Functional Components of Variation in Handwriting

J. O. Ramsay, McGill University

October 17, 1997

1 Introduction

Handwriting displays substantial and complex variation, both across and within individuals. At a casual level of observation, it is not hard to see why this should be so. The hand is a complex bio-mechanical system, and the movements made by the fingers, wrist, and forearm are produced the contraction of dozens of muscle groups, applied to the skeletal structures and the soft tissue that envelops them. The system exhibits various kinds of harmonic or oscillatory behavior, is subject to forces of elastic and viscous components of muscle, and must work in the context of gravitational and inertial forces. These mechanical components vary in many ways from one individual to another, as well as within an individual.

Moreover, handwriting is the consequence of a complex control process that activates muscle contraction. Trains of spike potentials cascade down a network of nerve fibers and arrive at motoneuron junctions to trigger or release contraction of muscle fibres. These neural events exhibit variability in both timing and amplitude, due in part to the variation in the activity at the cortical level, and in part to variation in transmission properties along the neural pathways.

This paper attempts to display at least some of this complex variation in terms of components of a linear differential equation that is estimated from replicated handwriting samples from a single individual. A number of steps are required as preliminaries, and these are typical of a functional data analysis, as described in Ramsay and Silverman (1997). The goal of the paper is perhaps more to indicate how a model of this type can be constructed and why it might be useful than to make a substantial contribution to the literature on handwriting itself.

1.1 Why a Linear Differential Equation?

Exploratory techniques such as principal components analysis will certainly offer us some insight into data such as these. But there are arguments in favor of a differential equation model. First, because handwriting is the result of a mechanical process, it necessarily possesses derivatives, and moreover we can expect from Newton's Third Law that looking at acceleration will be a way of seeing the forces involved. A differential equation describes processes by finding relationships between derivatives. For example, the position $x(t)$ of an undamped spring is determined by $D^2x(t) = -\omega^2 x(t)$, $\omega/(2\pi)$ being the frequency of oscillation. This is a second-order linear homogeneous differential equation, and more complex mechanical systems subject to frictional and other forces can be described in terms of second order linear equations of a more general nature.

This system is subject to external control through forces applied by muscle contractions, and it seems reasonable to suppose that we can see this control system at work more clearly through aspects of a differential equation model than through models taking into account only position.

A differential equation also exploits the intrinsic smoothness of the process by taking explicit account of derivative behavior. This is not to say, of course, that we do not need to allow for observational noise that may perturb the measured pen position in a nonsmooth manner.

Finally, it will be indicated that a differential equation model permits a rather richer error or stochastic structure to account for observed variation in handwriting samples than is available through traditional smooth-signal-plus-error models.

1.2 An Outline of the Steps Involved

After a preliminary inspection of the data, the first task is to estimate derivatives. Of the various methods available, this paper will make use of a spline smoothing module that permits smooth derivative estimates for an arbitrary derivative $D^m x$.

An essential next step is registration of the data. This involves aligning coordinate functions by allowing for smooth monotone but nonlinear transformations of clock times so as to allow for temporal variation across handwriting samples. A registration technique of Ramsay and Li (1997) based on a smooth monotone function family considered in Ramsay (1997) is used here.

Once the curves are registered, so that temporal variation is largely removed, the next step is to model amplitude variation from one sample to another by a linear differential equation. For these data, a homogeneous equation of the third order in velocity is estimated, using a technique called principal differential analysis (PDA) described in Ramsay (1996).

The final stage is the interpretation and evaluation of this equation and how well solutions of the equation can recover the original individual handwriting samples.

2 The Data and Preliminary Analyses

2.1 The Experiment

The handwriting sample consists of 50 replications by one person writing the Chinese characters for "statistical science", displayed in Figure 1. Each replicate consists of three functions, one for each of X-, Y-, and Z-motion, corresponding to lateral and vertical movement in the writing surface, and lifting of the pen off the surface, respectively.

