# 11 The Dynamics of Handwriting Printed Characters

### 11.1 Recording handwriting in real time

The way we handwrite characters is a deeply individual matter, as bank tellers who ask for your signature and graphologists who claim to be able to study your personality from your handwriting know well. The handwriting samples that they work with are static, in the sense that they consider the trace left behind well after the signature is formed, and thus are at one remove from the person who actually did the original writing. In this sense, any attempts to identify an individual, let alone to claim to reconstruct aspects of their personalities, have the flavor of archaeological digs.

What if we could use the online time course of the formation of a signature? Would we not see things as the signature unfolds in time that could not be observed in the static image? Could we see, for example, when a person was nervous, in a hurry, suffering from the onset of Parkinsonism, or rejoicing in a state of profound tranquillity and peace? Surely we could discover new ways by which a handwriting sample characterizes a specific individual, and perhaps use this to make forgery harder than it is now.

In this chapter we use what we call a dynamic model for handwriting. We demonstrate how the model can be fitted to the writing of a particular individual using repeated samples of their printing. We also investigate how well the model separates one person from another.

Our first task, however, is a brief and nontechnical account of some simple dynamic models. Those familiar with differential equations may well be happy to skip ahead, but many readers will find this next section important.

#### 11.2 An introduction to dynamic models

The term *dynamic* implies change. When we speak of the dynamics of a function of time, we are discussing some aspect of the change in curve values over small changes in time, and we therefore focus on one or more of the derivatives of the curve. Chapter 6 described the dynamics of growth, and indeed we defined growth there as the rate of change of height.

A dynamic model therefore involves one or more derivatives with respect to time. Because a number of orders of derivatives may be involved, we use the notation  $D^m x(t)$  to denote the *m*th derivative of the function x(t). This is more convenient than using a separate symbol for each derivative, as we did in Chapter 6, or the classic notation

$$\frac{d^m x}{dt^m},$$

which is too typographically bulky to perpetuate here.

The most common form of dynamic model is an equation linking two or more orders of derivatives. In our present notation, the fundamental equation of growth that we developed in Chapter 6 is

$$D^2 x(t) = \beta(t) D x(t), \qquad (11.1)$$

and this equation links the first derivative to the second by the functional factor  $\beta(t)$ . It is an example of a *linear* differential equation and has the structure of a standard regression model, albeit one expressed in functional terms:

- the acceleration  $D^2x(t)$  is the dependent variable,
- the velocity Dx(t) is the independent variable,
- $\beta(t)$  is the regression coefficient, and
- the residual or error, not shown in the model (11.1), is zero.

The regression model is functional in that the variables and the coefficient  $\beta(t)$  all depend on t. But if we fix time t, and we have in hand N replications  $x_i(t)$  of the curve, you can well imagine that ordinary regression analysis would be one practical way to estimate the value of  $\beta(t)$  at a fixed time  $t_j$ . As the dependent variable in a standard regression, you would use the N values  $y_i = D^2 x_i(t_j)$  for  $i = 1, \ldots, N$ . The corresponding independent variable values would be  $z_i = Dx_i(t_j)$ , and so you would estimate the constant  $\beta(t_j)$  as the coefficient  $b = \sum y_i z_i / \sum z_i^2$  resulting from regressing y on z without an intercept. And you would be quite right!

We briefly review how a differential equation determines the behavior of a function, by considering a second-order linear differential equation, restricted to having constant coefficients:

$$D^{2}x(t) = \beta_{0}x(t) + \beta_{1}Dx(t).$$
(11.2)

Process	Equation	Coefficients	
		$\beta_0$	$\beta_1$
Linear motion	x(t) = a + bt	0	0
Exponential growth/decay	$x(t) = a + be^{\gamma t}$	0	$\gamma$
Harmonic motion	$x(t) = a\sin\omega t + b\cos\omega t$	$-\omega^2$	0
Damped harmonic motion	$x(t) = e^{\gamma t} (a \sin \omega t + b \cos \omega t)$	$-\omega^2$	$\gamma$

Table 11.1. Some processes defined by a second-order linear differential equation with constant coefficients

Table 11.1 relates some special cases of the equation to some familiar functional models and physical processes.

