Lecture Notes for Math 425
Fall 2002

Alan Demlow

September 4, 2002

Chapter 1

Preliminaries on Differential Equations

1.1 A two-point boundary value problem

During this course, we shall study the numerical analysis of the *two-point boundary value problem* (or TPBVP)

$$\begin{aligned} -(a(x)u'(x))' &= f(x) \text{ for all } x \in (0,1), \\ u(0) &= u(1) = 0, \end{aligned} \tag{1.1}$$

where we assume that $A_0 \geq a(x) \geq a_0 > 0$. Actually, we will consider a number of variations on this problem, including problems with variable coefficients multiplying u'' and u, problems with different boundary conditions at $x = 0$ and $x = 1$, time-dependent problems such as the heat equation, and nonlinear versions of (1.1). For the present, however, we shall concentrate on (1.1) as we begin our investigation of numerical methods for approximating solutions to different types of differential equations.

The boundary-value problem (1.1) may (under some simplifying assumptions) be used to model the deflection of a string with material properties a under a load f. However, for our purposes, its main interest lies in its ability to serve as a *model problem* for types of partial differential equations which arise in many applications such as structural engineering (what are the stresses in an airplane wing?) or stationary heat conduction (what is the equilibrium distribution of heat in a body?). The actual differential equations arising in applications are often much more complicated than (1.1), but they share many features with this simple model problem. Our goal, then, is to understand in detail how numerical methods for approximating solutions to our simple two-point boundary value problem and some of its cousins work, with the hope being that the salient features of numerical methods for this problem will translate to more complicated problems. (One should also note that it is possible to solve (1.1) exactly by simply integrating, so it isn't really necessary to develop high-powered numerical techniques to approximate its solutions!)

1.2 Three forms of a TPBVP

Before beginning our discussion of numerical methods for approximating solutions to a two-point boundary value problem, we first need to discuss different ways of representing differential equations. We begin by noting that (1.1) is called the *strong form* of the differential equation. We also remark here that we shall assume, unless otherwise noted, that all functions in our arguments (u, a, f, etc.) are continuous and have enough continuous derivatives to write any quantities of interest. It is interesting and physically important to consider what happens when certain quantities of interest are not continuous or not smooth. However, technicalities become more burdensome when analyzing this case, so we shall generally assume we can differentiate or use the properties of continuous functions whenever we wish.

1.2.1 The Weak Form

Let v be a smooth function such that $v(0) = v(1) = 0$. We then multiply both sides of (1.1) by v, integrate by parts, and use the fact that $v(0) = v(1) = 0$ to find that

$$\begin{aligned} -\int_0^1 (a(x)u'(x))'v(x)\,dx &= -a(x)u'(x)v(x)|_0^1 + \int_0^1 a(x)u'(x)v'(x)\,dx \\ &= \int_0^1 a(x)u'(x)v'(x)\,dx \\ &= \int_0^1 f(x)v(x)\,dx\,. \end{aligned} \tag{1.2}$$

We denote by $\mathcal{A}_0$ the set of all functions which are smooth enough and which are zero at the boundary points 0 and 1. We also define the bilinear form $\mathcal{L}(\cdot,\cdot)$ by

$$\mathcal{L}(u,v) = \int_0^1 au'v'\,dx$$

and the inner product $(\cdot,\cdot)$ by

$$(f,v) = \int_0^1 fv\,dx\,.$$

Then we may rewrite (1.1) as: Find a function $u \in \mathcal{A}_0$ such that

$$\mathcal{L}(u,v) = (f,v) \text{ for all } v \in \mathcal{A}_{\prime}. \tag{1.3}$$

We shall call (1.3) the *weak form* of the differential equation (1.2). We have already shown that if u satisfies the strong form equation (1.1), then it satisfies the weak form equation above. But what about the converse? That is, if u solves the weak form equations, does it also satisfy the strong form equations? The answer is yes, it does.

