

Analysis of juggling data: Alignment, extraction, and modeling of juggling cycles*

Sebastian Kurtek

Department of Statistics
The Ohio State University, Columbus, OH 43210
e-mail: kurtek.1@stat.osu.edu

Qian Xie and Anuj Srivastava

Department of Statistics
Florida State University, Tallahassee, FL 32306
e-mail: qxie@stat.fsu.edu; anuj@stat.fsu.edu

Abstract: In this paper we present results from alignment, extraction, and statistical analysis of juggling trajectories using an elastic functional data analysis framework. This framework, specifically adapted for analyzing cyclostationary signals using an elastic Riemannian metric, was introduced recently by Kurtek et al. [2]. It relies on a special representation of curves called the square-root velocity function to pose the alignment problem as an optimization over the re-parametrization space. The cost function for alignment is a proper metric and is used to separate phase and amplitude components of juggling cycles. We present results of segmenting juggling trials into cycles, separating phase and amplitude components of cycles, and developing principal component analysis (PCA) based statistical models for these individual components.

Keywords and phrases: Cyclostationary processes, elastic functional data analysis, phase-amplitude separation, cycle extraction.

Received August 2013.

1. Introduction

In this paper we present results on using elastic functional data analysis for performing statistical analysis of the juggling data described earlier in the data introduction paper by Ramsay et al. [3]. As described in that paper, this dataset consists of measurements of the 3D position of the tip of the right index finger of the juggler as he juggles three balls. This includes a total of ten juggling trajectories with approximately 11–13 cycles per trial. Since the juggler’s motion is repetitive, it is natural to treat the trials as observations of a cyclostationary process. A cyclostationary process is a stationary stochastic process whose underlying probability distribution is cyclic in time. Then, the goal of the analysis

*Main article [10.1214/14-EJS937](https://doi.org/10.1214/14-EJS937).

is to formulate a statistical model for the basic unit, a single cycle, and estimate it from the observed data.

The general approach used in this paper was developed by Srivastava et al. [4, 1, 5] and has been summarized in the accompanying paper on the analysis of AneuRisk65 data by Xie et al. [6]. The adaption of this framework to analysis of cyclostationary signals was done by Kurtek et al. [2] and is used here directly. We simply remind the reader that this approach is based on the use of a special function called the square-root velocity function (SRVF) of curves. For a curve $f : [0, 1] \rightarrow \mathbb{R}^3$, its SRVF is defined by a function $q : [0, 1] \rightarrow \mathbb{R}^3$ as $q(t) = \dot{f}(t)/\sqrt{|\dot{f}(t)|}$ [4]. In case of real-valued functions $g : [0, 1] \rightarrow \mathbb{R}$, the SRVF takes the form $q(t) = \text{sign}(\dot{g}(t))\sqrt{|\dot{g}(t)|}$ and is termed the square-root slope function (SRSF). For any two curves, f_1 and f_2 , represented by their SRVF functions q_1 and q_2 , the alignment or registration problem is formulated as:

$$\inf_{h \in \mathcal{H}} \|q_1 - (q_2 \circ h)\sqrt{\dot{h}}\|, \quad (1.1)$$

where $h \in \mathcal{H}$ is a re-parameterization function, and \mathcal{H} is the space of all such functions.

2. Experimental results

There are five types of results presented in this paper:

1. We segment each long, periodic juggling trial into corresponding cycles.
2. We separate the amplitude and phase variabilities in the segmented cycles using SRVFs.
3. We compute summary statistics such as the mean and covariance of the amplitude component.
4. We perform dimension reduction using PCA in an appropriate space.
5. We define Gaussian models and validate them through random sampling.

2.1. Cycle segmentation

As a first step in our analysis, we want to segment the long, approximately cyclostationary juggling trajectories into their corresponding cycles. To do this, we utilize the tangential velocity functions for all of the trajectories as defined in Ramsay et al. [3]. We use an automatic segmentation algorithm that has been previously introduced by Kurtek et al. [2] and is based on the SRSF representation of functional data. In summary, it works as follows:

1. Manually select a set of cycles from the long signal.
2. Temporally align the chosen cycles and compute their average or median. This forms the cycle template.
3. Slide the cycle template along the long, periodic signal with a pre-specified window, which is application specific.

