18

Differential equations and operators

18.1 Introduction

The derivatives of functional observations have played a strong role from the beginning of this book. For example, we chose to work with acceleration directly rather than height for growth curves and handwriting coordinate functions, and to inspect functions $(\pi/6)^2 D\,\texttt{Temp} + D^3\texttt{Temp}$ for temperature profiles. We used $D^2\beta$ as a measure of curvature in an estimated regression function β so as to regularize or smooth the estimate, and applied this idea in functional principal components analysis, canonical correlation, and various types of linear models. When the objective was a smooth estimate of a derivative $D^m x$, we used $D^{m+2}x$ to define the roughness penalty. Thus, derivatives can be used both as the object of inquiry and as tools for stabilizing solutions.

In Chapter 17, we introduced the idea of incorporating derivatives into linear models for functional data. We saw that this permitted a model for the simultaneous variation in a function and one or more of its derivatives, and in the oil refinery example in Section 17.2, the approach came up with an elegant little model with only two parameters that fit the data beautifully.

It is time to look more systematically at how derivatives might be employed in modelling functional data. Are there other ways of using derivatives, for example? Can we use mixtures of derivatives instead of simple derivatives? Can we extend models so that derivatives can be used on either the covariate or response side? Can our smoothing and regularization techniques be extended in useful ways? Are new methods of analysis making explicit use of derivative information possible?

This chapter provides some background on differential equations and their use in applications. Readers either considering differential equations for the first time or whose memories of their first contact has dimmed may appreciate this material. We begin with the simplest of input/output systems commonly described by a differential equation. After considering possible extensions, we review how linear differential operators may be used in various ways and some basic theory. The last three sections, on constraint functionals, Green's functions and reproducing kernels, are more advanced. They may therefore be profitable to those already having a working knowledge of this field. We nevertheless consider these topics to be of potential importance for statistical applications, and they play a role in subsequent chapters.

18.2 Exploring a simple linear differential equation

An input/output system has an input function u that in some way modifies an output function x. Perhaps you might like to return to the refinery data in Figure 1.4 for an example.

Here is the simplest prototype for such equations:

$$Dx(t) = -\beta x(t) + \alpha u(t) + \epsilon(t). \tag{18.1}$$

This is a functional linear model in which the dependent variable is the derivative of output x, and the two independent variables are x itself and input function u. To keep things as simple as possible, we have specified that the regression coefficient functions are constant. Function ϵ allows for noise and other forms of ignorable variation in the functional data. It is a useful convention to place a minus before terms on the right side involving output function x; most real-life systems modelled by differential equations have positive values of β if we do this, reflecting their natural tendency to return to their resting state.

We could, however, make things even simpler by dropping u from the equation. Situations do arise where the goal is to model the behavior of a function x and its derivatives without considering any external influences. The no-input version of the equation,

$$Dx(t) = -\beta x(t) + \epsilon(t), \tag{18.2}$$

is said to be *homogeneous*, while (18.1) is called *nonhomogeneous.* Input function u, when it is present, is often called a *forcing function*, and the homogeneous version of the equation is said to be *forced* by αu.

Let x_0 be a solution to the homogeneous equation. Given parameter β and assuming that the noise function ϵ is zero, a moment of reflection reveals that the solution to $Dx_0 = -\beta x_0$, is

$$x_0(t) = Ce^{-\beta t}$$

for some nonzero constant C. If we knew the value of x_0 at time $t = 0$, then $C = x_0(0)$ and the solution is completely determined.

It is a bit harder to work out the solution of (18.1), or $Dx = -\beta x + \alpha u$, but here it is:

$$x(t) = Ce^{-\beta t} + \alpha \int_0^t e^{-\beta(t-s)} u(s)\, ds \ . \tag{18.3}$$

As with the homogeneous equation, constant C is simply $x(0)$.

A graph helps us to see the role played by the two parameters α and β. Engineers often study how an industrial process reacts to changes in its inputs by stepping these inputs up or down abruptly. Accordingly, let $u(t) = 0$ for $0 \leq t \leq 1$, and $u(t) = 1$ for $t > 1$. Also, let's set $x(0) = C = 1$. Then solution (18.3) becomes

$$\begin{aligned} x(t) &= e^{-\beta t}, 0 \leq t \leq 1, \\ &= e^{-\beta t} + (\alpha/\beta)[1 - e^{-\beta(t-1)}], t > 1. \end{aligned}$$

Figure 18.1 shows the solution x for $\beta = 2$, and $\beta = 4$, while fixing $\alpha/\beta = 2$. Over the first half of the interval, x behaves like x_0, and we see that the solution decays to zero in about $4/\beta$ time units. Over the second half of the interval, the solution grows at an exponentially decreasing rate towards an upper asymptote of α/β, often called the *gain* of the system. Again, the gain level is achieved in about $4/\beta$ time units. The role of β is now clear; it determines the rate of change in x in response to a step change in u.

We can summarize the roles of these two parameters by comparing α to the volume control on a radio playing a song carried by radio signal u; the bigger α, the louder the sound. The bass/treble control, on the other hand, corresponds to β; the larger β, the higher the frequency of what you hear.

We may rearrange differential equation (18.1) to put it in the form

$$Lx(t) = \beta x(t) + Dx(t) - \alpha u(t) - \epsilon(t) \ . \tag{18.4}$$

Function x is a solution of the original equation when $\epsilon = 0$ if and only if $Lx = 0$. We call $L = \beta I + D$, where I is the identity operator, or $Ix = x$, a *linear differential operator*, in this case with constant coefficients. This alternative expression of the differential equation is handy, as we now know, for defining roughness penalties, and using the roughness penalty

$$\mathtt{PEN}(x) = \int [Lx(t)]^2\, dt$$

is equivalent to penalizing the failure of x to satisfy the differential equation $Dx = -\beta x$ corresponding to operator L.

