

MOX–Report No. 42/2013

Analysis of Juggling Data: an Application of K-mean Alignment

BERNARDI, M.; SANGALLI, L.M.; SECCHI, P.; VANTINI, S.

MOX, Dipartimento di Matematica “F. Brioschi”
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it

<http://mox.polimi.it>

Analysis of Juggling Data: an Application of K-mean Alignment

Mara Bernardi, Laura M. Sangalli, Piercesare Secchi and Simone Vantini

MOX - Department of Mathematics, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano, Italy
marasabina.bernardi@mail.polimi.it, laura.sangalli@polimi.it,
piercesare.secchi@polimi.it, simone.vantini@polimi.it

Keywords: K-mean alignment, Registration, Functional clustering, Juggling Trains

Abstract

We analyze the juggling data by means of the k -mean alignment algorithm using cycles as the experimental units of the analysis. Allowing for affine warping, we detect two clusters distinguishing between mainly-planar trajectories and trajectories tilted toward the body of the juggler in the lower part of the cycle. In particular we detect an anomalous presence of tilted trajectories among the record third cycles. We also find warping functions to be clustered according to records suggesting that each record is performed at a different pace and thus associated to a different typical cycle-duration.

1 Pre-processing

We analyzed the juggling data described in [Ramsay et al. \(2013b\)](#) from two perspectives: records as experimental units of the analysis and cycles - composing each record - as experimental units of the analysis. In this manuscript we focus on the second approach and thus we deal with 113 three-dimensional curves indicating different trajectories of the juggler's right hand forefinger. We arbitrarily indicate as *cycle* the period between two subsequent releases of a ball and, for biological reasons, we identify the moment when a ball is released with the moment when the tangential acceleration is maximal. These moments identify the end of a cycle and the beginning of the following one. As an example, in [Figure 1](#) we report the evolution across time of the tangential acceleration for the first record. Colored stars represent stationary points. In particular, blue stars represent the ones used to cut the record into cycles.

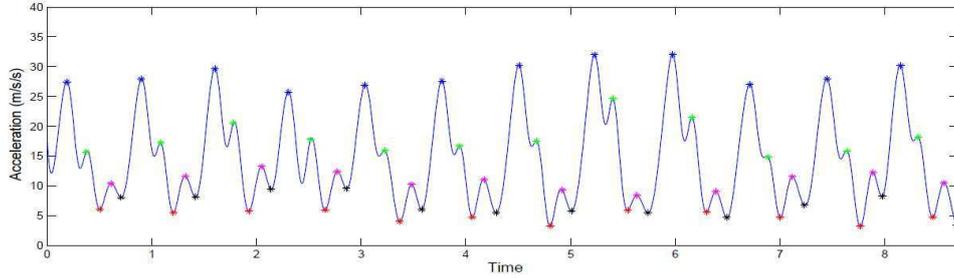


Figure 1: Tangential acceleration of the first record.