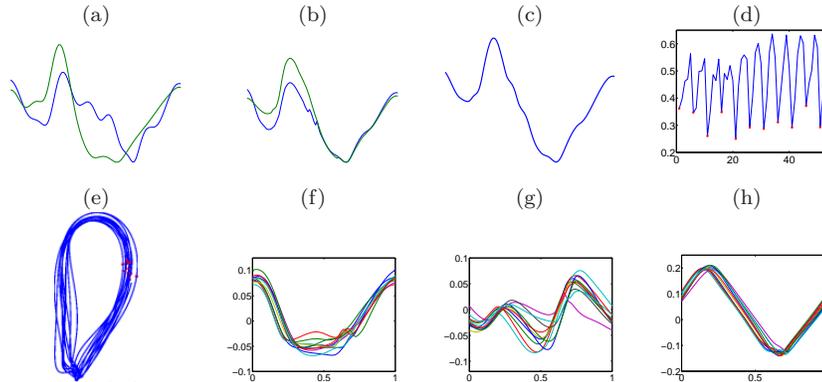


FIG 1. Initial data processing via segmentation of juggling trajectories into corresponding cycles.

4. At each step, optimally match the template cycle to the part of the signal in the current window.
5. Record the phase distance generated based on this alignment. This should result in a harmonic cost function.
6. Find all of the minima of this cost function. These points correspond to the break points between the segmented cycles.

We present the results of this algorithm in Figure 1. In Figure 1(a)–(c), we display two manually selected cycles, their temporal alignment and the resulting cycle template, respectively. The harmonic cost function computed by sliding this template along the long signal is displayed in Figure 1(d). The minima of this function are marked in red and correspond to the segmented cycles. Figure 1(e) shows one juggling trajectory in the xz plane with the beginning of each cycle marked in red. Finally, Figure 1(f)–(h) shows the x , y , and z coordinate functions of the segmented cycles. We note that this algorithm was able to find all of the cycles (in all of the juggling trials) efficiently and accurately.

2.2. Registration and statistical analysis of juggling cycles

After segmentation, if needed, one can temporally register all of the juggling cycles by considering each coordinate function separately. That is, for each coordinate of each cycle we estimate a different warping function using the SRSF representation. An example of coordinate functions and their corresponding SRSFs are presented in Figure 2. The results of this alignment are shown in Figure 3. We are able to achieve nice alignment of features using this approach. However, it is not clear which registration should be used to perform the subsequent statistical analysis. In particular, the juggling cycles are 3D curves and thus the registration should be performed jointly on all three coordinates. Thus, for the remainder of this paper, our approach will be to represent the three-dimensional

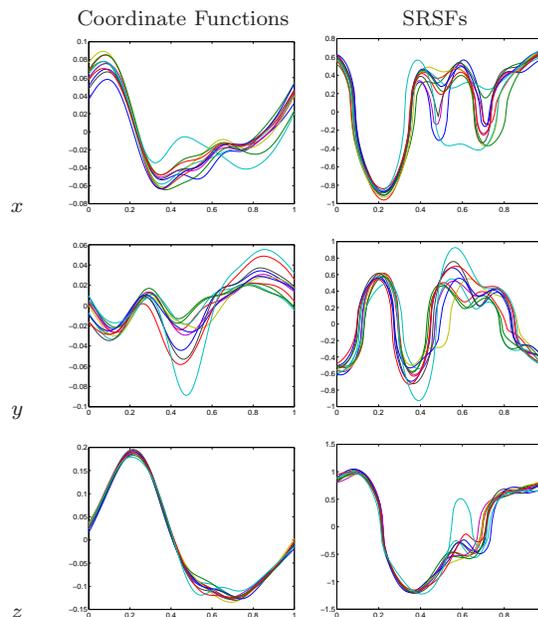


FIG 2. *Coordinate functions and their corresponding SRSFs.*

cycles by their SRVFs, estimate a joint warping function for all coordinates as described in [4], and finally perform statistical analysis of these curves and their corresponding warping functions.

We begin with the statistical analysis of all cycles in juggling trial 10. The cycles in this juggling trial, before and after temporal alignment, are displayed in Figure 4. The phase variability is displayed in colors where dark blue corresponds to the beginning of a cycle and dark red to its end. Note that by temporally aligning these cycles we do not change their 3D coordinates. We only change the correspondence of points along the curves. We see an improvement in correspondence using our alignment method. That is, after alignment, points with similar colors correspond to similar locations along the cycle. We can quantify the improvement in cycle alignment by computing the cumulative variance before and after alignment. The cumulative variance in this trial decreased by 58%, from 9.05 to 3.77, due to optimal alignment of the cycles.