The constants a and b in the table are arbitrary. We see that equation (11.2) covers three basic dynamic processes of science. If both coefficients are zero, we have the linear motion exhibited by bodies that are free of any external force. But if  $-\beta_0$  is a positive number, we see the other type of stationary motion, that of perpetual oscillation.<sup>1</sup> Introducing a nonzero value for  $\beta_1$ , however, results in exponential growth or decay, without oscillation if  $\beta_0 = 0$ , and superimposed on harmonic motion otherwise.

Note, too, the models in Table 11.1 also define the simultaneous behavior of a certain number of derivatives. In fact, the characteristics of both the first and second derivatives are essentially specialized versions of the behaviors of the functions themselves. In this sense, then, these models are really about the dynamics of the processes.

In (11.2) we considered the special case of constant coefficients. What difference does it make if the coefficients  $\beta_0(t)$  and  $\beta_1(t)$  also change with time? If the change is not rapid, we can consider the corresponding differential equation as describing a system that has an evolving dynamics, in the sense that its frequency of oscillation and its rate of exponential growth or decay are themselves changing through time. The larger the value of  $-\beta_0(t)$  the more rapidly the system will oscillate near time t.

In this chapter and subsequently, we use linear differential equation models of order m in the general form

$$D^{m}x(t) = \alpha(t) + \sum_{j=0}^{m-1} \beta_j(t) D^j x(t).$$
(11.3)

There are *m* coefficient functions  $\beta_j(t)$  that define the equations, but in specific applications we may want to fix the values of some of these. In particular we may set one or more to zero.

<sup>&</sup>lt;sup>1</sup>Coefficient  $-\beta_0$  can also be negative, of course, and in this case the sines and cosines in the last two rows of Table 11.1 must be replaced by their hyperbolic counterparts. But the positive case is seen much more often in applications.



Figure 11.1. Twenty registered printings of the characters "fda."

In addition to the coefficients  $\beta_j(t)$ , the form (11.3) contains the function  $\alpha(t)$ , called the *forcing function* in many fields that use differential equations. The function  $\alpha(t)$  often reflects external or exogenous influences on the system not captured by the main part of the equation, or that part of the derivative  $D^m x(t)$  not captured by the simultaneous variation in the lower-order derivatives. From a regression analysis perspective, we may regard  $\alpha(t)$  as the constant or intercept term. If  $\alpha(t) = 0$ , the differential equation is said to be *homogeneous*, and otherwise is *nonhomogeneous*.

#### 11.3 One subject's printing data

The data are the X-, Y- and Z-coordinates of the tip of the pen captured 200 times a second while one subject, designated "JR," prints the characters "fda" N = 20 times. The X-coordinate is the left-to-right position on the writing surface. Coordinate Y is the up-and-down position on the writing surface, and Z is the position upward from the writing surface. Of course, static records give very little information about Z at all—we can only see the X and Y values corresponding to times when Z is zero, and at times when Z is nonzero we have no data at all. The additional richness of a dynamic record is considerable.

Extensive preprocessing is required before we are ready to fit a differential equation. The times of the beginning of "f" and the end of "a" for each



Figure 11.2. The average of 20 registered printings of "fda." The average position of the pen was on or very near the writing surface for the solid lines, and lifted off for the dotted lines.

record must be identified, and the coordinate system in which measurements were taken must be rotated and translated to the (X, Y, Z) system that we described above. In addition we register, or time-warp, the records  $[x_i(t), y_i(t), z_i(t)], i = 1, ..., N$  to a common template  $[x_0(t), y_0(t), z_0(t)]$ . The details of the registration step are described in Chapter 7, and we assume that we can take off from where we left the handwriting problem there. Figure 11.1 shows the trace in the X-Y plane of the 20 functional records, after registration.

Figure 11.2 displays the mean characters for this subject. Most of the registration process does not affect the individual static records plotted in Figure 11.1, but, as we saw in Chapter 6, registration is crucial in the estimation of the mean. The regions where the average position of the pen is clearly above the writing surface are shown in Figure 11.2 as dotted lines, and we see that there are four such intervals. The characters are formed from five strokes on the writing surface (two for "f," two for "d," and one for "a") along with the four off the surface. The average time taken to print these characters was 2.345 seconds, and corresponds to an average of 0.26 seconds per stroke. Note the two sudden changes of direction or cusps between the main part of the "f" and its cross-stroke, and at the beginning and end of the downstroke for "d." There is a lot of energy in such sudden events, and they may be hard for a dynamic model to capture.