Proposition 1.2.1 *u satisfies the weak form differential equation (1.3) if and only if u also satisfies the strong form differential equation (1.1).*

Proof. We have already shown that if u satisfies (1.1), then u also satisfies (1.3). To show the converse, we proceed with a proof by contradiction. Thus we assume that u satisfies the weak form (1.3) but does not satisfy the strong form (1.1). Then there is a point $x_0 \in (0,1)$ such that $-(a(x_0)u'(x_0))' \neq f(x_0)$, or, without loss of generality, $-(a(x_0)u'(x_0))' - f(x_0) > 0$. (The argument would proceed in precisely the same fashion if instead $-(a(x_0)u'(x_0))' - f(x_0) < 0$.) Since we

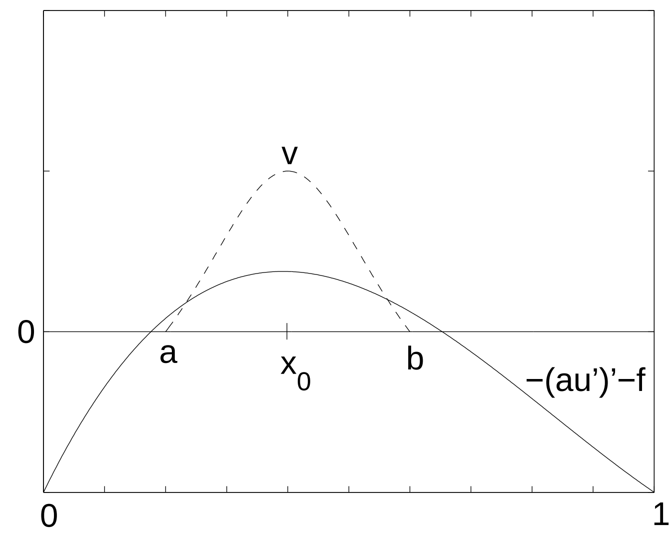

Figure 1.1: A "bump" function

have assumed all functions of interest are continuous, $-(au')' - f$ is continuous, and in particular $-(a(x)u'(x))' - f(x) > 0$ in some interval (a, b) containing x_0. We next let $v(x) \in \mathcal{A}_0$ be a "bump" function which is positive on (a, b) and 0 elsewhere, as pictured in Figure 1.1.

Thus $(-(a(x)u'(x))' - f(x))v(x)$ is strictly positive on (a, b) and 0 elsewhere. Thus

$$\int_0^1 (-(a(x)u'(x))' - f(x))v(x)\,dx > 0,$$

and integrating by parts gives us

$$\int_0^1 (a(x)u'(x)v'(x) - f(x)v(x))\,dx > 0.$$

(Recall that since $v \in \mathcal{A}_0$, the boundary terms in the integration by parts disappear.) In particular,

$$\int_0^1 au'v'\,dx \neq \int_0^1 fv\,dx\,.$$

But we have assumed that u satisfies (1.3), so this is a contradiction.

1.2.2 A Minimization Problem

A guiding principal in much of physics, chemistry, and engineering is the principal of minimization of energy. In general, physical systems seek a configuration which minimizes their potential energy. This principal leads us to our third form of the two-point boundary value problem (1.1). If we assume that (1.1) is modelling the deflection of a string as previously described, then we may define the "energy" of the string with deflection v as

$$\mathcal{E}(v) = \frac{1}{2}\int_0^1 a(v')^2\,dx - \int_0^1 fv\,dx\,. \tag{1.4}$$

We let $\mathcal{A}_0$ be the same admissible set of functions we have described previously, that is, functions v which satisfy the boundary conditions $v(0) = v(1) = 0$ and which are "smooth enough". Then our minimization problem is: Find a function $u \in \mathcal{A}_0$ such that

$$\mathcal{E}(u) = min_{v\in\mathcal{A}_0}\mathcal{E}(v). \tag{1.5}$$

As in the case of the weak form, we may ask whether the minimization problem (1.5) is equivalent to our original two-point boundary value problem (1.1). The answer is yes.

Proposition 1.2.2 *The strong form, weak form, and minimization form of the two-point boundary value problem are equivalent. That is, if u is a "smooth enough" function which solves any one of the problems, then u also solves the other two.*

Proof. We shall proceed by showing that the weak form and the minimization form are equivalent. Since we have already shown equivalence between the weak and strong forms, this will complete our proof.

