20

Green's functions and reproducing kernels

20.1 Introduction

We now introduce two concepts that are useful for both computation and theory. Green's functions are important because they permit the solution for a nonhomogeneous linear differential equation $Lx = u$ to be explicitly represented and calculated, no matter what the forcing function u. Well, this is a slight overstatement, since what we mean is that the explicit solution is available provided that we know the solution to the corresponding homogeneous equation $Lx = 0$. But it is often the case that we do, and even if we only have available an approximation to the homogeneous solution, it can still be the case that we want to compute solutions for a wide range of forcing functions. Green's functions can, therefore, make a real difference in applications.

A reproducing kernel is a somewhat more theoretical concept, but many texts, such as Gu (2002) and Wahba (1990) use the notion freely, and one often encounters the term *reproducing kernel Hilbert space* in the literature using spline functions. In fact, the term has a standard abbreviation, namely *RKHS*. So it can be useful to know what it means. We try in this chapter to demystify reproducing kernels by showing their relationship to Green's functions.

In Chapter 21 we will use both Green's functions and reproducing kernels to develop new designer bases associated with any specific choice of linear differential operator L. These bases will, like B-splines, be nonzero only over a small number of adjacent intervals. That is, they have band-structured coefficient matrices, and permit smoothing in order n operations.

20.2 The Green's function for solving a linear differential equation

It often happens in engineering and science that the researcher has a homogeneous linear differential equation corresponding to a linear differential operator L that has either been worked out using fundamental principles or determined empirically. This equation describes the internal or endogenous dynamics of some system. What he or she wants to know, however, is that the consequences will be of adding an external or exogenous influence u to the system, and there are a wide variety of these potentially available. For example, a rocket may be a well-understood system as long as it is on the launching pad, but what will happen when it is in flight under the influence of atmospheric turbulence?

That is, what we want to know is the solution of the nonhomogeneous equation

$$Lx = u \tag{20.1}$$

for known L but arbitrary u? In effect, we want to reverse the effect of applying operator L because we have a *forcing function* u and we want to find x.

Of course, we recognize that the solution is not unique; if we add to any solution x some linear combination of the functions $\xi_j \in \ker L$ that span the null space, $\ker L$, of L, then this function also satisfies the equation.

But let us assume that the investigator has a set of known conditions defining a constraint operator B, and that these satisfy the complementarity condition (18.31). Typically, these will be initial value conditions describing, for example, the status of the rocket on the launching pad. Let m be the order of the equation. Then these constraints will be in the form

$$Bx = \mathbf{b} \tag{20.2}$$

for some known fixed m-vector $\mathbf{b}$.

Define the matrix $\mathbf{A}$ as the result of applying constraint operator B to each of the ξ_j's in turn:

$$\mathbf{A} = B\boldsymbol{\xi}', \tag{20.3}$$

so that the element in row i and column j of $\mathbf{A}$ is the ith element of vector $B\xi_j$. Since every ξ in $\ker L$ can be written as

$$\xi(t) = \sum_j c_j \xi_j(t) = \boldsymbol{\xi}'\mathbf{c}$$

for an m-vector of coefficients $\mathbf{c}$, then by the definition of $\mathbf{A}$ we have that

$$B\xi = \mathbf{b} = \mathbf{Ac}.$$

The conditions we have specified ensure that $\mathbf{A}$ is invertible, and consequently we have that

$$\mathbf{c} = \mathbf{A}^{-1}\mathbf{b}.$$

Now suppose that ν satisfies $L\nu = u$ and also $B\nu = 0$. That is, $\nu \in \ker B$, and in this sense is the complement of $\xi \in \ker L$. Then

$$x(t) = \xi(t) + \nu(t)$$

satisfies

$$Lx = u \quad \text{subject to} \quad Bx = \mathbf{b}.$$

Consequently, if we can solve the problem

$$L\nu = u \quad \text{subject to} \quad \nu \in \ker B, \tag{20.4}$$

we can find a solution subject to the more general constraint $Bx = \mathbf{b}$.

20.2.1 The definition of the Green's function

It can be shown that there exists a bivariate function $G(t; s)$ called the *Green's function*, associated with the pair of operators (B, L) that satisfies

$$\nu(t) = \int G(t; s) Lx(s)\, ds \quad \text{for} \quad \nu \in \ker B. \tag{20.5}$$

Thus, for $L\nu = u$, the Green's function defines an integral transform

$$\mathcal{G}u = \int G(t; s) u(s)\, ds \tag{20.6}$$

that inverts the linear differential operator L. That is, $\mathcal{G}L\nu = \nu$, given that $B\nu = 0$.