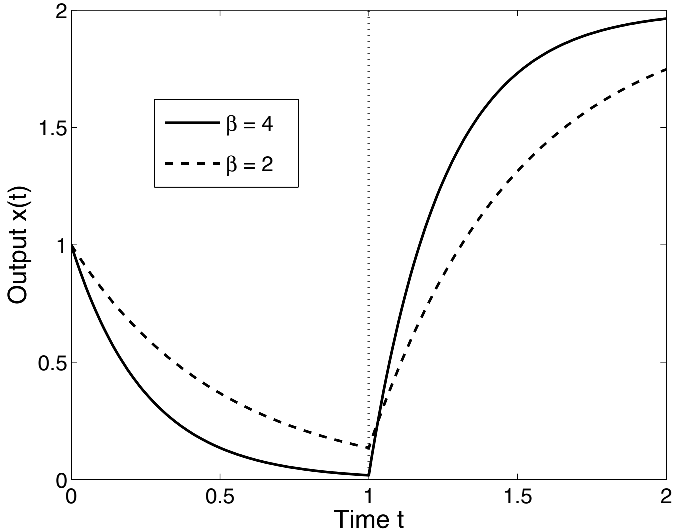

Figure 18.1. The solid and dashed lines are two solutions (18.3) to a first-order constant-coefficient differential equation for two different values of the rate constant β.

18.3 Beyond the constant coefficient first-order linear equation

18.3.1 Nonconstant coefficients

Returning to Figure 18.1, we might be struck by an anti-symmetry: The rate of decay over the first interval, $Dx = -\beta e^{-\beta t}$ is the negative of the rate of increase over the second, $Dx = \beta e^{-\beta t}$. Many systems, however, increase more rapidly than they decrease, or vice versa. We acquire common cold symptoms within hours and take days to recover from them, for example. This suggests that allowing β to vary over time might be useful, and similar arguments could be made for α. Then (18.1) becomes

$$Dx(t) = -\beta(t)x(t) + \alpha(t)u(t) + \epsilon(t). \tag{18.5}$$

The solution to (18.5) is

$$x(t) = Cx_0(t) + \int_0^t \alpha(s)u(s)x_0(t)/x_0(s)\,ds, \tag{18.6}$$

where

$$x_0(t) = \exp[-\int_0^t \beta(s)\,ds].$$

The functional ratio $x_0(t)/x_0(s)$ that occurs in the second term of (18.6) defines the Green's function for the differential equation; see Section 20.2 for details.

18.3.2 Higher order equations

More generally, the order of the derivative on the left side of (18.5) may be m, with lower order derivatives appearing in the right side:

$$\begin{aligned} D^m x(t) &= -\beta_0(t)x(t) - \beta_1(t)Dx(t) - \ldots - \beta_{m-1}(t)D^{m-1}x(t) \\ &\quad +\alpha(t)u(t) + \epsilon(t) \\ &= -\sum_{j=0}^{m-1} \beta_j(t)D^j x(t) + \alpha(t)u(t) + \epsilon(t). \end{aligned} \tag{18.7}$$

These higher order systems are needed when there are more than two time scales for events. This means that, in the case of a second order system, there is a long-term trend, medium-term changes, and sharper shorter-term events.

Figure 18.2 shows the forced second order equation

$$D^2x(t) = -4.04x(t) - 0.4Dx(t) + 2u(t), \tag{18.8}$$

where forcing function $u(t)$ steps from 0 to 1 at time $t = 2\pi$. The corresponding homogeneous solution is

$$x_0(t) = e^{-0.2t}[\sin(2t) + \cos(2t)].$$

There are three time scales involved here. The longest scale is the overall oscillation level, first about 0 and then later about 0.5. The medium scale trend is the exponential decay in the amplitude of the oscillation, and of course the shortest scale is the oscillation with period π.

Consider handwriting; Ramsay (2000) observed that there were features in script at four time scales:

1. The overall spatial position of the script, that is, the line on which it is written, requiring some considerable seconds per line.

2. The movement of the script from left to right within a line, taking place over several seconds.

3. The strokes and loops within the script, produced about eight times a second.

4. Sharper transient effects due to the pen striking or leaving the paper, lasting of the order of 10 milliseconds.

The differential equation developed in this study consequently was of the third order.

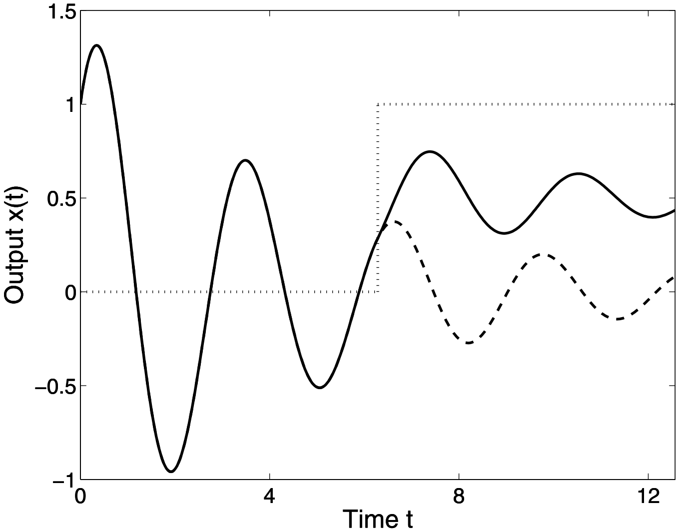

Figure 18.2. The solid line is the solution to the second order equation (18.8). The dashed line is the corresponding homogeneous equation solution, and the dotted line indicates the step function u forcing the equation.