#### 11.4 A differential equation for handwriting

We now want to estimate a linear differential equation for each of the three coordinate functions. We use a third-order equation, m = 3. The third derivative is sometimes called "jerk."

To make our task a bit easier, we simplify our equation by fixing  $\beta_0(t) = 0$ . Without this constraint, the equation would have to be recalibrated for any translation of the coordinate values. The resulting equation is, in the case of the X-coordinate for record i,

$$D^{3}x_{i}(t) = \alpha_{x}(t) + \beta_{x1}(t)Dx_{i}(t) + \beta_{x2}(t)D^{2}x_{i}(t) + \epsilon_{xi}(t).$$
(11.4)

There are two coefficient functions  $\beta_{x1}(t)$  and  $\beta_{x2}(t)$ , as well as the forcing or intercept function  $\alpha_x(t)$ . In effect, this is a second-order nonhomogeneous linear differential equation in pen velocity, so we can think of velocity as our basic observed variable. The residual function  $\epsilon_{xi}(t)$  varies from replicate to replicate, and represents variation in the third derivative in each curve that is not accounted for by the model. There are, of course, corresponding coefficient, forcing, and residual functions associated with coordinates Yand Z. In particular, the forcing function for coordinate Z is the aspect that allows the pen to lift off the paper, because when the pen is in contact with the paper  $z_i$  and all its derivatives are zero.

We carry out one additional preprocessing step, by removing the linear trend in the X-coordinate as the hand moves from left to right. In effect, this positions the origin for X in a moving coordinate frame that can be thought of as at the center of the wrist. If the slope of the linear trend is v, the adjusted X-coordinate will satisfy the same model as the original, with a multiple of  $v\beta_{x1}$  added to the forcing function. So this linear correction will only have an important effect on the model if there is substantial variability in the rate of moving from left to right, which in practice there is not.

How do we estimate an equation such as (11.4)? Our first task is to find a good nonparametric estimate of the derivative functions using the 20 replications. These function estimates are then used to estimate the two coefficient functions  $\beta_{x1}(t)$  and  $\beta_{x2}(t)$  and the forcing function  $\alpha_x(t)$ . Returning to the regression perspective, a successful equation will mean that the residual function  $\epsilon_{xi}(t)$  is relatively small for all records and all t. The natural approach will be ordinary least squares in the sense that we choose to minimize, in the X-coordinate case,

$$SSE_X = \sum_{i=1}^N \int_0^T \epsilon_{xi}^2(t) \, dt.$$

The adequacy of the fit can be assessed by comparing the residuals to the third derivative, which acts as the dependent variable in the regression analysis. The technique for minimizing  $SSE_X$  with respect to  $\alpha_x(t)$ ,  $\beta_{x1}(t)$ , and



Figure 11.3. The top panel shows the function  $-\beta_{x1}(t)$  for the differential equation describing the motion of the pen in the horizontal or X direction. The dashed-dotted line indicates the average value, and corresponds to a horizontal oscillation every 0.58 seconds. The bottom panel shows the corresponding function  $\beta_{x2}(t)$ , and this tends to oscillate about zero. It controls the instantaneous exponential growth or decay in the instantaneous oscillation. The shaded areas correspond to periods when the pen is lifted off the paper.

 $\beta_{x2}(t)$  was developed by Ramsay (1996b), who called the method *principal* differential analysis because of its close conceptual relationship to principal component analysis.

Figure 11.3 displays the two estimated coefficient functions for the Xcoordinate. Although it is hard to see much to interpret in these functions, one can compare them to the equation for harmonic motion in Table 11.1. We notice immediately that there is considerable variability in both functions about the average value, also displayed in the plot. This variability is due to the control of the hand by the contracting and relaxing muscles, and these in turn are controlled by neural activation arriving from the motor cortex of the brain. The rapid local variations in the plots are easily ignored "by eye," but perhaps suggest that a regularization term could be



Figure 11.4. The forcing function  $\alpha_x(t)$  for the differential equation describing the motion of the pen in the horizontal or X direction. The dashed-dotted line indicates the average value of the third derivative, to give an idea of the relative size of  $\alpha_x(t)$ . Episodes of forcing occur when  $\alpha_x(t)$  deviates strongly from zero. The shaded areas correspond to periods when the pen is lifted off the paper.

added to the criterion  $SSE_X$ . The formal inclusion of regularization is an interesting topic for future investigation.