Before giving a general recipe for computing the Green's function G, let's look at a few specific examples. The first is nearly trivial: If our interval is $[0, T]$ and our constraint operator is the initial value constraint $B_0 x = x(0)$, then for $L = D$,

$$G(t; s) = 1, s \le t, \text{ and } 0 \text{ otherwise.}$$

That is, for ν such that $\nu(0) = 0$,

$$\nu(t) = \int_0^t D\nu(s)\, ds = \int_0^t u(s)\, ds.$$

Now consider the first order constant coefficient equation (18.1). Looking at the solution (18.3) for $\alpha(t) = 1$, we see by inspection that

$$G(t; s) = e^{-\beta(t-s)}, s \le t, \text{ and } 0 \text{ otherwise.}$$

Progressing from this situation to the variable coefficient version (18.5) is now easy:

$$G(t:s) = \xi(t)/\xi(s), s \leq t, \text{ and } 0 \text{ otherwise.}$$

20.2.2 A matrix analogue of the Green's function

Readers of this book may be familiar enough with matrix algebra to welcome a closely related concept in that domain. Suppose that we have, for $n > m$ an $n-m$ by n matrix $\mathbf{L}$ of rank $n-m$. If n is very large, then we approach the functional situation where $n \to \infty$.

Then there exists a subspace of n-vectors $\boldsymbol{\xi} \in \ker \mathbf{L}$ such that

$$\mathbf{L}\boldsymbol{\xi} = 0,$$

and that space is of dimension m. This means that we can construct a n by m matrix $\mathbf{Z}$ whose columns span this subspace such that $\mathbf{LZ} = 0$.

Also, we can always find an m by n matrix $\mathbf{B}$ of rank m such that there exists a space of dimension m of n vectors $\boldsymbol{\nu}$ such that

$$\mathbf{B}\boldsymbol{\nu} = 0 \ ;$$

and, moreover, such that the only vector $\mathbf{x}$ satisfying simultaneously $\mathbf{Lx} = 0$ and $\mathbf{Bx} = 0$ is $\mathbf{x} = 0$. For example, one way to compute such a matrix $\mathbf{B}$ is through the singular value decomposition of $\mathbf{L}$, but there are many other ways in which to define $\mathbf{B}$, which is not uniquely defined, just as the defining conditions and operator B for differential equations are not unique.

Corresponding to a particular choice of $\mathbf{B}$, we can find an n by $n-m$ matrix $\mathbf{N}$ such that $\mathbf{BN} = 0$.

Now suppose that we have an arbitrary n-vector $\mathbf{u}$. Then it follows that

$$\boldsymbol{\nu} = \mathbf{N}(\mathbf{LN})^{-1}\mathbf{u} \tag{20.7}$$

solves the equation

$$\mathbf{L}\boldsymbol{\nu} = \mathbf{u}$$

and, moreover, $\boldsymbol{\nu} \in \ker \mathbf{B}$ since $\mathbf{BN} = 0$. Matrix

$$\mathbf{G} = \mathbf{N}(\mathbf{LN})^{-1} \tag{20.8}$$

is the analogue of the Green's function $G(s;t)$.

A special choice of $\mathbf{B}$ leads to an interesting result. Let $\mathbf{B}$ be chosen so that $\mathbf{N} = \mathbf{L}'$. In that case, $\mathbf{G} = \mathbf{L}'(\mathbf{LL}')^{-1}$ and $\mathbf{G}'$ is the pseudo-inverse of $\mathbf{L}$.

20.2.3 A recipe for the Green's function

We can now offer a recipe for constructing the Green's function for any linear differential operator L of the form (18.12) and the initial value constraint B_I of the corresponding order. First, compute the Wronskian matrix

$\mathbf{W}(t)$ defined in (18.26). Secondly, define the functions

$$\mathbf{v}(t) = (v_1(t), \ldots, v_m(t))'$$

to be the vector containing the elements of the last row of $\mathbf{W}^{-1}$. Then, it turns out that initial value constraint Green's function $G_0(t; s)$ is

$$G_0(t; s) = \sum_{j=1}^{m} \xi_j(t) v_j(s) = \boldsymbol{\xi}(t)'\mathbf{v}(s), s \le t, \text{ and 0 otherwise.} \tag{20.9}$$

