A local spectral method: the finite element method for BVPs

We spent last week learning about spectral methods for boundary value problems (BVPs). By contrast with finite difference methods, we saw that these spectral methods are derived by seeking a solution that minimizes the least squares error from the true solution. We then saw that we could compute this optimal solution using an energy inner product. We also considered some examples involving *global* spectral methods, which involved basis functions that were defined globally over the domain $[a, b]$.

If some of this is sounding hazy, now is a good opportunity to spend 5-10 minutes reviewing the Week 11 notes!

In this lecture, we will continue to consider spectral methods, but will focus on a very important sub-class of spectral method referred to as the *finite element method*. Like any spectral method, the finite element method finds the best least-squares solution onto some space $\mathcal{V}$ that we are approximating onto. The difference from the global spectral methods we explored last time is that the finite element method is defined in terms of *locally* defined basis functions. In this week's typed notes, we will first motivate why we are interested in using locally defined basis functions, and will then derive a specific finite element method.

1 *Motivation: why would we want locally defined basis functions?*

In this section, we will motivate the choice of locally defined basis functions. To do this, let us recall that since we are using a spectral method, the equation we need to solve to compute our numerical solution is

$$\begin{bmatrix} (\phi_{i_1},\phi_{i_1})_E & (\phi_{i_2},\phi_{i_1})_E & \cdots & (\phi_{i_{n-1}},\phi_{i_1})_E & (\phi_{i_n},\phi_{i_1})_E \\ (\phi_{i_1},\phi_{i_2})_E & (\phi_{i_2},\phi_{i_2})_E & \cdots & (\phi_{i_{n-1}},\phi_{i_2})_E & (\phi_{i_n},\phi_{i_2})_E \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ (\phi_{i_1},\phi_{i_{n-1}})_E & (\phi_{i_{n-1}},\phi_{i_1})_E & \cdots & (\phi_{i_{n-1}},\phi_{i_{n-1}})_E & (\phi_{i_n},\phi_{i_{n-1}})_E \\ (\phi_{i_1},\phi_{i_n})_E & (\phi_{i_n},\phi_{i_n})_E & \cdots & (\phi_{i_{n-1}},\phi_{i_n})_E & (\phi_{i_n},\phi_{i_n})_E \end{bmatrix} \begin{bmatrix} c_{i_1} \\ c_{i_2} \\ \vdots \\ c_{i_{n-1}} \\ c_{i_n} \end{bmatrix} = \begin{bmatrix} -(f,\phi_{i_1})_s \\ -(f,\phi_{i_2})_s \\ \vdots \\ -(f,\phi_{i_{n-1}})_s \\ -(f,\phi_{i_n})_s \end{bmatrix} \tag{1}$$

which we may for succinctness write as $\boldsymbol{G}\boldsymbol{c} = \boldsymbol{b}$. Note that for most basis choices, the matrix $\boldsymbol{G}$ will have many nonzero entries and it will therefore be very costly to solve the system (1) for $\boldsymbol{c}$. There are important exceptions to this rule of thumb. For example, we saw in last week's lecture that the basis of sine functions for $\mathcal{V} = \mathcal{T}_0^n$ led to a diagonal $\boldsymbol{G}$, which leads to the easiest possible linear system to solve. For the most part, however, a given basis (including the Lagrange polynomials we considered when working with $\mathcal{P}_0^n$) will give a $\boldsymbol{G}$ that is quite cumbersome to work with.

Remember that this equation came from i) selecting a space $\mathcal{V}$ on which to approximate our numerical solution, ii) writing the numerical solution as a linear combination of basis functions for $\mathcal{V}$, and iii) using the energy inner product and the orthogonality relation enjoyed by the best least-squares solution to arrive at equations for the unknown coefficients $c_{i_1}, \ldots, c_{i_n}$.

The crux of the finite element method is to get around this challenge by observing that the energy inner product involves integrals over the interval $[a, b]$. Thus, if we can choose a set of basis functions that are only nonzero for a small portion of the interval $[a, b]$, we can produce a linear system $\boldsymbol{G}$ that is dominated by zeros, and thus much easier to work with.