Once the cycles are aligned, we can analyze their amplitude variation, which is closely related to their shape. That is, we can compute the average and covariance and perform PCA to reveal the structure in the aligned data. The first two principal directions of variation are visualized in Figure 5. We display the path generated by this direction within one standard deviation of the mean. In each case, the mean is marked in green. We also display the point-wise magnitude of variation on the mean cycle. We note that the most variation along the first principal direction is seen when the index finger nears the bottom of the cycle. The highest variation in the second direction is in the location of the beginning of

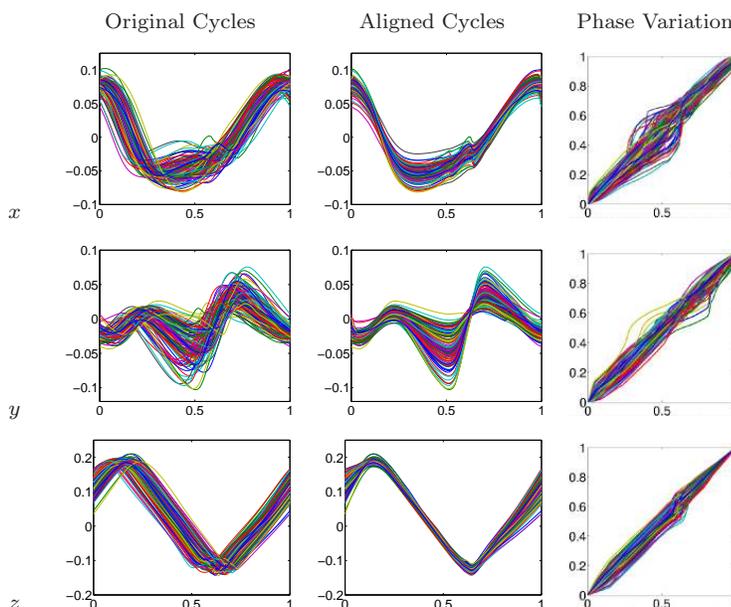


FIG 3. Registration of individual coordinate functions for all juggling cycles.

each cycle along the y direction. Overall, the juggler’s finger follows a fairly stable path along each cycle. Using the average cycle and these principal directions of variation, we impose a Gaussian model on cycle amplitude. Using this model, we can generate random samples, which are displayed in the bottom row of Figure 5. We note that all of the random samples generated based on our stochastic model are valid instances and closely resemble the cycles in the given data.

We can perform similar analysis on the estimated temporal registration functions, which represent the phase variation in the given data. In Figure 6, we display the two principal directions of variation in the registration functions,

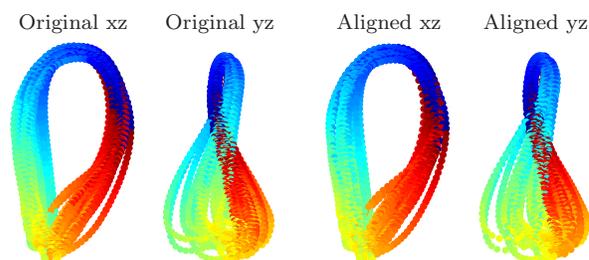


FIG 4. Registration of juggling cycles for one of the trials in the dataset. We provide two different views for improved display. The cumulative variance decreased from 9.05 to 3.77 due to cycle registration.