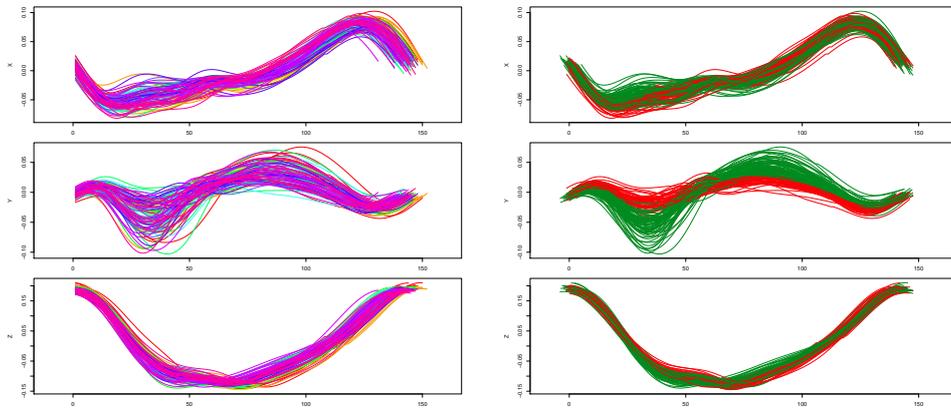


Figure 2: Left panels: the 113 cycles. The three panels show the X, Y, and Z components of the curves, respectively. Right panels: Aligned and clustered cycles provided by the k -mean alignment algorithm performed with affine warping functions and $k = 2$. The three panels show the X, Y, and Z components of the aligned cycles. The cycles of the first cluster are in green, while those of the second cluster are in red.

This procedure was applied to all ten records thus obtaining a set of 113 cycles. As suggested in [Ramsay et al. \(2013b\)](#), the origin of each cycle has been set at the time when the tangential acceleration is maximal, i.e., the starting points of the 113 curves. Note that all cycles present different durations and thus the final time instants differ across cycles. In the left panels of [Figure 2](#) the X, Y, and Z components of the 113 cycles are reported.

2 K-mean Alignment

To look for clusters of trajectories in presence of phase variability we applied the k -mean alignment algorithm, detailed in [Sangalli et al. \(2010\)](#) and summarized in [Sangalli et al. \(2013\)](#), to the 113 trajectories. Since data pre-

processing centered and partially rotated data but did not rescaled them, we will consider two cycles as similar if they are identical up to a multiplicative factor along each component. Therefore, we shall use the following similarity index:

$$\rho(f_i, f_j) = \frac{1}{3} \sum_{p=1}^3 \frac{\int f_{ip}(t) f_{jp}(t) dt}{\sqrt{\int f_{ip}(t)^2 dt} \sqrt{\int f_{jp}(t)^2 dt}}, \quad (1)$$

where f_i and f_j are cycles. Indeed, this similarity index assigns similarity equal to 1 (its maximal value) to couples of curves that differ only for a positive multiplying factor along each component:

$$\rho(f_i, f_j) = 1 \Leftrightarrow \forall p \in \{1, 2, 3\} \exists a_p \in \mathbb{R}^+ : f_{ip}(t) = a_p f_{jp}(t). \quad (2)$$

Since the physical phenomenon does not suggest any particular group of warping functions to be the best suited to the analysis, we run the analysis using different groups of warping functions coherent with the previous similarity index:

$$\begin{aligned} \mathcal{H}_{\text{affine}} &= \{h : h(t) = mt + q \text{ with } m \in \mathbb{R}^+, q \in \mathbb{R}\}, \\ \mathcal{H}_{\text{shift}} &= \{h : h(t) = t + q \text{ with } q \in \mathbb{R}\}, \\ \mathcal{H}_{\text{dilation}} &= \{h : h(t) = mt \text{ with } m \in \mathbb{R}^+\}, \\ \mathcal{H}_{\text{identity}} &= \{h : h(t) = t\}. \end{aligned}$$

The analysis here presented has been performed using `fdakma` R package downloadable from CRAN ([Patriarca et al. \(2013\)](#)).

The left panel of Figure 3 shows the results of the k -mean alignment algorithm applied with different choices for the number k of clusters and the group \mathcal{H} of warping functions. For each couple (k, \mathcal{H}) the mean similarity between the aligned curves and their respective templates is reported. The gray dot on the left represents the mean similarity between the unaligned curves and their mean which acts as a lower bound for the algorithm performance. The mean similarities achieved by using $\mathcal{H}_{\text{affine}}$, $\mathcal{H}_{\text{shift}}$, $\mathcal{H}_{\text{dilation}}$, and $\mathcal{H}_{\text{identity}}$ are reported in orange, blue, green, and black, respectively. Note that, as already pointed out in [Sangalli et al. \(2010\)](#) and in [Sangalli et al. \(2013\)](#), running the k -mean alignment without allowing for warping (i.e, choosing $\mathcal{H}_{\text{identity}}$) is equivalent to perform a simple functional k -mean clustering, while setting $k = 1$ is equivalent to perform a simple continuous alignment with just one template. As described in [Sangalli et al. \(2010\)](#) and in [Sangalli et al. \(2013\)](#), being the curves not defined on the entire real axis, the integrals in (1) are computed over the intersection of the domains of \mathbf{f}_i and \mathbf{f}_j , and the cluster templates are estimated by means of local polynomial regression.