18.3.3 Systems of equations

Often the processes that we study produce more than one output, and so we need several output functions x_i. As an example, suppose that $\beta(t)$ in (18.7) is also affected by $u(t)$, and that we can develop a differential equation that defines its behavior. We now have two differential equations, one for x and one for β.

Or, as another example, suppose that an engineer develops a feedback loop for the process permitting the output x to have an effect on the input u. For example, a doctor adjusts the medication u of the patient according to changes in the symptoms x. Then u and x can each be expressed as a differential equation, and in each equation the other variable now plays the role of an input. That is,

$$\begin{aligned} Dx(t) &= -\beta_x(t)x(t) + \alpha_x(t)u(t) \\ Du(t) &= -\beta_u(t)u(t) + \alpha_u(t)x(t). \end{aligned} \tag{18.9}$$

In fact, any differential equation of order m can be expressed as a system of m first-order equations. For a second order system,

$$D^2x(t) = -\beta_0(t)x(t) - \beta_1(t)Dx(t), \tag{18.10}$$

for example, define $y(t) = Dx(t)$. Then we have the equivalent system of two linear differential equations

$$\begin{aligned} Dx(t) &= y(t) \\ Dy(t) &= -\beta_1(t)y(t) - \beta_0(t)x(t). \end{aligned} \tag{18.11}$$

18.3.4 Beyond linearity

Equation (18.7) is a linear differential equation in the sense that each derivative or input function is multiplied by a coefficient function, and the products added to yield the output. That is, it is linear in the same sense that the models in Chapters 12 to 16 are linear.

The general form of a nonlinear differential equation of the first order is

$$Dx(t) = f[t, x(t), u(t)]$$

for some function f.

Linear differential equations are easier to work with. They have solutions for all values of t, and their properties are much better understood by mathematicians than nonlinear equations. However, simple nonlinear systems can define remarkable and often complex behavior in a solution x. The world of *nonlinear dynamics* is vast and fascinating, but unfortunately beyond the scope of this book.

The term "linear" is often used in engineering and elsewhere to refer only to linear constant coefficient systems. In this restricted case, the use of Laplace transformations leads to expressing the behavior of solutions in terms of *transfer functions.*

18.4 Some applications of linear differential equations and operators

In this section, we review a number of ways in which linear differential equations and operators are useful in functional data analysis. Many of these we have already encountered, but a few new ones are also suggested. We will assume that the linear differential operator is in the form

$$Lx = \sum_{j=0}^{m-1} \beta_j D^j x + D^m x. \tag{18.12}$$

18.4.1 Differential operators to produce new functional observations

Derivatives of various orders and mixtures of them are of immediate interest in many applications. We have already noted that there is much

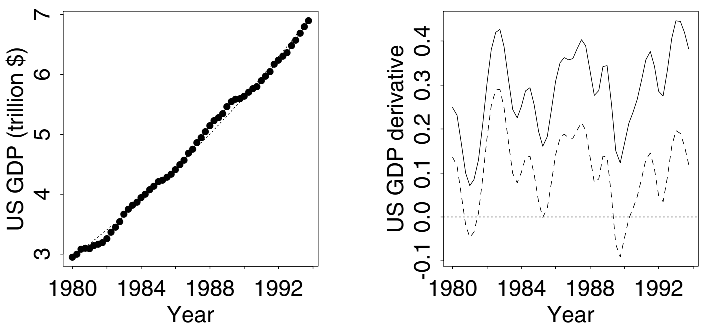

Figure 18.3. The left panel shows the gross domestic product of the United States in trillion US dollars. The solid curve mostly obscured by the dots is a polynomial smoothing spline constructed with a penalty on the integrated squared fourth derivative, and the dotted curve is a purely exponential trend fit by least squares. The solid curve in the right panel is the estimated first derivative of GDP. The dashed curve in this panel is the value of the differential operator $L = \beta\,\mathtt{GDP} + D\,\mathtt{GDP}$.

to be learned about human growth by examining acceleration profiles. There is an analogy with mechanical systems; a version of Newton's third law, $a(t) = F(t)/M$, asserts that the application of some force $F(t)$ at time t on an object with mass M has an immediate impact on acceleration $a(t)$. However, force has only an indirect impact on velocity, through $v(t) = v_0 + M^{-1} \int_0^t F(u)du$, and an even less direct impact on what we directly observe, namely position, $s(t) = s_0 + v_0 t + M^{-1} \int_0^t \int_0^u F(z)\,dz\,du$. From the standpoint of mechanics, the world that we experience is two integrals away from reality! The release of adrenal hormones during puberty tends to play the role of the force function F, and so does a muscle contraction with respect to position of a limb or other part of the body.

18.4.2 The gross domestic product data

The gross domestic product (GDP) of a country is the financial value of all goods and services produced in that country, whether by the private sector of the economy or by government. Like most economic measures, GDP tends to exhibit a percentage change each year in times of domestic and international stability. Although this change can fluctuate considerably from year to year, over long periods the fluctuations tend to even out for most countries and the long-range trend in GDP tends to be roughly exponential.