Each sample pen position was recorded by an OPTOTRACK tracking system that recorded the position in 3-space of a small infra-red emitting diode fastened near the writing tip of the pen. The position sampling rate was 400 per second. Samples varied in duration, and a preliminary normalization and linear interpolation were performed so that each record consisted of 601 values of the three coordinate functions, corresponding to six seconds, this being roughly the average duration of the samples. Figure 2 shows the three coordinate functions.

The notation $x_i(t)$ will be used in the following to denote any one of the coordinate functions for record i. When a specific coordinate is specified, $x_{ik}(t)$ will be used to indicate the kth coordinate function, $k = 1$ for X, $k = 2$ for Y, and $k = 3$ for Z.

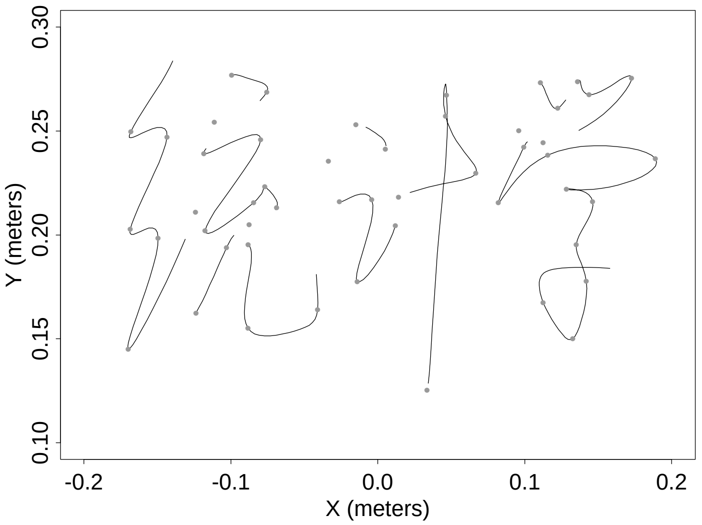

Figure 1: The average script, after registration, for "statistical science" in Chinese. Dots indicate the positions of peaks in the acceleration vector.

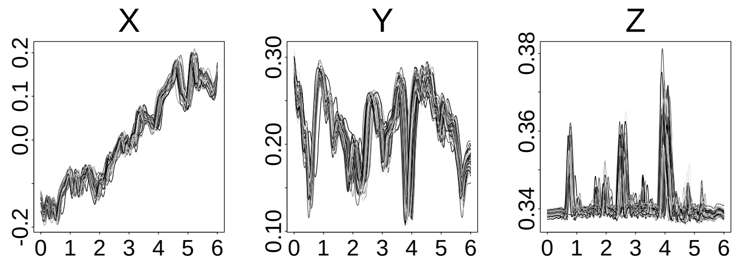

Figure 2: The 50 coordinate functions after initial normalization and interpolation to 601 equally-spaced values. The time scale is in seconds, and the position scale is in meters.

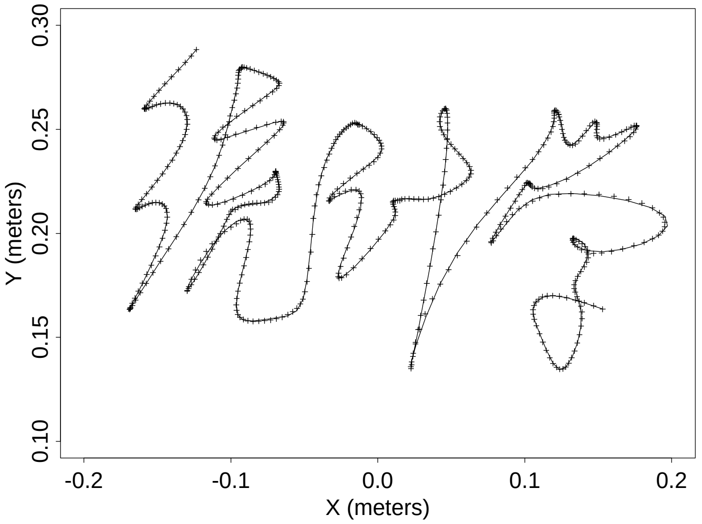

Figure 3: The points indicate the observed values for the first record, and the line shows the estimate of the smooth curve achieved by using least squares approximation by order 9 b-spline basis functions defined by 101 equally-spaced knots.