Function  $-\beta_{x1}(t)$  plays the role of  $\omega^2$  in the harmonic equation and, since the period of oscillation in a harmonic system is  $2\pi/\omega$ , the larger the value of  $\beta_{x1}(t)$  at some time point t, the faster the velocity is oscillating at that time. The average value of  $-\beta_{x1}(t)$  is 259, corresponding to an oscillation every  $2\pi/\sqrt{259} = 0.39$  seconds. This means that the hand is producing a horizontal stroke once each 0.20 seconds, on the average, which agrees closely with what we observed in Figure 11.2.

On the other hand, coefficient function  $\beta_2(t)$  varies about a value relatively close to zero. It determines the instantaneous exponential growth  $(\beta_{x2}(t) > 0)$  or decay  $(\beta_{x2}(t) < 0)$  in the oscillations.

Corresponding analyses were performed for the other two coordinates. The dynamics of the Y-coordinate resemble those of the X-coordinate, in that the average value of  $-\beta_{y2}(t)$  is 277, and this also corresponds to a period of about 0.38 seconds. The Z-coordinate, however, has an average period of 0.29 seconds.

Figure 11.4 shows the estimated forcing function  $\alpha_x(t)$  for the Xcoordinate. We focus our attention on times when  $\alpha_x(t)$  deviates strongly from zero, indicating times when the homogeneous version of the equation will not capture the intensity of the dynamics. The first peak is in the curved part of the "f" downstroke, when the pen is changing direction and probably accelerating. The next substantial deviation coincides with the pen leaving the paper to cross over to begin the "d" downstroke. There is



Figure 11.5. The residual functions  $\epsilon_{xi}(t)$  for the X-coordinate. Shaded areas indicate periods when the pen is off the writing surface.

another forced point at the cusp at the end of the "d" downstroke, again just as the pen leaves the writing surface to begin the loop part of "d." We see the largest deviation as the pen leaves the writing surface to cross over to begin "a." In summary, forcing events coincide either with points of sharp curvature or cusps, or with the pen leaving the writing surface. The change in the frictional forces as the pen leaves the surface seems to be an important part of the dynamics.

#### 11.5 Assessing the fit of the equation

Now we want to see how well this equation fits the data. One way to do this is to work with the regression concept, and calculate the squared multiple correlation measure, or proportion of variability explained,

$$R_X^2 = 1 - \frac{\sum_{i=1}^N \int_0^T \epsilon_{xi}^2(t) dt}{\sum_{i=1}^N \int_0^T D^3 x_i^2(t) dt}.$$
(11.5)

The values we obtain are 0.991, 0.994, and 0.994 for the X-, Y-, and Z-coordinates, respectively, indicating a very good fit in all three cases.

As the value of  $R_X^2$  indicates, the residual functions are much smaller overall than the original third derivatives  $D^3x_i$ . However, integrating across time in (11.5) risks missing something interesting that might occur at some specific points in time. Figure 11.5 plots the 20 residual functions for the X-coordinate, and we see that these are small relative to the size of the third derivative, and that they are concentrated around zero. They seem to behave as random "noise" functions that do not contain any systematic variability that we have failed to fit. This investigation shows that the dynamic model generally fits the data extremely well, and invites the suggestion that the coefficient functions characterize the particular subject in some way, and hence can be used as the basis of a classification method in preference to direct consideration of the handwriting itself. This is the theme of our next section.

However, we do notice that there are some sharp excursions in the forcing functions, with a couple of the largest being associated with the beginnings of intervals when the pen is off the paper. It may be that the change in frictional forces plus the effect of raising the pen can have a noticeable effect on printing dynamics in the X-Y plane. Maybe things would be simpler if we only used cursive handwriting, and you can consult Ramsay (2000) to compare these results with that situation.