We obtained quarterly GDP values for 15 countries in the Organization for Economic Cooperation and Development (OECD) for the years 1980

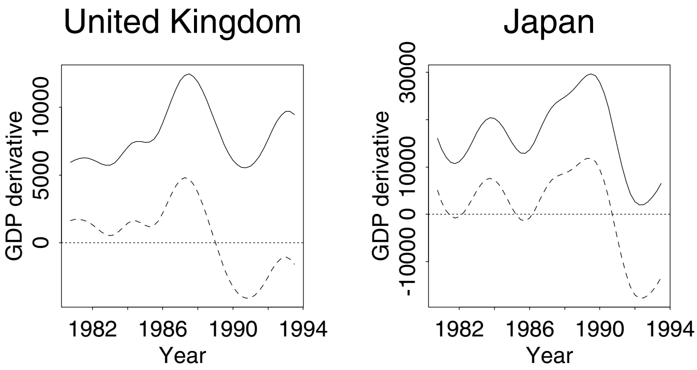

Figure 18.4. The solid curves are the derivatives of GDP of the United Kingdom and Japan estimated by order 4 smoothing splines. The dashed curves are the corresponding values of the differential operator $L = \beta x + Dx$.

through 1994 (OECD, 1995). The values for any country are expressed in its own currency, and thus scales are not comparable across countries. Also, there are strong seasonal effects in GDP values reported by some countries, whereas others smooth them out before reporting.

The left panel of Figure 18.3 displays the GDP of the United States. The seasonal trend, if any, is hardly visible, and the solid line indicates a smooth of the data using a penalty on D^4 `GDP`. It also shows a best fitting exponential trend, $C\exp(\gamma t)$, with rate constant $\gamma = 0.038$. Thus, over this period the U.S. economy tended to grow at about 4% per year. The right panel displays the first derivative of GDP as a solid line. The economy advanced especially rapidly in 1983, 1987 and 1993, but there were slowdowns in 1981, 1985 and 1990.

If we define Lx to be $\beta x + Dx$, then we may say even more compactly that $Lx = 0$ implies exponential growth. When studying processes that exhibit exponential growth or decay to some extent, it can be helpful to look at Lx defined in this way; the extent to which the result is a nonzero function with substantial variation is a measure of departure from exponential growth, just as the appearance of a nonzero phase in D^2x for a mechanical system indicates the application of a force.

The right panel of Figure 18.3 shows the result of applying this differential operator to the U.S. GDP data. The result is clearly not zero; there seem to be three cycles of shorter term growth in GDP that depart from the longer-term exponential trend. Figure 18.4 shows the comparable curves for the United Kingdom and Japan, and we note that the U.K. had only one boom period with an uncertain recovery after the recession, while Japan experienced a deep and late recession.

18.4.3 Differential operators to regularize or smooth models

Although we have covered this topic elsewhere, we should still point out that we may substitute Lx for D^2x in any of the regularization schemes covered so far. Why? The answer lies in the homogeneous equation $Lx = 0$; functions satisfying this equation are deemed to be *ultrasmooth* in the sense that we choose to ignore any component of variation of this form in calculating roughness or irregularity. In the case of the operator D^2, linear trend is considered to be so smooth that any function may have an arbitrary amount of it, since the penalty term $\lambda \int (D^2x)^2$ is unaffected. Suppose, on the other hand, that we are working with a process that is predominantly exponential growth with rate parameter β. We may choose in this case to do nonparametric regression with the fitting criterion

$$\mathtt{PENSSE}_\lambda(x) = n^{-1} \sum_{j=1}^{n} [y_j - x(t_j)]^2 + \lambda \int [\beta x(t) + Dx(t)]^2 \, dt$$

in order to leave untouched any component of variation of this form.

More generally, suppose we observe a set of discrete data values generated by the process

$$y_j = x(t_j) + \epsilon_j,$$

where, as in previous chapters, x is some unobserved smooth function that we wish to estimate by means of nonparametric regression, and ϵ_j is a disturbance or error assumed to be independently distributed over j and to have mean zero and finite variance. Suppose, moreover, that we employ the general smoothing criterion

$$\mathtt{PENSSE}_\lambda(\hat{x}) = n^{-1} \sum_{j} [y_j - \hat{x}(t_j)]^2 + \lambda \int (L\hat{x})^2(t) \, dt \tag{18.13}$$

for some differential operator L.

It is not difficult to show (see Wahba, 1990) that, if we choose $\hat{x}$ to minimize $\mathtt{PENSSE}_\lambda$, then the integrated squared bias

$$\mathtt{Bias}^2(\hat{x}) = \{ \int \mathrm{E}[\hat{x}(t) - x(t)] \, dt \}^2$$

cannot exceed $\int (Lx)^2(t) \, dt$. This is useful, because if we choose L so as to approximate $Lx = 0$, then the bias is likely to be small. It then follows that we can use a relatively large value of the smoothing parameter λ, leading to lower variance, without introducing excessive bias. Also, we can achieve a small value of the integrated mean squared error

$$\mathtt{IMSE}(\hat{x}) = \int \mathrm{E}[\hat{x}(t) - x(t)]^2 \, dt$$

since

$$\mathtt{IMSE}(\hat{x}) = \mathtt{Bias}^2(\hat{x}) + \mathtt{Var}(\hat{x}),$$

where

$$\texttt{Var}(\hat{x}) = \int \mathrm{E}\{\hat{x}(t) - \mathrm{E}[\hat{x}(t)]\}^2 \, dt.$$

The conclusion, therefore, is that if we have any prior knowledge at all about the predominant shape of x, it is worth choosing a linear differential operator L so as to annihilate functions having that shape. We show how to construct customized spline smoothers of this type in the next two chapters.