2.2 Smoothing and Derivative Estimation

The OPTOTRACK system is rather accurate, so that the signal-to-noise ratio is high. The standard error of pen position was estimated to be about 0.4 mm. Nevertheless, the signal is very complex, and, moreover, estimates of derivatives up to order three were required. Consequently, the high sampling rate essential, and a certain amount of smoothing of the raw data was also required.

After some experimentation, it was felt that the records were adequately represented by approximating each set of observed coordinate values in terms of a set of order 9 B-spline functions defined by 101 equally spaced knots, $0, 0.06, \ldots, 5.94, 6.00$. These specifications define 108 b-spline functions, and the approximation is a piece-wise polynomial of degree 8. Fitting was by least squares. The quality of the approximation of the data for the first record is displayed in Figure 3.

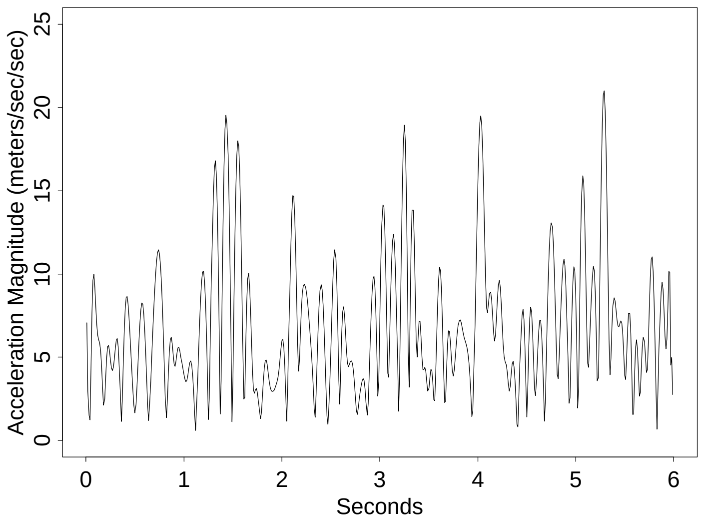

Figure 4: The magnitude of the acceleration vector for the first record.

2.3 Looking at Acceleration

As indicated above, we would expect that acceleration might reveal some interesting aspects of the control processes driving this system. Figure 4 displays the magnitude of the acceleration vector, $(D^2X + D^2Y + D^2Z)^{1/2}$, for the first record. This plot, typical of all records, indicates some rather remarkable characteristics of handwriting.

About 50 acceleration pulses are clearly visible, corresponding to about 8 pulses per second. These are separated by near-zero acceleration events, indicating that the many muscles involved are acting in a remarkably co-ordinated fashion, performing with impressive consistency and unity. By comparison, an expert typist can produce about 500 key strokes per minute, or around 8 per second, each involving an up- and a down-stroke. This rate is also comparable to the number of phonemes per second produced in normal speech. Thus, this rate of event production ranks with the best that we can achieve.

The magnitude of the acceleration pulses are likewise worth noting; if sustained, an acceleration of 20 m/s^2 would put a satellite in orbit in about seven minutes, or move a car from 0 to 100 km/h in about one second. It seems at least as dramatic that these pulses are separated by instants of near zero force, or linear motion, achievable only if all the muscles involved simultaneously relax.

3 Registration of the Samples

We see in these data that, while individual records retain the level of detail in Figure 4, the timing of the events involved varies substantially from record to record. If we were, for example, to attempt to estimate the average acceleration function by simple cross-sectional averaging, the time variation would cause the sharp acceleration peaks troughs to largely disappear, and thus be unrepresentative of individual curves. This temporal variation is to be expected; higher velocities are exhibited for some characters and records than for others, due perhaps to factors such as speed, arousal, arm position, motivation level, and etc.

