Electronic Journal of Statistics
Vol. 8 (2014) 1874–1878
ISSN: 1935-7524
DOI: 10.1214/14-EJS937REJ

# Functional data analysis of juggling trajectories: Rejoinder*

**James O. Ramsay**

Received August 2013.

## 1. Introductory remarks

My appreciation of the innovation, effort, and commitment to the data that has led to these six fine papers, not to mention my own small contribution, goes beyond what I can adequately express in this rejoinder; which is such in the sense of celebrating my return to this outstanding team of researchers and their future research on this fascinating problem of the identification of phase variation in functional data. These collected manuscripts apply five registration methods, three warping families, five collatoral data analyses and two new models; and my commenting on all the new insights and remaining challenges would surely exhaust the patience of the reader.

Perhaps the editor will tolerate my referring to each team by the string of first letters of author names, such as RGK for my own analysis, Ramsay, Gribble and Kurtek (2013).

What do the data and the biomechanical processes that generated them invite the statistician to do? A complete investigation should recognize the three levels of time in the data: (1) across records, (2) within records across juggling cycles, and (3) within cycles. It would be essential, too, to recognize and faithfully represent the information found in at least two orders of derivatives, since the velocity and acceleration vectors at any time indicate the direction of the ball and the force required by the juggling task, respectively. That is, juggling dynamics is easily as important in this problem as juggling statics; and was the motivation for RGK's focus on tangential motion rather than purely positional information. Phase variation, too, demands effort from the juggler, and the warping functions used to represent this effort must also respect the system's limited energy budget available for the task. In this and many application of registration methods, phase variation should be viewed and studied as a substantively interesting part of the total picture, rather than as and ignorable nuisance. Finally, the data analyst has inevitably to tell the story via the technology of statistical graphics, and some strategies work better than others. To expect all this and more from each paper is, of course, unreasonable; but collectively much of this agenda has been achieved.

*Main article 10.1214/14-EJS937.

## 2. Representing phase variation by warping function $h$

Bernardi, Sangalli, Secchi and Vantini (2013) [BSSV] adopted a minimalist approach in which their two-parameter $h$'s were line segments. Because these segments were unconstrained at their boundaries, their approach was unique in not requiring common domains for the registered functions. Of course warping functions this elementary are quite unable to deal with the biphasic nature of the the juggling cycle, but it was surprising how revealing these simple warps were nonetheless, and I was reminded that many applications would only require one- or two-parameter warps, such as would be achieved by expressing the log-derivative $W$ of $h$ as a second or third degree polynomial.

Poss and Wagner (2013) [PW], Tolver, Sorensen, Muller and Mousavi (2013) [TSMM] and RGK opted for the B-spline expansion of $W = \log Dh$ defined by Ramsay and Silverman (2005) and installed in their software. This choice has many advantages, including as much flexibility as the data require, which was considerable for the full record registration of RGK, control over the differentiability of $h$ that can be augmented by regularization, and the possibility of incorporating covariates and input functions in the definition of $h$, in addition to the low dimensional possibilities already suggested.

The estimation of $h$ via the dynamic time warping (DTW) algorithm by Kurtek, Xie and Srivastava (2013) [KXS] and Lu and Marron (2013) [LM] comes in for some negative commentary, I'm afraid. This algorithm, closely related to that for isotonic regression (which also computes monotone functions), was originally defined by Sakoe and Chiba (1978) for registering discrete sequences of phonemes in records of speech. It is an exhaustive search method that assigns a cost value to insertions, deletions or non-changes at the nodes of the lattice of all possible pairs from members of two sequences, and which then assembles a minimum-cost non-decreasing trajectory from the result. The final trajectory increases in discrete jumps to reach the upper right of the lattice from its start at the lower left. Applied to potentially continuous or at least high density sampling points, the resulting warping function is anything but smooth, and we see the consequence of this in the plots of warping functions resulting from the use of the square-root velocity function method as instantiated in the Matlab and R code distributed from Professor Srivastava's lab. These discrete jumps have particularly deadly consequences for derivative estimates, but also discernably distort position functions such as those for the Z coordinates in Figure 3 of KXS. The overly aggressive nature of DTW seems, when applied to the spike train data, to manufacture peaks where none are obvious in the plots of the original data. It is a shame that an otherwise exciting registration method is marred by the use of DTW in a context that was never intended in the first place.

The objective functions used in registration methods need not, in principle, be tied to any particular family $\mathcal{H}$ of warping functions. BSSV optimize a correlational measure motivated by the possibility that curves vary in shape that is dominated by a simple multiplication by a positive constant. I was not sure what the domain of the integration was in their equation (2.1), but suspect that it was the intersection of two domains of the two functions being compared since that would be consistent with their use of linear warps. The minimum eigenvalue criterion used by RGK and TSMM is motivated by the same proportionality argument. PW used an iterated optimization of squared error loss algorithm developed by Kneip and Ramsay (2008) [KR], a criterion already known to distort peaks but nevertheless used by them for reasons of computational convenience.