We note at the outset that we shall be a little bit sloppy at a few points in our proof, that is, we shall give the flavor of the correct proof while neglecting some technical details. We begin by assuming that u satisfies the minimization problem (1.5). We then let $v \in \mathcal{A}_0$ and $t \in \mathbb{R}$ and notice that then $u + tv \in \mathcal{A}_0$. We then may compute that

$$\begin{aligned}\mathcal{E}(u+tv) &= \tfrac{1}{2}\textstyle\int_0^1 a(u'+tv')^2\,dx - \int_0^1 f(u+tv)\,dx\\ &= \tfrac{1}{2}\textstyle\int_0^1 au'^2\,dx + t\int_0^1 au'v'\,dx + \frac{t^2}{2}\int_0^1 a(v')^2\,dx\\ &\quad - \textstyle\int_0^1 fu\,dx - t\int_0^1 fv\,dx\,.\end{aligned} \tag{1.6}$$

Since u and v are fixed, we may think of $\mathcal{E}(u + tv)$ as a function of t–in fact, it's a differentiable function of t. But we have assumed that $\mathcal{E}$ is minimal at u, that is, $\mathcal{E}(u) \le \mathcal{E}(u + tv)$. Thus by Fermat's theorem, $\frac{d}{dt}\mathcal{E}(u+tv)|_{t=0} = 0$. Using our expression for $\mathcal{E}(u + tv)$ from (1.6), we find the *Euler-Lagrange equations* for the problem (1.5):

$$\begin{aligned}\tfrac{d}{dt}\mathcal{E}(u+tv)|_{t=0} &= (\textstyle\int_0^1 au'v'\,dx + t\int_0^1 a(v')^2\,dx - \int_0^1 fv\,dx)|_{t=0}\\ &= \textstyle\int_0^1 au'v'\,dx - \int_0^1 fv\,dx = 0,\end{aligned}$$

or rewriting,

$$\int_0^1 au'v'\,dx = \int_0^1 fv\,dx$$

for any $v \in \mathcal{A}_0$. This is precisely the weak form (1.3), so we have proven that if u minimizes $\mathcal{E}$, then it also solves the weak form.

We now assume that u solves the weak form, and show that it also minimizes $\mathcal{E}$. The proof may appear to be unmotivated. We will seemingly pull an expression out of thin air, then manipulate it using the weak form to reach the statement we wish to prove. The statement we pull out of thin air is:

$$-\frac{1}{2}\int_0^1 a(u'-v')^2\,dx \le 0,$$

which is true because $a > 0$ and $(u'-v')^2 \ge 0$, so the integral of the product of these two things must be nonnegative as well. Expanding this expression, we find that

$$-\frac{1}{2}\int_0^1 a(u')^2\,dx + \int_0^1 au'v'\,dx - \frac{1}{2}\int_0^1 a(v')^2\,dx \le 0. \tag{1.7}$$

Manipulating (1.7) and using the weak form (1.3), we find

$$\begin{aligned} -\tfrac{1}{2}\textstyle\int_0^1 a(u')^2\,dx &\le -\textstyle\int_0^1 au'v'\,dx + \tfrac{1}{2}\textstyle\int_0^1 a(v')^2\,dx \\ &= -\textstyle\int_0^1 fv\,dx + \tfrac{1}{2}\textstyle\int_0^1 a(v')^2\,dx = \mathcal{E}(v). \end{aligned}$$

Again using the weak form (1.3), we find that

$$\begin{aligned} -\tfrac{1}{2}\textstyle\int_0^1 a(u')^2\,dx &= \tfrac{1}{2}\textstyle\int_0^1 a(u')^2\,dx - \textstyle\int_0^1 a(u')^2\,dx \\ &= \tfrac{1}{2}\textstyle\int_0^1 a(u')^2\,dx - \textstyle\int_0^1 fv\,dx = \mathcal{E}(u). \end{aligned}$$

Combining the last two expressions above yields

$$\mathcal{E}(u) \le \mathcal{E}(v),$$

which is what we were seeking to prove.