The vertical displacement between the black curve and the others suggests the presence of phase variability. In detail, the slightly higher performances achieved using the group of affinity (i.e., orange curve) seem to suggest this latter group as the most suitable. This choice is consistent with a possible wrong detection of the starting point of the cycle and a possible different velocity across cycles. The right panel of Figure 3 details the orange curve by showing the boxplots of the similarities between each aligned curve and its respective template. Focussing on the orange curve, the elbow observed for $k = 2$ suggests the existence of two clusters. No significant gain in the mean similarity is indeed obtained by introducing an extra cluster (i.e., $k = 3$). We thus now focus on the clustering and alignment obtained by setting the number of clusters equal to 2 (i.e., $k = 2$) and by considering positive affinities (i.e., $\mathcal{H}_{\text{affine}}$) as the group of warping functions. The three right panels of Figure 2 show the X, Y, and Z components of the aligned cycles. The cycles of the first cluster are colored in green, while those of the second cluster in red.

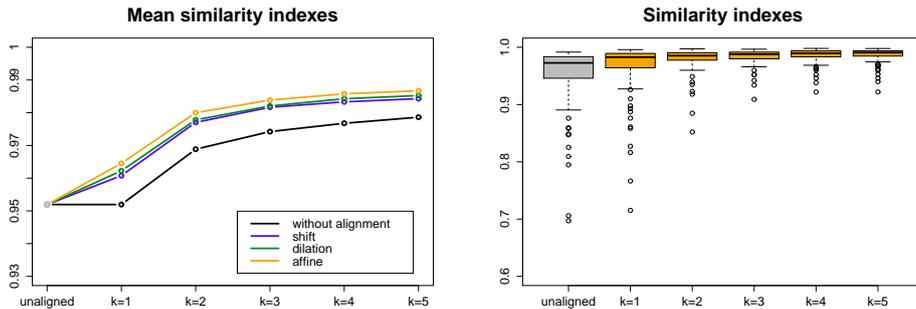


Figure 3: Performance of the k -mean alignment algorithm. The left panel shows the mean similarity between the aligned curves and their respective templates obtained with different values of k and different classes of warping functions. The right panel displays the boxplots of the similarity indexes between the aligned curves and their respective templates with the group of warping functions $\mathcal{H}_{\text{affine}}$ and different values of k .

3 Analysis of the clusters

Differences between the two clusters are observed in all three components, though the major ones pertain to the front-back component (i.e., Y-axis) where the trajectories of the first cluster (i.e, green) present a more oscillating behavior. As it is even clearer in the left panel of Figure 4, the trajectories of the second cluster (i.e, red) are mainly planar curves in the

X-Z plane, while the trajectories of the first cluster are tilted toward the body of the juggler in the lower part of the cycle. The same differences are captured by the two template curves reported in the right panel of Figure 4.