However, squared error loss applied to the SRVF transformation was used by KXS and LM with far more impressive justification, namely that this is an invariant measure that can be used to develop equivalence classes of registered functions, with geodesic distance measures between these classes or orbits that lead to the Karcher mean concept and to many other good things. This is an exciting approach, and one anchored in solid differential geometry. Moving from DTW to smoother warping family such as that of RS will present computational challenges due to the derivative of the SRVF transformation having bipolar singularities at peak and valley locations, but I'm confident that these difficulties will be effectively dealt with, as has already been the case in applications of $\mathcal{L}_\infty$ loss measures.

Brunel and Park (2013) [BP] offer a completely different but no less valuable attack on the registration problem. Their use of Frenet-Serret frames, arc length, curvature and torsion, as coordinates for three-dimensional space curves bypasses the need for registration entirely, and therefore automatically reflects pure amplitude variation. Their beautiful exposition of their methodology will be appreciated by readers wanting to try this out on their own data. We need to learn how to interpret the resulting representations, as well how to find stably estimate three orders of derivatives; but it seems sure that this approach has much to offer.

On the other hand, the X, Y and Z coordinates used to record the data are related to architecture of the human body, and Z is indispensable because of the critical role played by gravitational force. It might have been worth, however, looking at other coordinates systems, such as polar or elliptical coordinates, in the X-Y plane. In short, multiple choices coordinate systems are entirely permissible, and defaulting to the coordinates used to record the data can seriously handicap an analysis. Indeed, the same may be said for time $t$ as a parameterization of these space curves. It would not be difficult estimate to the integral of the norm of the force vector from knowledge of the mass of the balls, and this might provide valuable insight into how amplitude and phase variation are related to work done over juggling cycles.

There is no particular reason for the usual practice of normalizing the domains of two curves prior to registration, and the impact of this practice on derivatives is a good reason not to do so. For example, the variation in the lengths of the 113 juggling cycles conveys important information about how the ball is thrown, and seems likely to be related to the shape of the juggling trajectories in the Y-Z plane. Domain normalization was even more problematic for the carotid artery shape data. Current software needs to extended so as to permit registrations with domains with different starting and ending points.

BSSV, TSMM and PW, as did KR before them, demonstrate that registration is a more powerful tool when used to represent phase variation around variation defined by a model rather than by a single template function. PW's finding that principal components lined neatly with coordinate directions was dramatic, and their plots of the estimated deformation functions show that the phase variation around these components was impressively small, amounting at most to about 4% of the cycle time. The revealing post-hoc decompositions offered by LM and KXS also point to the value of folding together model fitting and registration. KR admitted that their algorithm does not always perform well, so that more work is required and is already in progress, and we can expect to hear more on this important idea soon.

The analysis by TSMM was a particularly bold attempt to decompose each record into a periodic cycle-length component with constant norm and a longer range variation over the whole record. Perhaps the failure of the constant norm to be stable across records is related to the impact the wilder loops found in some records and not others. In any case, and notwithstanding, the elliptical cycles in their Figures 2 and 4 delighted this reader.

## 3. Other aspects, follow-up analyses and observations

The three-dimensional nature of the trajectories led to some impressive attempts at two-dimensional graphical displays. Coding variation around a mean curve in terms of color in Figure 8 of KXS did not seem quite as informative as the displays of samples of curves in the same figure, and color didn't seem to add much to their Figures 4 and 7 either. But color-coding clusters in Figure 3 of LM and most of the figures in BSSV worked well.

Plotting warping functions $h$ is not that informative since they usually cluster tightly around the diagonal, but plots of deformation functions $d(t) = h(t) - t$ or of the log derivatives $W(t) = \log Dh(t)$ as in Figure 3 of PW makes much more effective use of the space in a graph.

The plot of the SRVF curves in the right panel of Figure 2 of KXS was especially welcome. It shows that this transformation enhances the slope of these curves where they cross the zero line, corresponding the locations of peaks and valleys in the untransformed curves. At the same time, it evens out the variation in the slope of the original curves. This methodology is comes close to being an automatic peak/valley landmark registration procedure, something that most of our nonstatistical clients will find relatively easy to understand.

An especially important follow-up to a registration exercise is the exploration of covariation between phase and amplitude. PW offered the only analysis of this kind, based on principal components scores derived from PCAs of the registered cycles and the log-derivatives of the warps. One is of course not surprised to see phase changes related to variation in how the ball is thrown, such as are indicated by the regression coefficients in their Table 5. Figure 6 of BSSV also shows that the third cycle in each record, when the first-thrown ball is caught, is mainly in cluster 2 containing the tighter loops. The juggler presumably has more control of the trajectory of the first ball, and this should also affect the shape of the third warp.

The most involved second stage analyses of amplitude variation are provided by the LM team, using new methods developed for object-oriented data and surveyed in Marron and Alonso (2013), as well as PCA. It's exciting see the greater clarity offered by these approaches showcased in this context.