Consequently, the registration process seeks for record i a smooth strictly monotone function h_i such that, for example, x_1 at time $h_1(t)$ is exhibiting about the same behavior as x_2 at time $h_2(t)$, so that the values $x_i[h_i(t)]$ can be viewed as comparable. Function h_i is called in the engineering literature a warping function, and can be viewed as a transformation from clock time to system time. Naturally, all three coordinate functions for a given record are altered by a single common warping function.

The warping functions used in this analysis were developed by Ramsay (1997), and have the form

$$h_i(t) = C_i \int_0^t \exp[\int_0^u w_i(v)\, dv]\, du. \tag{1}$$

The function w_i is completely unconstrained, except for being integrable, and therefore easier to estimated than h_i itself. The constant C_i is used to ensure that $h_i(6) = 6$. The identity transformation corresponds to $w_i = 0$, and a constant value for w_i implies an exponential transformation. In fact, it can be shown that all twice-differentiable strictly monotone functions can be put in this form.

The registration algorithm used involves the following features:

1. Each w_i is expressed as a piece-wise constant function defined by a set of 20 equally-spaced knots.

2. Each sample is registered towards a single target function x_0 by minimizing the smallest eigenvalue of

$$\sum_{k=1}^{3} \begin{bmatrix} \int x_{0k}^2(t)\,dt & \int x_{0k}(t)x_{ik}(t)\,dt \\ \int x_{0k}(t)x_{ik}(t)\,dt & \int x_{ik}^2(t)\,dt \end{bmatrix} \tag{2}$$

with respect to the values $c_{ik\ell}$ specifying the height of w_{ik} over interval ℓ. If the smallest eigenvalue is near zero, then the warped coordinate and the target have a nearly linear relationship.

3. A penalty term was also appended to this criterion to control the size of function w_i, thus ensuring that the monotone functions were kept smooth.

4. Initially the simple cross-sectional mean is used for x_0. But after a first registration cycle, this was replaced by the cross-sectional mean of the warped functions, and a second registration cycle is performed.

This process, called Procrustes fitting, can be repeated until convergence in the cross-sectional mean is achieved. This usually requires only two cycles in practice. Further technical details can be found in Ramsay and Li (1997).

Figure 1 was constructed from the cross-sectional mean resulting from two cycles. Figure 5 displays all 50 acceleration magnitude functions. We see a remarkable stability in both timing and amplitude for the first two characters after registration, but rather less stability for the final character. This reflects real differences in the way the third character is formed rather than any failure of the registration process. One is nevertheless left with an impression of a process showing great complexity and stability at the same time. Figure 6 shows the 50 third derivative estimates, and even at this depth of differentiation we see much stability.

4 The Differential Equation Model

4.1 The Model

The differential equation model proposed for the data is as follows. Let L be the linear differential operator defined by