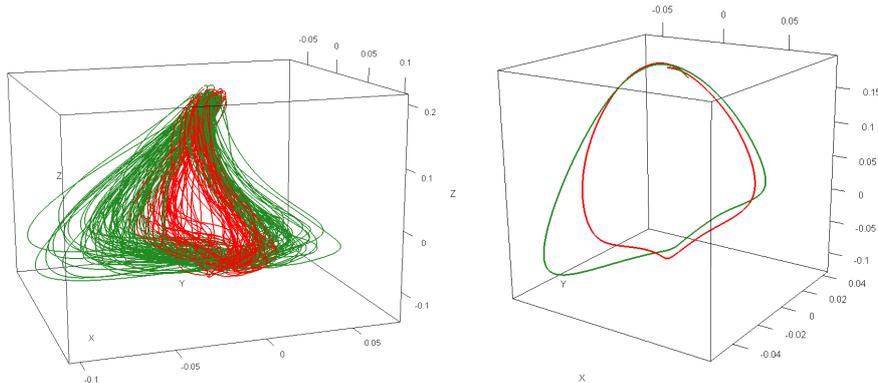


Figure 4: Left panel: 3D plot of the cycles colored according to the clusters provided by the k -mean alignment algorithm performed with affine warping functions and $k = 2$. The cycles of the first cluster are in green, while those of the second cluster are in red. Right panel: the corresponding template curves.

We now want to test if there is any relation between cluster assignments and records or cluster assignments and position in the sequence. Figure 5 shows cluster assignment for each cycle. Rows are associated to records and columns to positions in the sequence. In Figure 6 we report confidence intervals for the proportion of cycles belonging to the second cluster across records (right panel) and across positions in the sequence (left panel). We found no statistical differences among the proportions across records (i.e., curves of each clusters seem randomly spread across records), while we found a higher proportion of curves belonging to the second cluster among the third cycles of each record (i.e., third column in Figure 5). Note that in the third cycle for the first time the juggler needs to catch and then throw a ball with his right hand. Indeed, the first two cycles are warm-up cycles in which no ball is caught, and only balls previously handled by the juggler are thrown.

4 Analysis of the warping functions

We now focus on the analysis of warping functions and in particular on possible associations between warping functions and records or warping functions and position in the sequence. Analyzing the mean shift and dilation across records and positions in the sequence (Figure 7) we found an opposite scenario with respect to the one illustrated in the previous section.

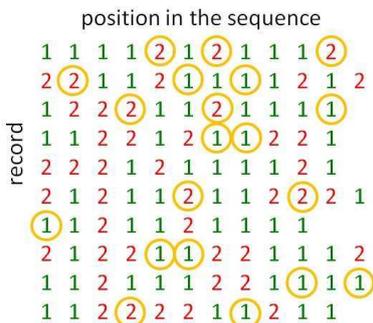


Figure 5: Clustering assignments across records (i.e., rows) and positions in the sequence (i.e., columns). Mismatches between our classification and the one obtained by [Lu and Marron \(2013\)](#) are pointed out in orange.

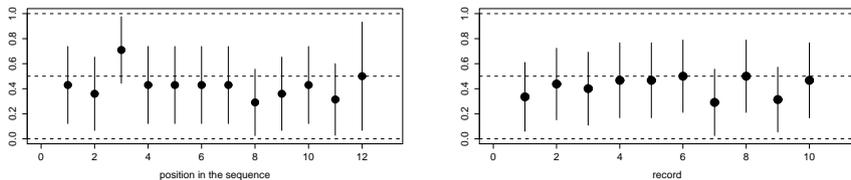


Figure 6: Confidence (95%) intervals for the proportion of cycles belonging to the second clusters across records (right panel) and across position in the sequence (left panel).

Indeed we found no association between warping functions and position in the sequence, while we observed significant differences across records. As an example, in [Figure 8](#) we point out with different colors over the gray background the warping function associated to cycles belonging to records 1, 9, 2, and 5, respectively. Indeed, all warping functions of record 1 have negative intercepts, while those of record 9 have positive intercepts: the cycles of record 1 have been anticipated to be aligned to the other ones, while those of record 9 have been delayed. All warping functions of record 2 have slopes lower than 1, while those of record 5 have slopes greater than one: the cycles of record 2 have been sped up to be aligned, while those of record 5 have been slowed down. Our analysis suggests that each record has been performed at a different pace (i.e., cycle durations are homogeneous within records but not across records).