This insight about the role of L in the regularization process also leads to the following interesting question: Can we use the information in N replications x_i of functional observations such as growth or temperature curves to *estimate* an operator L that comes close in some sense to satisfying $Lx_i = 0$? If so, then we should certainly use this information to improve on our smoothing techniques. This matter is taken up in Chapter 21.

18.4.4 Differential operators to partition variation

Linear differential operators L of the form (18.12) of degree m have m linearly independent solutions ξ_j of the homogeneous equation $L\xi_j = 0$. There is no unique way of choosing these m functions ξ_j, but any choice is related by a linear transformation to any other choice. The set of all functions z for which $Lz = 0$ is called the *null space* of L, and the functions ξ_j form a basis for this space. The notation $\ker L$ is often used to indicate this null space.

Consider, for example, the derivative operator $L = D^m$: The monomials $\{1, t, \dots t^{m-1}\}$ are a basis for the null space, as is the set of m polynomials formed by any nonsingular linear transformation of these. Likewise the functions $\{1, e^{-\beta t}\}$ are a solution set for $\beta Dx + D^2x = 0$. And $\{1, \sin\omega t, \cos\omega t\}$ were cited as the solution set or null space functions for $Lx = \omega^2 Dx + D^3 x = 0$ in Chapter 1.

This means, then, that we can use linear differential operators L to partition functional variation in the sense that Lx splits x into two parts, the first consisting of what is in x that can be expressed in terms of a linear combination of the null space functions ξ_j, and the second being whatever is orthogonal to these functions.

This partitioning of variation is just what happens, as we already know from Section 4.4, with basis functions ϕ_k and the projection operator P that expands x as a linear combination of these basis functions. That is,

$$Px = \hat{x} = \sum_{k=1}^{m} c_k \phi_k$$

also splits any function x into the component $\hat{x}$ that is an optimal combination of the basis functions in a least squares sense, and an orthogonal residual component $x - \hat{x} = (I - P)x$. The complementary projection operator $Q = I - P$ therefore satisfies the linear homogeneous equation $Q\hat{x} = 0$, as well as the m equations $Q\phi_k = 0$. Thus the projection operator Q and the differential operator L have analogous properties.

But there are some important differences, too. First, the projection operator P does not pay any attention to derivative information, whereas L does. Second, we have the closely related fact that Q is chosen to make Qx small, while L is chosen to make Lx small. Since Lx involves derivatives up to order m, making Lx small inevitably means paying attention to the size of $D^m x$. If we think there is important information in derivatives, then it seems right to exploit this in splitting variation.

It is particularly easy to compare the two operators, differential and projection, in situations where there is an orthonormal basis expansion for the function space in question. Consider, for example, the space of infinitely differentiable periodic functions defined on the interval $[0, 1]$ that would be natural to model our temperature and precipitation records. A function x has the Fourier expansion

$$x(t) = c_0 + \sum_{k=1}^{\infty} [c_{2k-1} \sin(2\pi kt) + c_{2k} \cos(2\pi kt)].$$

Suppose our two operators L and Q are of order 3 and designed to eliminate the first three terms of the expansion, that is, vertically shifted sinusoidal variation of period 1. Then

$$Qx(t) = \sum_{k=2}^{\infty} [c_{2k-1} \sin(2\pi kt) + c_{2k} \cos(2\pi kt)]$$

while

$$\begin{aligned} Lx &= 4\pi^2 Dx + D^3 x \\ &= \sum_{k=2}^{\infty} 8\pi^3 k(k^2 - 1)[-c_{2k-1} \cos(2\pi kt) + c_{2k} \sin(2\pi kt)]. \end{aligned}$$

Note that applying Q does not change the expansion beyond the third term, while L multiplies each successive pair of sines and cosines by an ever-increasing factor proportional to $k(k^2 - 1)$. Thus, L actually accentuates high-frequency variation while Q leaves it untouched; functions that are passed through L are going to come out rougher than those passing through Q.

The consequences for smoothing are especially important: If we penalize the size of $\|Lx\|^2$ in spline smoothing by minimizing the criterion (18.13), the roughening action of L means that high-frequency components are forced to be smaller than they would be in the original function, or than they would be if we penalized using Q by using the criterion

$$\texttt{PENSSE}_\lambda^Q(\hat{x}) = n^{-1} \sum_j [y_j - \hat{x}(t_j)]^2 + \lambda \int (Q\hat{x})^2(t)\, dt. \tag{18.14}$$

But customizing a regularization process is only one reason for splitting functional variation, and in Chapter 19 we look at a differential operator analogue of principal components analysis, called *principal differential analysis*, that can prove to be a valuable exploratory tool.

18.4.5 Operators to define solutions to problems

We have already considered a number of situations in Chapter 6 requiring smoothing functions x that had constraints such as positivity, monotonicity, values in (0,1), and so forth. We saw there that functions having these constraints can often be expressed as solutions to linear or nonlinear differential equations. This insight helped us to modify conventional linear least squares smoothing methods to accommodate these constraints.

18.5 Some linear differential equation facts

So far in this chapter, we have set the scene for the use of linear differential operators and equations in FDA. We now move on to a more detailed discussion of techniques and ideas that we use in this and the following chapters. Readers with some familiarity with the theory of linear ordinary differential equations may wish to skip on to the next two chapters, and refer back to this material only where necessary.