$$Lx(t) = w_1(t)Dx(t) + w_2(t)D^2x(t) + D^3x(t). \tag{3}$$

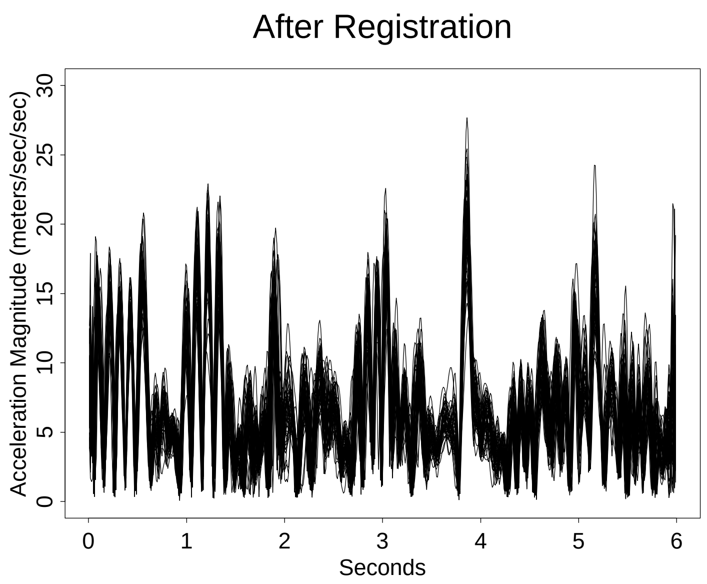

Figure 5: The 50 acceleration magnitude functions after registration.

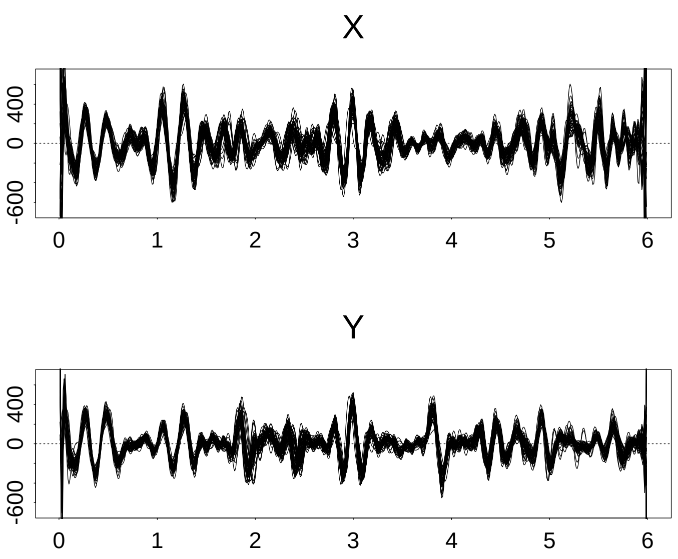

Figure 6: The 50 third derivative estimates after registration.

The model is

$$Lx_i = f_i. \tag{4}$$

The operator L is defined by the two coefficient functions w_1 and w_2, weighting the first two derivatives in a manner that varies over time t and is relative to the size of D^3x. The function f_i is called the forcing function, and corresponds to the residual term in regression analysis. It reflects variation that cannot be explained by the homogeneous differential equation $Lx_i = 0$, and its size should be viewed in relation to that of D^3x. A separate operator L is required for each coordinate. Equation (4) can be viewed as a second order equation in velocity Dx, and thus is invariant with respect to translations in any of the coordinates.

In order to simplify the analysis somewhat, the linear trend in the X-coordinate was first removed by regressing these functions on time and replacing them by their residuals. In effect, this places the origin of the system within the hand-arm structure rather than in a fixed point, and this could be viewed as a reasonable change of coordinates. This transformation also centers the velocity functions about 0.

4.2 Estimating the Equation

Each operator L_k is identified by using the data to estimate the two coordinate functions w_{1k} and w_{2k}. An approach called principal differential analysis, or PDA, developed by Ramsay (1996) proceeds by expanding each coefficient function in terms of a fixed number of basis functions. The fitting criterion to be minimized is

$$\sum_{i=1}^{50} \int_0^6 f_i^2(t)\, dt. \tag{5}$$

That is, the fit aims to minimize in the sense specified by this criterion the discrepancy between the actual curves and the model $Lx_i = 0$. The fitting criterion can be seen to be quadratic in the coefficients defining the expansions of the basis functions, and the PDA process is simply one of optimizing (5) with respect to these coefficients, a procedure that is noniterative and reasonably rapid. For these data, where a high level of detail is required in the coefficient functions, the expansion each coefficient function was in terms of the same 108 order 9 B-spline functions used to approximate the raw observations.