5 Discussion

Our findings pertaining to the amplitude variability agree with [Lu and Marron \(2013\)](#), [Kurttek et al. \(2013\)](#), and [Poss and Wagner \(2013\)](#). Our cluster-

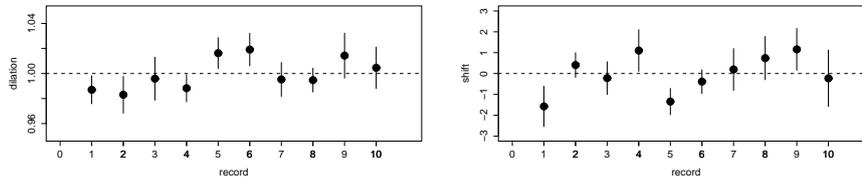


Figure 7: Confidence (95%) intervals for the mean of the dilations (left panel) and the shifts (right panel) of the warping functions.

ing structure is indeed very similar to the one detected in [Lu and Marron \(2013\)](#): only 20 cycles among the 113 have been assigned to different clusters by the two analyses. In [Figure 5](#) the mismatches are pointed out by means of orange circles. The differences between the two clusters shown in the right panels of [Figures 2 and 4](#) are also the same gathered by the first functional principal component detected in both [Kurtek et al. \(2013\)](#) and [Poss and Wagner \(2013\)](#). Moreover, [Poss and Wagner \(2013\)](#) look for possible association between scores and the position in the sequence. Since their first principal component seems to be related to our clustering structure, it would be interesting to know if they find an anomaly in the first principal component score distribution in correspondence of third cycles analogous to the one we find for the cluster assignment.

On the contrary, with respect to phase variability, our results disagree with the ones reported in [Kurtek et al. \(2013\)](#) and [Ramsay et al. \(2013a\)](#). In details, [Kurtek et al. \(2013\)](#) point out a “compensation” effect within each cycle, suggesting that cycles that started faster then slowed down or cycles that started slower then sped up. Possibly, because of the affine warping we used, we did not detect any similar effect. Finally, our analysis of the warping functions suggests that cycle durations are homogeneous within records but not across records. This contrasts with the “return to the base frequency” effect reported in [Ramsay et al. \(2013a\)](#) where early cycles at the beginning of each record are found to be followed by late cycles at the end of the record.

Acknowledgements

All authors are grateful to MBI Mathematical Biosciences Institute <http://mbi.osu.edu/> for support. L.M. Sangalli acknowledges funding by the research program Dote Ricercatore Politecnico di Milano - Regione Lombardia, project: Functional data analysis for life sciences, and by MIUR Ministero dell’Istruzione dell’Università e della Ricerca, *FIRB Futuro in Ricerca* starting grant SNAPLE: Statistical and Numerical methods for the Analysis of Problems in Life sciences and Engineering <http://mox.polimi.it/users/sangalli/firbSNAPLE.html>.