18.5.1 Derivatives are rougher

First, it is useful to point out a few things of general importance. For example, taking a derivative is generally a roughening operation, as we have observed in the context of periodic functions. This means that Dx has in general rather more curvature and variability than x. It is perhaps unfortunate that our intuitions about functions are shaped by our early exposure to polynomials, where derivatives are smoother than the original functions, and transcendental functions such as e^t and $\sin t$, where taking derivatives produces essentially no change in shape. In fact, the general situation is more like the growth curve accelerations in Figure 1.2, which are much more variable than the height curves in Figure 1.1, or the roughening effect of applying the third order linear differential operator to temperature functions displayed in Figure 1.7.

By contrast, the operation of partial integration essentially reverses the process of differentiation (except for the constant of integration), and therefore is a smoothing operation. It is convenient to use the notation $D^{-1}x$ for

$$D^{-1}x(t) = \int_{t_0}^{t} x(s)\, ds,$$

relying on context to specify the lower limit of integration t_0. This means, of course, that $D^{-1}Dx = x$.

18.5.2 Finding a linear differential operator that annihilates known functions

We have already cited a number of examples where we had a set of known functions $\{\xi_1, \ldots, \xi_m\}$ and where at the same time we were aware of the operator L that solved the homogeneous linear differential equations $L\xi_j = 0, j = 1, \ldots, m$. Suppose, however, that we have the ξ_j's in mind but that the L that annihilates them is not obvious, and we want to find it.

The process of identifying the linear differential operator that sets m linearly independent functions to 0, as well as other aspects of working with linear differential operators, can be exhibited through the following example: Suppose we are considering an amplitude-modulated sinusoidal signal with fixed period ω. Such a signal would be of the form

$$x(t) = A(t)[c_1 \sin(\omega t) + c_2 \cos(\omega t)]. \tag{18.15}$$

The function A determines the amplitude pattern. If A is regarded as a known time-varying function, the constants c_1 and c_2 determine the overall size of the amplitude of the signal and also the phase of the signal.

Our aim, for given ω and $A(t)$, is to find a differential operator L such that the null space of L consists of all functions of the form 18.15. Because these functions form a linear space of dimension 2, we seek an annihilating operator of order 2, of the form

$$Lx = \beta_0 x + \beta_1 Dx + D^2 x.$$

The task is to calculate the two weight functions β_0 and β_1.

First, let's do a few things to streamline the notation. Define the vector functions $\boldsymbol{\xi}(t)$ and $\boldsymbol{\beta}(t)$ as

$$\boldsymbol{\xi}(t) = \begin{bmatrix} A(t)\sin(\omega t) \\ A(t)\cos(\omega t) \end{bmatrix} \text{ and } \boldsymbol{\beta}(t) = \begin{bmatrix} \beta_0(t) \\ \beta_1(t) \end{bmatrix}. \tag{18.16}$$

Also, let us use $\mathtt{S}(t)$ to stand for $\sin(\omega t)$ and $\mathtt{C}(t)$ for $\cos(\omega t)$. Then

$$\boldsymbol{\xi} = \begin{bmatrix} A\mathtt{S} \\ A\mathtt{C} \end{bmatrix}. \tag{18.17}$$

The required differential operator L satisfies the vector equation $L\boldsymbol{\xi} = 0$.

Recall that the first and second derivatives of $\mathtt{S}$ are $\omega\mathtt{C}$ and $-\omega^2\mathtt{S}$, respectively, and that those of $\mathtt{C}$ are $-\omega\mathtt{S}$ and $-\omega^2\mathtt{C}$, respectively. Then the first two derivatives of ζ are, after a bit of simplification,

$$D\boldsymbol{\xi} = \begin{bmatrix} (DA)\mathtt{S} + \omega A\mathtt{C} \\ (DA)\mathtt{C} - \omega A\mathtt{S} \end{bmatrix}$$

and

$$D^2\boldsymbol{\xi} = \begin{bmatrix} (D^2A)\mathtt{S} + 2\omega(DA)\mathtt{C} - \omega^2 A\mathtt{S} \\ (D^2A)\mathtt{C} - 2\omega(DA)\mathtt{S} - \omega^2 A\mathtt{C} \end{bmatrix}. \tag{18.18}$$

The relation $L\boldsymbol{\xi} = 0$ can be expressed as follows, by taking the second derivatives over to the other side of the equation:

$$\beta_0\boldsymbol{\xi} + \beta_1 D\boldsymbol{\xi} = -D^2\boldsymbol{\xi} \tag{18.19}$$

or, in matrix notation

$$\begin{bmatrix} \boldsymbol{\xi} & D\boldsymbol{\xi} \end{bmatrix} \boldsymbol{\beta} = -D^2\boldsymbol{\xi}. \tag{18.20}$$

This is a linear matrix equation for the unknown weight functions β_0 and β_1, and its solution is simple provided that the matrix

$$\mathbf{W}(t) = \begin{bmatrix} \boldsymbol{\xi}(t) & D\boldsymbol{\xi}(t) \end{bmatrix} \tag{18.21}$$

is nowhere singular, or in other words that its determinant $|\mathbf{W}(t)|$ does not vanish for any value of the argument t. This coefficient matrix, which plays an important role in linear differential operator theory, is called the *Wronskian matrix*, and its determinant is called the *Wronskian* for the system.