Let's see how this works for

$$L = \beta D + D^2.$$

The space $\ker L$ is spanned by the two functions $\xi_1(t) = 1$ and $\xi_2(t) = \exp(-\beta t)$. The Wronskian matrix is

$$\mathbf{W}(t) = \begin{bmatrix} \xi_1(t) & D\xi_1(t) \\ \xi_2(t) & D\xi_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \exp(-\beta t) & -\beta \exp(-\beta t) \end{bmatrix}$$

and consequently

$$\mathbf{W}^{-1}(t) = \begin{bmatrix} 1 & 0 \\ \beta^{-1} & -\beta^{-1} \exp(\beta t) \end{bmatrix},$$

from which we have

$$\mathbf{v}(s) = -\beta^{-1}[-1, \exp(\beta s)]'$$

and finally

$$G_0(t; s) = -\beta^{-1}[e^{-\beta(t-s)} - 1], s \le t, \text{ and 0 otherwise.} \tag{20.10}$$

We can verify that this is the required Green's function by integration by parts.

We do not discuss in any detail the case of any constraint functions B other than initial value constraints. Under boundary or periodic constraints, it may be that additional conditions are required on the function f or on the constraint values c, but nevertheless we can extend the basic ideas of Green's functions.

20.3 Reproducing kernels and Green's functions

A bivariate function called the *reproducing kernel* plays a central role in the theory of spline functions, and we will use reproducing kernels in Chapter 21 to define a basis function system $\boldsymbol{\phi}$ specific to any linear differential operator L used to define a roughness penalty.

20.3.1 What is a reproducing kernel?

We remarked in Section 18.6.3 that the concept of an *inner product* underlies perhaps about 95% of all applied mathematics and statistics. There is no possibility here of doing more than recalling the most basic elements of inner product spaces, and perhaps no particular need, either.

A *Hilbert space* is a collection of objects x for which there exists:

- linear combinations $ax_1 + bx_2$,
- an inner product $\langle x_1, x_2 \rangle$ for any pair x_1 and x_2
- a property called *completeness*, namely that convergent sequences of elements converge to elements within the space.

Both vectors and functions as used in applied work are typically elements of Hilbert spaces, and Section 18.6.3 gave some functional examples of useful inner products.

There is a sense, however, in which the Hilbert space is too loose a concept. This revolves around the linear mapping

$$\rho_t(x) = x(t),$$

which we called the *evaluation mapping* in Section 5.5. If a function x is smooth, we imagine that knowing $x(t)$ tells us a great deal about $x(t+\delta)$ when perturbation δ is sufficiently small. Unfortunately, such need not be the case for Hilbert spaces in general.

Consequently, we need to focus on the more specialized Hilbert space for which the evaluation map is *continuous*. It would be nice to imagine that these would be called something like smooth Hilbert spaces, or continuous Hilbert spaces, but alas, mathematics does not tend to generate its nomenclature in such a kindly way! Instead, spaces of this nature are called *reproducing kernel Hilbert spaces*, not surprisingly often abbreviated to *RKHS*s.

It is a basic theorem in functional analysis, called the *Riesz representation theorem*, that if a linear mapping $\rho(x)$ in a Hilbert space is continuous, then there exists a function k in the space such that

$$\rho(x) = \langle x, k \rangle \ .$$

Consequently, applying this idea to the evaluation map $\rho_t(x)$, there must exist a *bivariate* function $k(s,t)$ such that $k(\cdot, t)$ is in the space for any t, and that

$$\rho_t(x) = \langle x, k(\cdot, t) \rangle \ .$$

The term *reproducing kernel* comes from the consequence that

$$k(s,t) = \langle k(\cdot, s), k(\cdot, t) \rangle \ .$$

The existence of $k(s,t)$ has many wide-ranging consequences, and plays an especially important role in the history of the development of spline smoothing.

So, given that we have a Hilbert space with a continuous evaluation map, how do we find the reproducing kernel? The surprising result is: If you know the Green's function for the linear differential operator that defines inner products of the type described in Section 18.6.3, you are almost there!

There are two reproducing kernels $k(s,t)$ that we need to consider, one for each of the function subspaces $\ker B$ and $\ker L$. We now show how these can be calculated, and we will put them to work in Chapter 21.

20.3.2 The reproducing kernel for $\ker B$

The reproducing kernel for the $\ker B$ subspace, consisting of functions that satisfy $Bx = 0$, has a simple relationship to the Green's function G. First, however, we need to explain what a reproducing kernel is in this context.

