actual signal may be concentrated in a limited frequency range. In such cases, noise can dominate the meaningful signal, especially with high sampling rates that increase the number of frequency components. Applying filters, like a bandpass filter, could improve the metric's reliability by removing irrelevant noise frequencies.

However, noise can significantly impact the results if no filtering is applied. To address this the metric snCSD, instead of normalizing at each frequency, postpones normalization after summing the cross-spectral density (CSD) across frequencies. This allows amplitude information in the CSD to influence the final result. Specifically, snCSD$\in [0, 1]$ and can be defined as follows [1]:

$$\text{snCSD} = \frac{\sum_n |CSD_n(\mathbf{x}, \mathbf{y})| \cdot |CSD_n(\mathbf{x}, \mathbf{y})|}{\sum_n |CSD_n(\mathbf{x}, \mathbf{x})| \cdot |CSD_n(\mathbf{y}, \mathbf{y})|}, \tag{8}$$

where the summation is made over each n-frequency component of the CSD. Note that snCSD$\in [0, 1]$ tends to assume higher values every time $\mathbf{x}$ can accurately approximate $\mathbf{y}$, and vice-versa, therefore one can expect snCSD$\rightarrow 1$ if the signals are well synchronized.

2.4 ℓ_2 norm error

The ℓ_2 norm error is a measure of approximation accuracy. Given two series $\mathbf{x}$ and $\mathbf{y}$ of the same length, i.e. $N = M$ in this case, this metric computes the sum of each error $||\mathbf{x}_t - \mathbf{y}_t||_2$, that is:

$$\ell_2 = \sum_{t=1}^{N} ||\mathbf{x}_t - \mathbf{y}_t||_2, \tag{9}$$

where $||\cdot||_2$ is the Euclidean norm.

3 Skill Synthesis from Demonstrations

3.1 Haussdorff distance d_H

For the definition of d_H we used the following reference [5]. In particular, one can define d_H as:

$$d_H = \max \left\{ \sup_{\mathbf{x}_t \in \mathbf{x}} d(\mathbf{x}_t, \mathbf{y}), \quad \sup_{\mathbf{y}_t \in \mathbf{y}} d(\mathbf{x}, \mathbf{y}_t) \right\}, \tag{10}$$

where $d(\cdot, \cdot)$ can be defined as:

$$d(\mathbf{x}_t, \mathbf{y}) = \inf_{\mathbf{x}_t \in \mathbf{x}} ||\mathbf{x}_t - \mathbf{y}_t|| \text{ and } d(\mathbf{x}, \mathbf{y}_t) = \inf_{\mathbf{y}_t \in \mathbf{y}} ||\mathbf{x}_t - \mathbf{y}_t||. \tag{11}$$

3.2 Dynamic Time Warping (DTW) score d_{DTW}

For the definition d_{DTW} we used the following reference [6]. Given two generic series $\mathbf{x}$ and $\mathbf{y}$, d_{DTW} is defined as:

$$d_{DTW}(\mathbf{x}, \mathbf{y}) = \min_{\pi \in \mathcal{A}(\mathbf{x}, \mathbf{y})} \sum_{(i,k) \in pi} d(\boldsymbol{x}_i, \boldsymbol{y}_k), \tag{12}$$

where $\mathcal{A}(\mathbf{x}, \mathbf{y})$ is the set of all possible alignments between $\mathbf{x}$ and $\mathbf{y}$, while $d(\mathbf{x}_i, \mathbf{y}_k) = ||\mathbf{x}_i - \mathbf{y}_k||^2$ is the squared Euclidean distance.

References

[1] D. Hudson, T. J. Wiltshire, and M. Atzmueller, "multisyncpy: A python package for assessing multivariate coordination dynamics," *Behavior Research Methods*, vol. 55, no. 2, pp. 932–962, 2023.

[2] M. J. Richardson, R. L. Garcia, T. D. Frank, M. Gergor, and K. L. Marsh, "Measuring group synchrony: a cluster-phase method for analyzing multivariate movement time-series," *Frontiers in physiology*, vol. 3, p. 405, 2012.

[3] R. H. Stevens and T. L. Galloway, "Toward a quantitative description of the neurodynamic organizations of teams," *Social Neuroscience*, vol. 9, no. 2, pp. 160–173, 2014.

[4] R. E. White, "Signal and noise estimation from seismic reflection data using spectral coherence methods," *Proceedings of the IEEE*, vol. 72, no. 10, pp. 1340–1356, 1984.

[5] F. Donoso, J. Bustos-Salas, M. Torres-Torriti, and A. Guesalaga, "Mobile robot localization using the hausdorff distance," *Robotica*, vol. 26, pp. 129–141, 03 2008.

[6] M. Müller, "Dynamic time warping," *Information retrieval for music and motion*, pp. 69–84, 2007.