Substituting the specific functions $A\mathtt{S}$ and $A\mathtt{C}$ for this example for ξ_1 and ξ_2 gives

$$\mathbf{W} = \begin{bmatrix} A\mathtt{S} & (DA)\mathtt{S} + \omega A\mathtt{C} \\ A\mathtt{C} & (DA)\mathtt{C} - \omega A\mathtt{S} \end{bmatrix}. \tag{18.22}$$

Thus the Wronskian is

$$|\mathbf{W}| = A\mathtt{S}[(DA)\mathtt{C} - \omega A\mathtt{S}] - A\mathtt{C}[(DA)\mathtt{S} + \omega A\mathtt{C}] = -\omega A^2 \tag{18.23}$$

after some simplification. We have no worries about the singularity of $\mathbf{W}(t)$, then, so long as the amplitude function $A(t)$ does not vanish.

The solutions for the weight functions are then given by

$$\boldsymbol{\beta} = -\mathbf{W}^{-1}D^2\boldsymbol{\xi}.$$

This takes a couple of sheets of paper to work out, or may be solved using symbolic computation software such as Maple (Char et al. 1991) or Mathematica (Wolfram, 1991). Considerable simplification is possible because of the identity $\mathtt{S}^2 + \mathtt{C}^2 = 1$, and the final result is that

$$\boldsymbol{\beta} = \begin{bmatrix} \omega^2 + 2(DA/A)^2 - D^2A/A \\ -2DA/A \end{bmatrix},$$

so that, for any function x,

$$Lx = [\omega^2 + 2(DA/A)^2 - D^2A/A]x - 2[(DA)/A](Dx) + D^2x. \tag{18.24}$$

Note that the weight coefficients in (18.24) are, as we should expect, scale free in the sense that multiplying $A(t)$ by any constant does not change them.

Consider two simple possibilities for amplitude modulation functions. When $A(t)$ is a constant, both derivatives vanish, the operator reduces to $L = \omega^2 I + D^2$ and $Lx = 0$ is the equation for simple harmonic motion. On the other hand, if $A(t) = e^{-\lambda t}$ so that the signal damps out exponentially with rate λ, then things simplify to

$$\boldsymbol{\beta} = \left[\begin{array}{l} \omega^2 + \lambda^2 \\ 2\lambda \end{array} \right] \text{ or } Lx = (\omega^2 + \lambda^2)x + 2\lambda Dx + D^2 x. \tag{18.25}$$

This is the equation for damped harmonic motion with a damping coefficient 2λ.

The example illustrates the following general principles: First, the order m Wronskian matrix

$$\mathbf{W}(t) = \left[\begin{array}{cccc} \boldsymbol{\xi}(t) & D\boldsymbol{\xi}(t) & \dots & D^{m-1}\boldsymbol{\xi}(t) \end{array} \right] \tag{18.26}$$

must be invertible, implying that its determinant should not vanish over the range of t being considered. There are ways of dealing with isolated singularities, however. Second, finding the vector of weight functions $\boldsymbol{\beta} = (\beta_0(t), \dots, \beta_{m-1}(t))'$ is then is a matter of solving the system of m linear equations

$$\mathbf{W}(t)\boldsymbol{\beta}(t) = -D^m \boldsymbol{\xi}(t),$$

again with the possible aid of symbolic computation software.

18.5.3 Finding the functions ξ_j satisfying $L\xi_j = 0$

Let us now consider the problem converse to that considered in Section 18.5.2. Given a linear differential operator L of order m, we might wish to identify m linearly independent solutions ξ_j to the homogeneous equation $Lx = 0$. We can do this directly by elementary calculus in simple cases, but more generally there is a variety of analytic and numerical approaches to this problem. For full details, see a standard reference on numerical methods, such as Stoer and Bulirsch (2002).

Specifically, given (18.7), a common procedure is to use a numerical differential equation solving algorithm, such as one of the Runge-Kutta methods, to solve the equation for *initial value constraints*, described below, of the form $B_0 x = \mathbf{I}_i$, where $\mathbf{I}_i$ is the ith column of the identity matrix of order m. This will yield m linearly independent solutions ξ_i that can be used as a basis for obtaining all possible solutions.

18.6 Initial conditions, boundary conditions and other constraints

18.6.1 Why additional constraints are needed to define a solution

We have already noted that the space of solutions of the linear differential equation $Lx = 0$ is, in general, a function space of dimension m, called the null space of L, and denoted by $\ker L$. We now assume that the linearly independent functions $\xi_1, \ldots, \xi_m$ form a basis of the null space.

Any specific solution of $Lx = 0$ requires m additional pieces of information about x. For example, we can solve the equation $\beta Dx + D^2x = 0$, defining a shifted exponential, uniquely provided that we are able to specify that

$$x(0) = 0 \text{ and } Dx(0) = 1,$$

in which case

$$x(t) = \frac{1}{\beta}(1 - e^{-\beta t}).$$

Alternatively, $x(0) = 1$ and $Dx(0) = 0$ implies that $x_0 = 1$ and $\alpha = 0$, or simply that $x = 1$.

We introduce the notion of a *constraint operator* B to specify the m pieces of information about x that we require to identify a specific function x as the unique solution to $Lx = 0$. This operator simply *evaluates* x or its derivatives in m different ways. The most important example is the *initial value* operator used in the theory of ordinary differential equations defined over an interval $\mathcal{T} = [0, T]$,

$$\textbf{Initial Operator:}\; B_0 x = \begin{bmatrix} x(0) \\ Dx(0) \\ \vdots \\ D^{m-1}x(0) \end{bmatrix}. \tag{18.27}$$

When B_0x is set to an m-vector, initial value constraints are defined. In the example above, we considered the two cases $B_0x = (0, 1)'$ and $B_0x = (1, 0)'$, implying the two solutions given there.