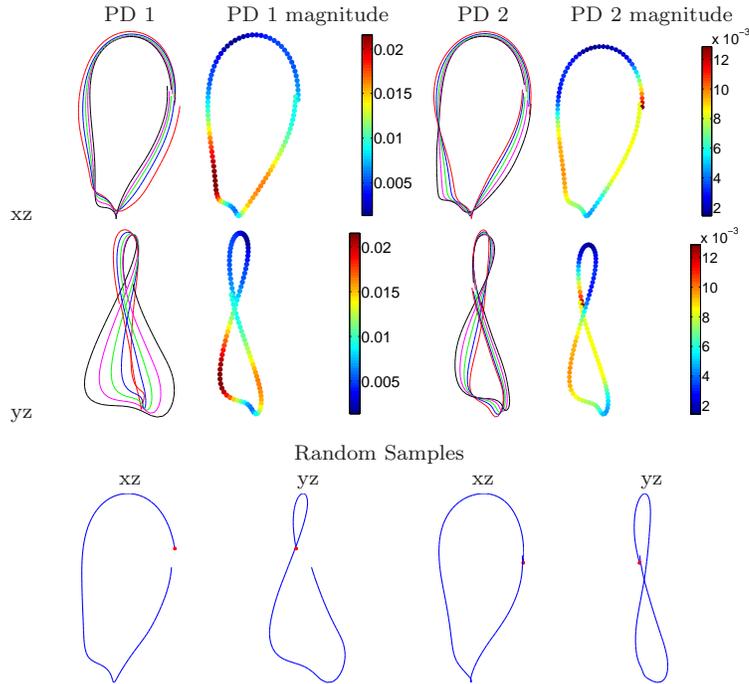


FIG 5. Principal directions of amplitude variation and random samples from a Gaussian model.

and their point-wise magnitudes on the average cycle. We observe that the main direction of variation represents a juggling cycle where the juggler starts either slower or faster compared to the mean juggling speed, catches up close to the midpoint of the cycle, and then again goes out of phase before once again catching up toward the end of the cycle. This type of phase variation is natural to juggling data as the juggler often tries to compensate for performing the action either too slow or too fast, sometimes even multiple times within one juggling cycle. Another interesting observation is that almost all of the phase variation is contained in the first direction. Overall, it is clear that the juggler seems very good at controlling the timing of each cycle as indicated by the estimated

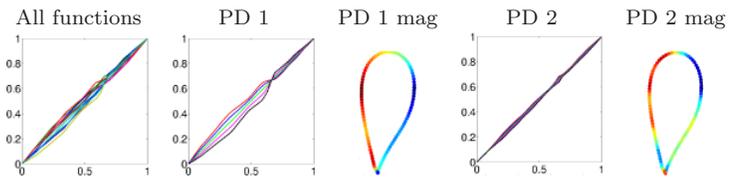


FIG 6. Principal directions of phase variation and their point-wise magnitudes (displayed on the mean cycle).

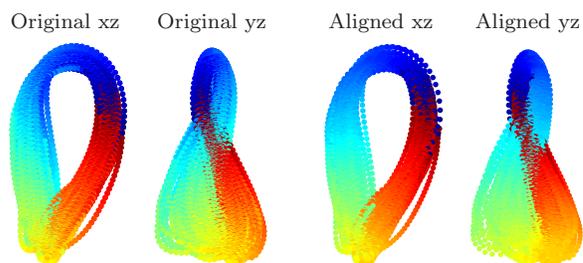


FIG 7. Registration of juggling cycles for all cycles in the dataset. We provide two different views for improved display. The cumulative variance decreased from 11.09 to 4.74 due to cycle registration.

warping functions, which all tend to be close to a function that represents no warping.

We repeated the above described analysis on all cycles from all juggling trials. The results are presented in Figures 7, 8 and 9, where we notice similar results as before. In Figure 7, there is clear improvement in point correspondences across different cycles after alignment. This is especially seen as the phase changes from

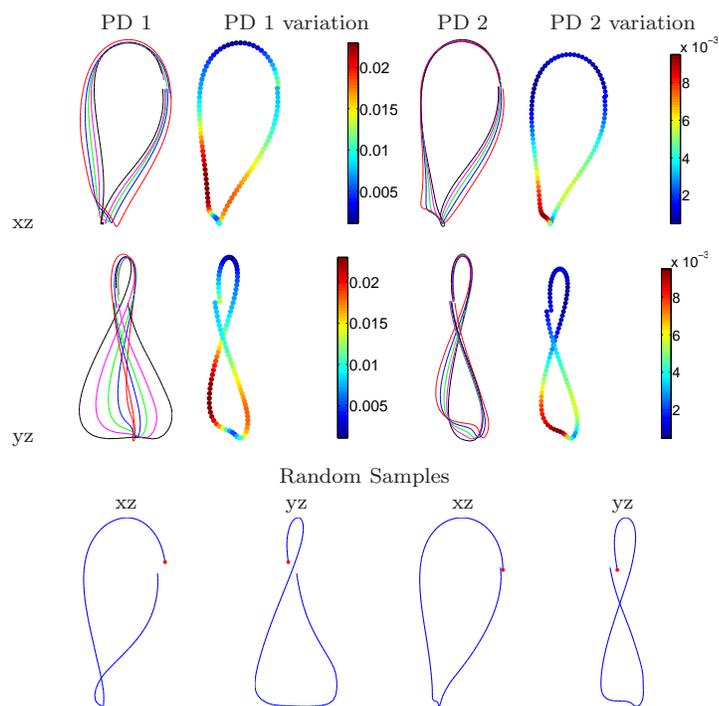


FIG 8. Principal directions of amplitude variation and random samples from a Gaussian model for all cycles.