We finally note that that we haven't shown that any solutions to our differential equations exist, or that if they do exist that they are unique. It is not that difficult to show in this simple case that exactly one solution to our differential equation does indeed exist; one may simply integrate (1.1). However, this is considered to be "cheating" because it doesn't work for most partial differential equations, so we shall leave this as an exercise to the reader.

1.3 Essential and natural boundary conditions

In writing down the differential equation (1.1), we specified that the value of u at the boundary points 0 and 1 of the interval $(0,1)$ be 0. In the weak and minimization forms of the two-point boundary value problem, we likewise specified explicitly that $u(0) = u(1) = 0$ by requiring that u come from the class $\mathcal{A}_0$ of admissible functions with $u(0) = u(1) = 0$. There are, however, many other possible boundary conditions we could specify, and they are not all treated the same, especially in the weak form. We shall not discuss here the treatment of boundary conditions in the minimization form of the problem.

In the strong form problem, there are three main types of boundary conditions we may impose: Dirichlet, Neumann, and Robin. When speaking about boundary conditions in the context of the weak form, we also make a distinction between *essential* and *natural* boundary conditions, which we define below. Letting x_{BP} be a boundary point (in our case, 0 or 1), we may summarize these types of boundary conditions in Table 1.1.

If we are modelling the deflection u of a string under a load, then Dirichlet conditions correspond to pinning down the ends of the string, Neumann conditions correspond to applying a fixed tension

Table 1.1:

Proper Name	Boundary Condition	Weak Type
Dirichlet	$u(x_{BP}) = u_{BP}$	essential
Neumann	$u'(x_{BP}) = \tilde{\nu}_{BP}$	natural
Robin	$u'(x_{BP}) + \tau_{BP} u(x_{BP}) = g_{BP}$	natural

at the ends, and Robin conditions correspond to requiring a sort of equilibrium between the tension and position at the end of the string. If we are modelling the temperature distribution in a body at thermal equilibrium (say, in a very thin rod in our present case), then Dirichlet conditions correspond to fixing the temperature at the ends of the body, Neumann conditions correspond to requiring that a certain amount of heat escapes or is pumped into the body, and Robin conditions correspond to a combination of the two. We note that we may require different boundary conditions at each end of the rod, e.g., we may require that $u(0) = 2$ and $u'(1) = -1$.

Notice that Dirichlet boundary conditions must be imposed in both the strong and weak formulations. In the strong case, we require that $u(0) = u(1) = 0$, and in the weak case we require that u be chosen from the class $\mathcal{A}_0$ of functions which all are zero at the boundary points. (In the case of nonzero Dirichlet boundary conditions, say $u(0) = u_0$ and $u(1) = u_1$, we would similarly take $\mathcal{A}_{Dir}$ to be the class of functions satisfying these boundary conditions, then choose u from $\mathcal{A}_{Dir}$.) We say that Dirichlet boundary conditions are *essential* because we must explicitly impose them in the the weak form: we restrict our class $\mathcal{A}_{Dir}$ to contain only those functions which satisfy the desired boundary conditions.

Neumann and Robin boundary conditions, on the other hand, are *natural* because they do not have to be imposed in the weak form; they are satisfied automatically, almost as if by magic. We do not further discuss Robin conditions here; we shall only consider the following Neumann problem.

$$\begin{gathered} -(au')' = f \text{ in } (0,1), \\ a(0)u'(0) = \nu_0, \quad a(1)u'(1) = \nu_1. \end{gathered} \tag{1.8}$$

Note that our Neumann conditions here look a little different that in Table 1.1. There we required $u'(0) = \tilde{\nu}_0$, whereas here we require $a(0)u'(0) = \nu_0$. Since $a > 0$, however, we may simply take $\tilde{\nu}_0 = \nu_0/a(0)$, so these are simply two ways of stating the same boundary condition.