The following *boundary value* operator is also of great importance in applications involving linear differential operators of even degree:

$$\textbf{Boundary Operator:}\; B_B x = \begin{bmatrix} x(0) \\ x(T) \\ \vdots \\ D^{(m-2)/2}x(0) \\ D^{(m-2)/2}x(T) \end{bmatrix}. \tag{18.28}$$

Specifying $B_B x = c$ gives the values of x and its first $(m-2)/2$ derivatives at both ends of the interval of interest.

The *periodic constraint* operator is

$$\textbf{Periodic Operator:}\quad B_P x = \begin{bmatrix} x(T) - x(0) \\ Dx(T) - Dx(0) \\ \vdots \\ D^{m-1}x(T) - D^{m-1}x(0) \end{bmatrix}. \tag{18.29}$$

Functions satisfying $B_P x = 0$ are periodic up to the derivative D^{m-1} over $\mathcal{T}$, and are said to obey periodic boundary conditions.

The *integral constraint* operator is

$$\textbf{Integral Operator:}\quad B_I x = \begin{bmatrix} \int \xi_1(t)x(t)\,dx \\ \int \xi_2(t)x(t)\,dx \\ \vdots \\ \int \xi_m(t)x(t)\,dx \end{bmatrix}, \tag{18.30}$$

where $\xi_1, \ldots, \xi_m$ are m linearly independent weight functions.

18.6.2 How L and B partition functions

Whatever constraint operator we use, consider the problem of expressing any particular function x as a sum of two components z and e, such that $Lz = 0$ and $Be = 0$. When can we carry out this partitioning in a unique way? This happens if and only if $x = 0$ is the only function satisfying both $Bx = 0$ and $Lx = 0$. Or, in algebraic notation,

$$\ker B \cap \ker L = 0. \tag{18.31}$$

Thus, the two operators B and L complement each other; the equation $Lx = 0$ defines a space of functions $\ker L$ that is of dimension m, and within this space B is a non-singular transformation. Or, looking at it the other way, the equation $Bx = 0$ defines a space of functions $\ker B$ of *codimension* m, within which L is a one-to-one transformation.

Note that the condition (18.31) can break down. Consider, for example, the operator $L = \omega^2 I + D^2$ on the interval $[0, T]$. The space $\ker L$ contains all linear combinations of $\sin \omega t$ and $\cos \omega t$. If $\omega = 2\pi k/T$ for some integer k and we use boundary constraints, all multiples of $\sin \omega t$ satisfy $B_B x = 0$, and so the condition (18.31) is violated. Some functions, namely those that satisfy $x(0) = x(T)$ and $Dx(0) = Dx(T)$, have infinitely many decompositions as $z + e$ with $Lz = Be = 0$, and are called the eigenfunctions of the differential operator.

18.6.3 The inner product defined by operators L and B

All the functional data analysis techniques and tools in this book depend on the notion of an inner product between two functions x and y. We have seen numerous examples of how a careful choice of inner product can produce more useful results, especially in controlling the roughness of estimated functions, such as functional principal components or regression functions. In these and other examples, it is important to use derivative information in defining an inner product.

Let us assume that the constraint operator is such that the orthogonality condition (18.31) is satisfied. We can define a large family of inner products as follows:

$$\langle x, y\rangle_{B,L} = (Bx)'(By) + \int (Lx)(t)(Ly)(t)\,dt \tag{18.32}$$

with the corresponding norm

$$\|x\|^2_{B,L} = (Bx)'(Bx) + \int (Lx)^2(t)\,dt. \tag{18.33}$$

The condition (18.31) ensures that this is a norm; the only function x for which $\|x\|_{B,L} = 0$ is zero itself, since this is the only function simultaneously satisfying $Bx = 0$ and $Lx = 0$.

In fact, this inner product works by splitting the function x into two parts:

$$x = z + e \text{ where } z \in \ker L \text{ and } e \in \ker B.$$

The first term in (18.33) simply measures the size of the component z, since $Be = 0$ and therefore $Bx = Bz$, while the second term depends only on the size of the component e since $Lx = Le$. The first term in (18.32) is essentially an inner product for the m-dimensional subspace in which z lives and which is defined by $Lz = 0$. The second term is an inner product for the function space of *codimension* m defined by $Be = 0$. Thus, we can write

$$\|x\|^2_{B,L} = \|z\|^2_B + \|e\|^2_L.$$

With this composite inner product in hand, that is, with a particular operator L and constraint operator B in mind, we can go back and revisit each of our functional data analytic techniques to see how they perform with this inner product. This is the central point explored by Ramsay and Dalzell (1991), to which we refer the reader for further discussion.

18.7 Further reading and notes

It is beyond the scope of this book to offer more than a cursory treatment of a topic as rich as the theory of differential equations, and there would be little point, since there are many fine texts on the topic. Texts on differential equations that are designed for engineering students tend to have two advantages. The amount mathematical detail is kept minimal and one gets to see differential equations applied to real world problems and is thereby helped to see them as conceptual as opposed to technical tools.

Some of our favorites references that are also classics are Coddington (1989), Coddington and Levinson (1955), Ince (1956) and Tenenbaum and Pollard (1963). For advice on a wide range of computational and otherwise practical matters we recommend Press et al. (1992).

For more general results for arbitrary constraint operators B, including the integral operator conditions that we need in the following section, see Dalzell and Ramsay (1993) and Heckman and Ramsay (2000).