A figure is helpful to illustrate this concept. We will define a set of "hat" basis functions later in this lecture that are depicted in figure 1. Notice that because they are only nonzero over a small sub-interval of $[a, b]$, many of the inner product terms in $\boldsymbol{G}$ would vanish. For example, the only inner products that are nonzero involving ϕ_2 are $(\phi_2, \phi_2)_E$, $(\phi_2, \phi_3)_E$, and $(\phi_3, \phi_2)_E$.

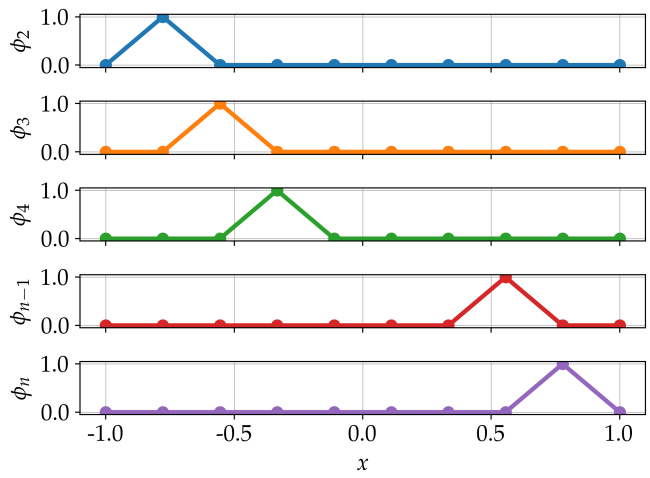

Figure 1: Five locally defined "hat" basis functions.

This philosophy driving finite element methods is sufficiently important that it gets its own orange box:

Philosophy behind the finite element method

The finite element method is a spectral method that uses locally defined functions to create a matrix $\boldsymbol{G}$ with predominately zero entries to facilitate a fast solution of $\boldsymbol{Gc} = \boldsymbol{b}$.

This goal of creating locally defined basis functions within this spectral method framework will be our focus for today. To facilitate this goal, we must ask ourselves some questions: what is a suitable space $\mathcal{V}$ that allows as a basis a set of locally defined functions? How do we define these basis functions mathematically? Finally, how do we use special $\mathcal{V}$ and set of local basis functions to solve the 1D Poisson problem? We will answer each of these questions in turn in this week's typed notes.

2 *Choosing the space $\mathcal{V}$ and the basis functions*

Figure 1 presents a tantalizing set of candidate basis functions for us to use: piecewise linear functions. But there are some complicating

issues that make it unclear whether this choice is appropriate. First, if we define piecewise functions, how many pieces do we use: 5? 20? 457? Can we systematically describe the number of pieces? Second, what is the appropriate space that defines these piecewise functions? Third, do these functions form a basis for the space comprised of piecewise linear functions?

Regarding the first issue—how to systematically define the number of pieces we are using to characterize our piecewise functions—we will break the interval $[a, b]$ up into n subintervals by defining a grid as $x_1, x_2, \ldots, x_n, x_{n+1}$ (each point x_j, $j = 1, \ldots, n+1$, is called a *node* in the finite element method). We define the number of pieces of the piecewise function in terms of these subintervals: a function $g(x)$ is called piecewise linear if

$$g(x) = a_i x + d_i \quad x \in [x_{i-1}, x_i], \quad i = 2, \ldots, n+1 \tag{2}$$

Said differently, $g(x)$ is piecewise linear if it can be written as a set of lines over each subinterval.

where a_i and d_i are constants defined over each subinterval.

Notice that the number of intervals n now becomes a parameter that we can tune. This is crucial, as it gives us a means to define convergence of finite element methods. Indeed, a natural question is to ask: will our finite element solution converge to the exact solution as n is increased?

2.1 *The appropriate space*

With respect to the second issue—the appropriate space—that is now clear from our definition of piecewise linear functions enabled by addressing the second issue. Let us define by $\mathcal{V}_n^L$ the space of n-piecewise linear functions *that also satisfy the requisite boundary conditions*. That is,

$$\begin{aligned}\mathcal{V}_n^L = \{g(x) : g(a) = g(b) = 0 \text{ and} \\ g(x) = a_i x + d_i, \text{ for } a_i, d_i \in \mathbb{R},\ x \in [x_{i-1}, x_i],\ i = 2, \ldots, n+1\}\end{aligned} \tag{3}$$

Do not forget that when confronted with intimidating mathematical notation, the key is to break it down into manageable words. Equation (3) says that $\mathcal{V}_n^L$ is defined by the set of functions g such that $g(x)$ satisfies the zero Dirichlet boundary conditions and can be written as a line over each sub-interval $x \in [x_{i-1}, x_i]$.