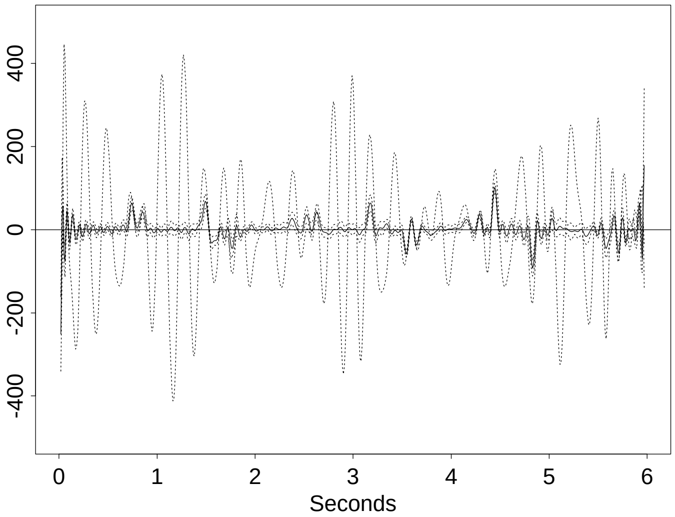

Figure 7: The solid line is the average forcing function f for the X-coordinate. Dotted lines immediately above and below it are pointwise 95% confidence limits. The larger dotted line indicates the average third derivative for comparison purposes.

4.3 The Results

If the homogeneous equation Lx_i is successful, the size of the f_i's should be small relative to that of D^3x, and should exhibit white noise characteristics. Figure 7 displays the mean forcing function for the X-coordinate, along with its pointwise 95% confidence limits. Also displayed for comparison purposes is the average of D^3x. We see that, on the average, the forcing function is small relative to the size of the third derivative over most of the time interval. However, we also see that there are several sharply localized events where the average forcing function deviates strongly from zero. Results for Y and Z were similar.

Figure 8 shows the coefficient functions for the X-coordinate. As is the case with regression and principal components analysis, direct interpretation of these functions may not be easy. However, we can make a crude comparison with what we would see were the system exhibiting purely harmonic motion and thus behaving like a spring or pendulum. In this case, we would see $w_2 = 0$, and w_1 take on a positive constant value. In fact, w_2, although

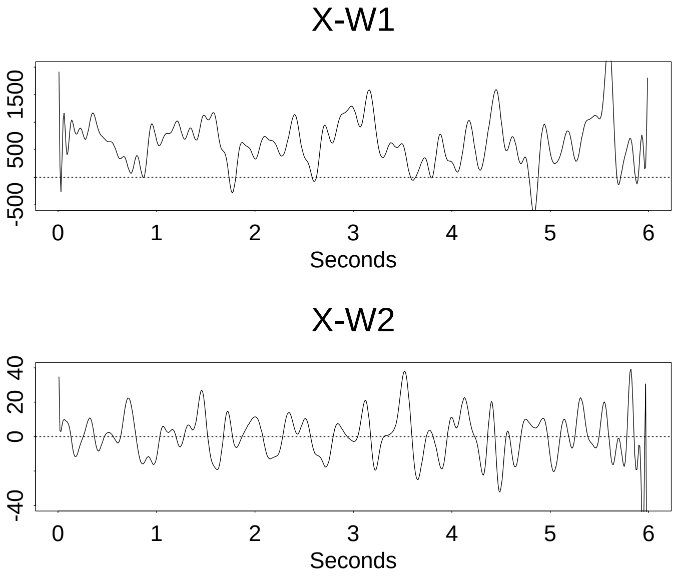

Figure 8: The estimated weight coefficient functions w_1 and w_2 for the X-coordinate.

obviously variable in important ways, is not greatly different from 0, and w_1 has values varying around about 700. This corresponds to a frequency of $\sqrt{700/(2\pi)} \approx 10cps$, or around 20 strokes per second. In fact, this is about what we see in the script. Again, results for the other two coordinates are rather similar.