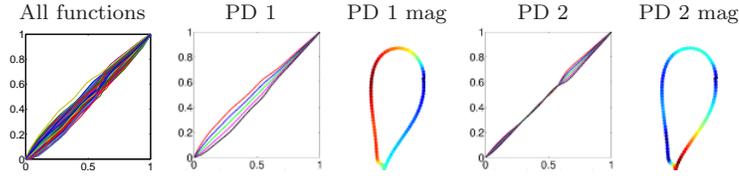


FIG 9. Principal directions of phase variation and their point-wise magnitudes (displayed on the mean cycle) for all cycles.

dark blue to light blue toward the beginning of the cycles, and then again from orange to light red toward the end of the cycles. The cumulative variance was equal to 11.09 before alignment and 4.74 after alignment. Thus, our registration framework provided a 57% reduction in the cumulative variance in this example. The principal direction of amplitude variation (Figure 8) is fairly similar to that displayed in trial 10. On the other hand, the second direction of variation is quite different and roughly shows up-down movement of the index finger in the bottom portion of the cycle. Again, the random samples generated using our model are representative of the original data. In Figure 9, we display the PCA results for the registration functions. The main direction of variation represents cycles that started faster than the mean juggling speed and then slowed down or started slower than the mean and then sped up. This is also a natural type of variability in this data. Once again, we would like to note the stability of the juggler's 3D coordinates of the finger as well as his timing of each cycle in the data. This is especially reflected in the similarity of principal directions of variation in the amplitude and phase components computed using cycles from a single trial (Figures 5 and 6) versus cycles from all trials (Figures 8 and 9).

3. Conclusion

In this paper, we have explored the use of elastic functional data analysis for studying the juggling data. The main tasks that are performed are: (1) extraction of cyclic units from cyclostationary juggling trials, (2) separation of the phase and amplitude components from these units using alignment of 3D curves, and (3) statistical modeling of these components using PCA techniques. As the results indicate, the elastic framework is successful in detecting and modeling the phase-amplitude variability in the given data. This conclusion is further supported by random sampling from the stochastic model with parameters estimated from the data.

Acknowledgements

The authors would like to thank the Mathematical Biosciences Institute at the Ohio State University for support and for hosting this workshop. The authors would also like to acknowledge the grants that in part supported this research: NSF IIS-1217515 and NSF 1208959.

References

- [1] KURTEK, S., SRIVASTAVA, A., KLASSEN, E., DING, Z. (2012), Statistical modeling of curves using shapes and related features, *Journal of the American Statistical Association*, **107:499**, 1152–1165. [MR3010902](#)
- [2] KURTEK, S., WU, W., CHRISTENSEN, G.E., SRIVASTAVA, A. (2013), Segmentation, alignment and statistical analysis of biosignals with application to disease classification, *Journal of Applied Statistics*, **40:7**, 1270–1288.
- [3] RAMSAY, J.O., GRIBBLE, P., KURTEK, S. (2014), Description and processing of functional data arising from juggling trajectories. *Electronic Journal of Statistics*, **8**, 1811–1816, Special Section on Statistics of Time Warpings and Phase Variations.
- [4] SRIVASTAVA, A., KLASSEN, E., JOSHI, S.H., JERMYN, I.H. (2011), Shape analysis of elastic curves in Euclidean spaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33:7**, 1415–1428.
- [5] SRIVASTAVA, A., WU, W., KURTEK, S., KLASSEN, E., MARRON, J.S. (2011), Registration of functional data using Fisher-Rao metric, [arXiv:1103.3817v2](#).
- [6] XIE, Q., KURTEK, S., SRIVASTAVA, A. (2014), Analysis of AneuRisk65 data: Elastic shape registration of curves, *Electronic Journal of Statistics*, **8**, 1920–1929, Special Section on Statistics of Time Warpings and Phase Variations.