Given any two functions x and y in $\ker B$, let us define the L-inner product

$$\langle x, y\rangle_L = \langle Lx, Ly\rangle = \int Lx(s)Ly(s)\,ds.$$

Let G_I be the Green's function as defined in Section 20.2.3, and define a function $k_2(t,s)$ such that, for all t,

$$Lk_2(t,\cdot) = G_I(t;\cdot) \text{ and } Bk_2(t,\cdot) = 0. \tag{20.11}$$

By the defining properties of Green's functions, this means that

$$k_2(t,s) = \int G_I(s;w)G_I(t;w)\,dw. \tag{20.12}$$

The function k_2 has an interesting property. Suppose that ν is any function in $\ker B$, and consider the L-inner product of $k_2(t,\cdot)$ and ν. We have, for all t,

$$\langle k_2(t,\cdot),\nu\rangle_L = \int Lk_2(t,s)L\nu(s)\,ds = \int G_I(t;s)L\nu(s)\,ds = \nu(t) \tag{20.13}$$

by the key property (20.5) of Green's functions. Thus, in the space $\ker B$ equipped with the L-inner product, taking the L-inner product of k_2 using its second argument with any function ν yields the value of ν at its first argument. Overall, taking the inner product with k_2 reproduces the function ν, and k_2 is called the reproducing kernel for this function space and inner product.

Chapter 21 shows that the reproducing kernel is the key to the important question, "Is there an optimal set of basis functions for smoothing data?" To answer this question, we need to use the important property

$$\langle k_2(s,\cdot), k_2(t,\cdot)\rangle_L = k_2(s,t), \tag{20.14}$$

which follows at once from (20.13) setting $\nu(\cdot) = k_2(s, \cdot)$ and appealing to the symmetry of the inner product.

We can put the expression (20.12) in a slightly more convenient form for the purpose of calculation. Recalling the definitions of the vector-valued functions $\boldsymbol{\xi}$ and $\mathbf{v}$ in Section 20.2.3, we have from (20.9), assuming that $s \leq t$, that

$$k_2(s,t) = \int_0^s [u(s)'v(w)][v(w)'u(t)]\,dw = u(s)'\mathbf{F}(s)u(t), \tag{20.15}$$

where the order m symmetric matrix-valued function $\mathbf{F}(s)$ is

$$\mathbf{F}(s) = \int_0^s v(w)v(w)'\,dw. \tag{20.16}$$

To deal with the case $s > t$, we use the property that $k_2(s,t) = k_2(t,s)$.

The matrix analogue of the reproducing kernel $k_2(s,t)$ is

$$\mathbf{K}_2 = \mathbf{G}\mathbf{G}',$$

since we see that for any $\nu \in \ker B$

$$\begin{aligned} \mathbf{K}_2 L' L \nu &= \mathbf{N}(\mathbf{L}\mathbf{N})^{-1}(\mathbf{N}'\mathbf{L}')^{-1}\mathbf{N}'\mathbf{L}'\mathbf{L}\nu \\ &= \mathbf{N}(\mathbf{L}\mathbf{N})^{-1}\mathbf{L}\nu \\ &= \nu \end{aligned} \tag{20.17}$$

as required.

20.3.3 *The reproducing kernel for* ker L

Suppose now that $f = \sum a_i \xi_i$ and $g = \sum b_i \xi_i$ are elements of $\ker L$. We can consider the B-inner product on the finite-dimensional space $\ker L$, defined by

$$\langle f, g \rangle_B = (Bf)'Bg = a'\mathbf{A}'\mathbf{A}b.$$

Define a function $k_1(t,s)$ by

$$k_1(t,s) = \boldsymbol{\xi}(t)'(\mathbf{A}'\mathbf{A})^{-1}u(s).$$

It is now easy to verify that, for any $f = \sum_i a_i \xi_i$,

$$\langle k_1(t,\cdot), f \rangle_B = \boldsymbol{\xi}(t)'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}a = \boldsymbol{\xi}(t)'a = a'\boldsymbol{\xi}(t) = f(t).$$

So k_1 is the reproducing kernel for the space $\ker L$ equipped with the B-inner product.

Finally, we consider the space of more general functions x equipped with the inner product $\langle \cdot, \cdot \rangle_{B,L}$ as defined in Section 18.6.3. It is easy to check from the properties we have set out that the reproducing kernel in this space is given by

$$k(s,t) = k_1(s,t) + k_2(s,t).$$

20.4 Further reading and notes

Although the theory of reproducing kernel Hilbert spaces is considered to be of relatively recent origin, and usually attributable to Aronszajn (1950), it is in fact grounded in the theory of Green's functions, a topic older by more than a century (Green, 1828). The concept of a reproducing kernel appears in most of the papers by G. Wahba, including Wahba (1990). More recently, reproducing kernels are used extensively in Gu (2002). The interested and highly motivated reader might want to consult a reference on functional analysis with an applied orientation, such as Aubin (2000).