2.2 *A basis for* $\mathcal{V}_n^L$

At last, we turn our attention to the final concern we raised: what is a good basis for $\mathcal{V}_n^L$? We will show that the functions plotted in figure 1 provide an answer to this question. To see this fact, let $q(x)$ be *any* piecewise linear function; *i.e.*, choose any $q(x) \in \mathcal{V}_n^L$. Then $q(x)$ is defined as a set of lines over $[x_1, x_2], [x_2, x_3], \ldots, [x_n, x_{n+1}]$ and has a set of nodal values $q(x_1), q(x_2), \ldots, q(x_{n+1})$.

For convenience, let us say that we want to find a set of basis functions $\{\phi_2, \ldots, \phi_n\}$ such that

Notice that since $q(x) \in \mathcal{V}_n^L$ and $\mathcal{V}_n^L$ satisfies the Dirichlet boundary conditions, $q(x_1) = q(x_{n+1}) = 0$. The fact that these nodal values are automatically zero is why they are not included in the sum 4.

$$q(x) = \sum_{j=2}^{n} q(x_j)\phi_j(x) \tag{4}$$

That is, we are looking for a basis that lets the coefficients in the expansion of a function be the nodal values of the function itself.

Notice that if we can pick a set of basis functions ϕ_j ($j = 2, \ldots, n$) that are piecewise linear and satisfy

$$\phi_j(x_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{5}$$

then we will have found our sought-after basis functions that enable the expression (4) to be used.

The property (5) does indeed facilitate the use of (4). Check this fact for yourself: notice that

$$q(x_i) = \sum_{j=2}^{n} q(x_j)\phi_j(x_i) \tag{6}$$

and thus all terms in the sum vanish except for the i^{th} term, which becomes $q(x_i)\phi_i(x_i) = q(x_i)$. Thus $q(x)$ will by construction have the correct nodal values.

Moreover, the values of $q(x)$ between nodes will also be correct, as two points uniquely define a line and the values $q(x_{i-1})$ and $q(x_i)$ are therefore all that is required to define the line over each interval $[x_{i-1}, x_i]$.

Now, what are these special basis functions that satisfy property (5)? One can verify that the expression for these basis functions is given by

$$\phi_i(x) = \begin{cases} \frac{1}{\Delta x}[x - a - (i-1)\Delta x] & x \in [x_{i-1}, x_i] \\ -\frac{1}{\Delta x}[x - a - (i+1)\Delta x] & x \in [x_i, x_{i+1}] \\ 0 & else \end{cases} \tag{7}$$

for $i = 2, \ldots, n$. Note that (7) assumes the grid spacing $\Delta x = x_j - x_{j-1}$ is constant (modifications to nonuniform grid spacings are tedious but conceptually straightforward). These basis functions have affectionately been dubbed "hat functions" by the finite element community. The reason for this moniker is evident from figure 1, which provides plots of these basis functions for $n = 9$.

Note that, while we have restricted ourselves to piecewise linear functions in this section, this is by no means the only option. Indeed, finite element methods can (and are!) constructed using other locally-defined functions (locally quadratic, cubic, *etc.*). So long as the method is a spectral method derived using local basis functions, it is a finite element method.

3 *Solving the 1D Poisson problem with our finite element method*

We now have all the requisite tools at our disposal to solve the 1D Poisson with the finite element method. Recall that despite the new name we use to describe our method, there is nothing new in our process compared with last week! We are simply using a spectral method to solve a BVP. The only difference is that we are approximating our solution on a space comprised of locally defined functions ($\mathcal{V}_n^L$) rather than globally defined functions (*e.g.*, $\mathcal{P}_0^n$). Because we are still using a spectral method, we may directly use the expression $\boldsymbol{Gc} = \boldsymbol{b}$ provided by (1).