The solution space to the nonhomogeneous equation $Lx = f_i$ is spanned by the three linearly independent functions u_q satisfying the homogeneous equation $Lx = 0$, among which is $u_1 = 1$; and by the function $\int G(t,w)\bar{f}(w)dw$, where in this case $\bar{f}$ is the mean forcing function such as shown in Figure 7, and G is the Green's function associated with L. Standard differential equation-solving techniques, such as the classic Picard's method, can be used to estimate these solution functions, and these can then be used as a basis for re-constituting the original functions. Figure 9 displays both the original X-Y coordinates and the re-constituted approximation for the first record. We see that the approximation is fairly satisfactory.

Since it is the task of the estimated differential operator to minimize the size of the forcing function, it is primarily these f_i's that should be taken as evidence that the model has worked or not. That the recovered position functions look good is encouraging, but if this were the main aim,

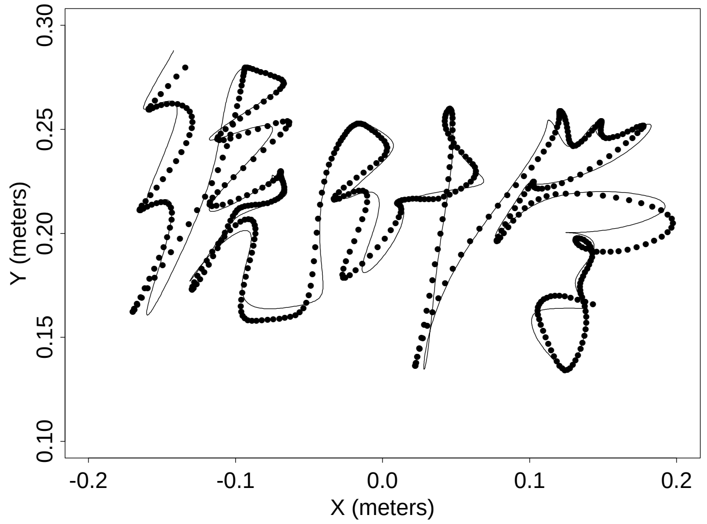

Figure 9: The observed script for the first record (X-Y plot only) is given by the points, and the approximation to this script based on solving the homogeneous differential equation is the solid line.

we would expect a functional version of principal components analysis, such as is described in Ramsay and Silverman (1997), to do a better job.

5 Discussion and Conclusions

The differential equation model has three functional degrees of freedom, corresponding to the two weight functions and the forcing function. These seem to do a reasonable job of accounting for these 50 actual curves, so that a fair amount of data compression has been achieved.

A number of modifications of this model seem plausible, but perhaps beyond the capacity of this small sample to support. The model used here has treated each coordinate separately, but it seems possible that we should aim at coupled systems of equations, so that variation in X can be modeled as connected with variation in Y, for example. This is possible to do in principle with PDA, but the existing model seems to be about as complex as required for such a modest sample.

One may also ask whether the correct coordinate system was used. Is the

Euclidean system the right one? Or would some aspects of a polar coordinate system show more simplicity? However, we can at least say that the order of the operator L annihilating the curves has been identified, and this would remain invariant with respect to any one-to-one change of coordinates.

Finally, what it would see that we have learned is the large amount of useful information available in derivatives. A linear differential equation system is only one possible approach to exploiting this information, but what seems to be best illustrated by these data can be seen directly in Figure 4 without any recourse to a model: The action in functional data may be more obvious at the level of derivatives than at the level of the curves themselves.

References

Ramsay, J.O. (1996) Principal differential analysis: Data reduction by differential operators. *Journal of the Royal Statistical Society, Series B,* **58,** 495-508.

Ramsay, J. O. (1997) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B*, to appear.

Ramsay, J. O. (1997) Curve registration. *Journal of the Royal Statistical Society, Series B*, to appear.

Ramsay, J. O. and Silverman, B. W. (1997) *Functional Data analysis.* New York: Springer.