# 11.6 Classifying writers by using their dynamic equations

We can now estimate a linear differential equation to describe the data of different people printing the same characters. How well does one person's equation model another person's data? We now introduce a second subject, called "CC," and consider a set of 20 replications of CC's printing of the characters "fda." In order to ensure that both dynamic models are defined on compatible time scales, CC's data are preprocessed by being registered to the mean curve of the registered JR data. Thus all the data are registered to the same template. After this preprocessing step, a dynamic model for CC's printing is estimated in the way set out above. We now apply the equation for subjects JR and CC to the data for themselves<sup>2</sup> and for each other.

Figure 11.6 shows the X-coordinate residual functions  $\epsilon_{xi}(t)$  resulting from applying the equation for subjects JR and CC to both sets of data. Corresponding results for the Y-coordinate are shown in Figure 11.7. What we see in the figures is that the residual functions are much larger, and

 $<sup>^{2}</sup>$ In the case where the equation is applied to the subject's own data, we reestimated the equation 20 times by dropping each record out in turn, estimating the equation for the remainder of the data, and then applying the equation to the excluded record. This standard leaving-one-out procedure gives a more honest estimate of how the equation will work in practice than the approach where the test record is included in the estimating set.



Figure 11.6. The residual functions for the X-coordinate resulting from applying both JR's and CC's differential operators to both sets of data.

also have strong systematic patterns, when they result from applying the equation estimated for one person to the data of the other. The self-applied residual functions for JR are rather smaller than those for CC, and two of the CC curves yield self-applied residual functions that are considerably larger in places. Subject CC seems to have altered his printing style in some important respect in these two anomalous cases. Thus, this technology also may be useful for detecting when people alter in some fundamental way how they print or write a sequence of characters.



Figure 11.7. The residual functions for the Y-coordinate resulting from applying both JR's and CC's differential operators to both sets of data.

Figure 11.8 investigates a simple numerical summary based on these results. We assessed the magnitude of the residual functions by computing the square root of their average squared values. As well as averaging across time, we average across all three coordinates in order to obtain a single number quantifying the residuals in Figures 11.6 and 11.7 and the corresponding residuals for the Z-coordinate. The figure uses box plots to show the distribution of these magnitudes for the four situations. We see that the JR operators decisively separate the magnitudes for the two subjects'



Figure 11.8. Box plots of the root mean square magnitudes of the residual functions resulting from applying both JR's and CC's differential operators to both sets of data. For each replication, the root mean square is calculated taking the average over time and all three coordinates.

printing. The largest value for JR's own data is about 12, and the smallest value for CC's data is about 45.

When the CC operators are used, the subtlety of the data becomes clearer. Using a cutoff value of 20, say, the 18 nonanomalous CC printings are clearly separated from the JR printings. On the other hand, the two anomalous printings are very badly modeled by the CC operators (estimated each time leaving out the individual datum in question). If we were using this simple numerical summary to classify the data, we would presumably categorize these two data as being written by neither CC or JR. If we look back to Figure 11.7, however, we can see that even the two anomalous curves yield residual curves that are near zero over the part of the range [0.3, 1.2]. We do not pursue this further in the present study, but it demonstrates that attempts to fool the dynamic model may not always be totally successful on closer examination.

## 11.7 What have we seen?

The methods of functional data analysis are especially well suited to studying the dynamics of processes that interest us. We saw this previously in our phase-plane plotting of the nondurable goods index, and now we see that a differential equation is a useful means of modeling this time-varying behavior. Of course, this is already well known in the natural sciences, where differential equations, such as Maxwell's equations for electromagnetic phenomena, emerge as the most elegant and natural means of expressing the laws of physics and chemistry.

But in the natural sciences differential equations emerged painstakingly after much experimentation and observation, and finally some deep thinking about the way the interplay of forces along with the law of conservation of energy might determine the results of these experiments. Now, however, we are evolving methods for estimating these equations directly from often noisy data, and in situations such as economics and biomechanics where fundamental laws will not be straightforward and may not even be possible. In this chapter we have put our empirically estimated differential equations to work to investigate an interesting practical problem, the identification of an individual by the dynamic characteristics of a sample of his or her behavior. The next chapter applies this idea to some rather more complex biomechanical data.