Recall from last week that the 1D Poisson problem we are considering is

$$\frac{d^2u}{dx^2} = f, \quad x \in [a, b] \tag{8}$$

$$u(a) = 0,\ u(b) = 0 \tag{9}$$

That said, it is worth re-reminding ourselves that the use of the Poisson problem is for simplicity of presentation, but is not prescriptive. Straightforward changes can be incorporated to develop a finite element method for the more general BVP

$$\alpha\frac{d^2u}{dx^2} + \beta\frac{du}{dx} + \gamma u = f, \quad x \in [a, b] \tag{10}$$

There are some simplifications that arise to this expression by virtue of our choice of $\mathcal{V}_n^L$ and our basis selection for this space. First, note that the coefficients $c_j = u_j$ (where u_j is an approximation to $u(x_j)$), by virtue of the property (5) that our basis functions satisfy which lets us write the approximation to $u(x)$ in terms of its nodal values.

Second, the terms in $\boldsymbol{G}$ and $\boldsymbol{b}$ simplify considerably. Let us consider the entries of $\boldsymbol{G}$ in detail by evaluating the inner products. Note that the hat functions defined in (7) have derivatives defined piecewise as

$$\phi_i'(x) = \begin{cases} \frac{1}{\Delta x} & x \in [x_{i-1}, x_i] \\ -\frac{1}{\Delta x} & x \in [x_i, x_{i+1}] \\ 0 & else \end{cases} \tag{12}$$

for $i = 2, \ldots, n$. We will use this expression to evaluate the inner product in terms in $\boldsymbol{G}$. From the definition of $\phi'(x)$ given in (12), it is clear that the only inner products that will be nonzero are

$$(\phi_{i-1}, \phi_i)_E, \quad (\phi_i, \phi_i)_E, \quad (\phi_i, \phi_{i+1})_E \tag{13}$$

for $i = 3, \ldots, n-1$. We can compute these analytically:

$$\begin{aligned} (\phi_{i-1}, \phi_i)_E &= \int_{a+(i-1)\Delta x}^{a+i\Delta x} \left(\frac{-1}{\Delta x}\right)\left(\frac{1}{\Delta x}\right) dx = -\frac{1}{\Delta x} \\ (\phi_i, \phi_i)_E &= \int_{a+(i-1)\Delta x}^{a+(i+1)\Delta x} \left(\frac{1}{\Delta x}\right)\left(\frac{1}{\Delta x}\right) dx = \frac{2}{\Delta x} \\ (\phi_i, \phi_{i+1})_E &= \int_{a+i\Delta x}^{a+(i+1)\Delta x} \left(\frac{-1}{\Delta x}\right)\left(\frac{1}{\Delta x}\right) dx = -\frac{1}{\Delta x} \end{aligned} \tag{14}$$

Regarding $\boldsymbol{b}$, we have that

$$(f, \phi_i) = \int_a^b f(x)\phi_i(x)dx = \int_{a+(i-1)\Delta x}^{a+(i+1)\Delta x} f(x)\phi_i(x)dx \tag{15}$$

for $i = 2, \ldots, n$.

We can put all of these results together to get the form that the linear system (1) takes when applying the finite element method (with piecewise linear functions). The result is

$$\frac{1}{\Delta x}\begin{bmatrix} -2 & 1 & \cdots & 0 & 0 \\ 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & \cdots & 1 & -2 \end{bmatrix}\begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} = \begin{bmatrix} \int_a^{a+2\Delta x} f(x)\phi_2(x)dx \\ \int_{a+\Delta x}^{a+3\Delta x} f(x)\phi_3(x)dx \\ \vdots \\ \int_{a+(n-2)\Delta x}^{a+n\Delta x} f(x)\phi_{n-1}(x)dx \\ \int_{a+(n-1)\Delta x}^{a+(n+1)\Delta x} f(x)\phi_n(x)dx \end{bmatrix} \tag{16}$$

Solving this matrix system gives us the values of our approximate solution at the nodes $x_1, \ldots, x_n$.

To facilitate our exploration of the inner products in $\boldsymbol{G}$, remember that the energy inner product is defined as

$$(f, g)_E = \int_a^b f'(x)g'(x)dx, \quad \forall f, g \in \mathcal{V} \tag{11}$$