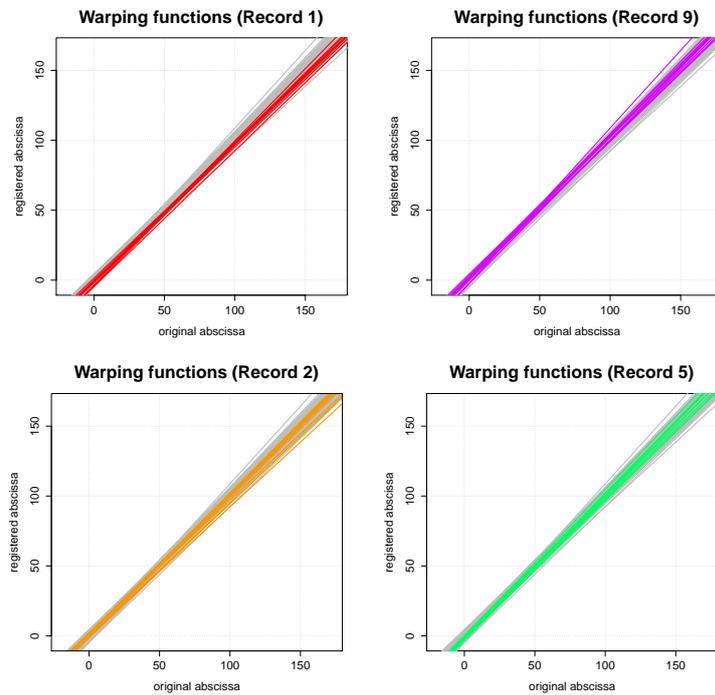


Figure 8: Warping function associated to cycles belonging to records 1, 9, 2, and 5, respectively.

References

- Kurtek, S., Xie, Q., and Srivastava, A. (2013), “Analysis of Juggling Data: Alignment, Extraction, and Modeling of Juggling Cycles,” *Special Section EJS*.
- Lu, X. and Marron, J. (2013), “Analysis of Juggling Data: object oriented data analysis of clustering in acceleration functions,” *Special Section EJS*.
- Patriarca, M., Sangalli, L. M., Secchi, P., Vantini, S., and Vitelli, V. (2013), *fdakma: Clustering and alignment of a given set of curves*, r package version 1.0.
- Poss, D. and Wagner, H. (2013), “Analysis of juggling data: Registering Data to Principal Components to explain Amplitude Variation,” *Special Section EJS*.
- Ramsay, J., Gribble, P., and Kurtek, S. (2013a), “Analysis of juggling data: landmark and continuous registration of juggling trajectories,” *Special Section EJS*.

- (2013b), “Description and processing of functional data arising from juggling trajectories,” *Special Section EJS*.
- Sangalli, L. M., Secchi, P., and Vantini, S. (2013), “Analysis of AneuRisk65 data: K-mean Alignment,” *Special Section EJS*.
- Sangalli, L. M., Secchi, P., Vantini, S., and Vitelli, V. (2010), “K-mean alignment for curve clustering,” *Computational Statistics and Data Analysis*, 54, 1219–1233.

MOX Technical Reports, last issues

Dipartimento di Matematica “F. Brioschi”,
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 42/2013** BERNARDI, M.; SANGALLI, L.M.; SECCHI, P.; VANTINI, S.
Analysis of Juggling Data: an Application of K-mean Alignment
- 41/2013** BERNARDI, M.; SANGALLI, L.M.; SECCHI, P.; VANTINI, S.
Analysis of Proteomics data: Block K-mean Alignment
- 40/2013** PACCARINI P.; ROZZA G.
Stabilized reduced basis method for parametrized advection-diffusion PDEs
- 39/2013** IEVA, F.; MARRA, G.; PAGANONI, A.M.; RADICE, R.
A semiparametric bivariate probit model for joint modeling of outcomes in STEMI patients
- 38/2013** BONIZZONI, F.; NOBILE, F.
Perturbation analysis for the Darcy problem with log-normal permeability
- 35/2013** CATTANEO, L.; FORMAGGIA, L.; IORI G. F.; SCOTTI, A.; ZUNINO, P.
Stabilized extended finite elements for the approximation of saddle point problems with unfitted interfaces
- 36/2013** FERRAN GARCIA, LUCA BONAVENTURA, MARTA NET, JUAN SANCHEZ
Exponential versus IMEX high-order time integrators for thermal convection in rotating spherical shells
- 37/2013** LEVER, V.; PORTA, G.; TAMELLINI, L.; RIVA, M.
Characterization of basin-scale systems under mechanical and geochemical compaction
- 33/2013** MENAFOGLIO, A; GUADAGNINI, A; SECCHI, P
A Kriging Approach based on Aitchison Geometry for the Characterization of Particle-Size Curves in Heterogeneous Aquifers
- 34/2013** TAVAKOLI, A.; ANTONIETTI, P.F.; VERANI, M.
Automatic computation of the impermeability of woven fabrics through image processing