We now let $\mathcal{A}$ be the class of functions which are sufficiently smooth *but with no boundary conditions imposed.* Multiplying (1.8) by any $v \in \mathcal{A}$ and integrating by parts, we find that

$$\begin{aligned} \textstyle\int_0^1 fv\,dx = -\int_0^1 (au')'v\,dx &= \textstyle\int_0^1 fv\,dx \\ &= \textstyle\int_0^1 au'v'\,dx - a(0)u'(0)v(0) + a(1)u'(1)v(1) \\ &= \textstyle\int_0^1 au'v'\,dx - \nu_0 v(0) + \nu_1 v(1). \end{aligned}$$

The following, then, is the weak form of the Neumann problem (1.8): Find $u \in \mathcal{A}$ such that

$$\int_0^1 au'v'\,dx - \nu_0 v(0) + \nu_1 v(1) = \int_0^1 fv\,dx \text{ for all } v \in \mathcal{A}. \tag{1.9}$$

We have shown that if u satisfies the strong form, it satisfies the weak form. Now we wish to show the converse. We may argue exactly as in the Dirichlet problem that $-(au')' = f$ in $(0,1)$ (we must only be careful to take our bump function $v \in \mathcal{A}_0$ so that boundary terms disappear). Thus we must only show that if u satisfies (1.9), then $a(0)u'(0) = \nu_0$ and $a(1)u'(1) = \nu_1$. Assuming that u satisfies (1.9), let us take $v \in \mathcal{A}$ such that $v(0) = 1$ and $v(1) = 0$ ($v(x) = 1 - x$ certainly does the trick). Then integrating by parts and collecting terms, we find that

$$\begin{aligned} 0 &= \int_0^1 au'v'\,dx - \nu_0 v(0) + \nu_1 v(1) - \int_0^1 fv\,dx \\ &= \int_0^1 (-(au')' - f)v\,dx + (a(0)u'(0) - \nu_0)v(0) - (a(1)u'(1) - \nu_1)v(1). \end{aligned}$$

We have already established that $-(au')' = f$, so $-(au')' - f = 0$, and we chose v so that $v(0) = 1$ and $v(0) = 0$. Continuing from the previous equation, we thus find that

$$a(0)u'(0) - \nu_0 = 0$$

or

$$a(0)u'(0) = \nu_0,$$

which is what we sought to show. We may similarly argue that $a(1)u'(1) = \nu_1$, so that the weak and strong forms of the Neumann problem are equivalent.

To emphasize the main points of this section, we quickly sum up our main definitions and findings. Boundary conditions are said to be *essential* if they must be imposed in the weak form, that is, if one must impose boundary conditions on members of the class $\mathcal{A}$ of admissible functions. Thus Dirichlet conditions are essential. Boundary conditions are *natural* if they are enforced automatically in the weak form, that is, if they need not be imposed on members of the class $\mathcal{A}$ of admissible functions. Neumann conditions are natural.

1.4 Inner products and norms of functions

In the mathematical analysis of numerical methods, it is handy to have a way to measure the size of a given function and the angle between two functions. Before discussing what these ideas might mean in the context of function spaces, we quickly recall from linear algebra the corresponding elementary concepts in Euclidean spaces. Let $\vec{a}$ and $\vec{b}$ be two vectors in $\mathbb{R}^n$, i.e., $\vec{a} = (a_1, a_2, ..., a_n)$ and $\vec{b} = (b_1, b_2, ..., b_n)$. The *inner product* (sometimes called a dot product in the context of Euclidean spaces) of $\vec{a}$ and $\vec{b}$ is then

$$(a, b) = \sum_{i=1}^{n} a_i b_i.$$

The size of a vector is then the Euclidean norm

$$||\vec{a}||_{Euc} = \sqrt{(a, a)},$$

and the distance between any two vectors in $\mathbb{R}^n$ is given by the familiar Euclidean distance

$$dist(\vec{a}, \vec{b}) = ||a - b||_{Euc} = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}.$$

Our intuition that the shortest distance between two points is a straight line is confirmed by the *triangle inequality*

$$||a + b||_{Euc} \leq ||a||_{Euc} + ||b||_{Euc},$$

which tells us that

$$dist(\vec{a}, \vec{b}) \leq dist(\vec{a}, \vec{c}) + dist(\vec{c}, \vec{b})$$

for any vector c. Finally, we may measure the angle θ between two vectors by

$$\cos\theta = \frac{(\vec{a}, \vec{b})}{||\vec{a}||_{Euc}||\vec{b}||_{Euc}},$$

and we say that two vectors are orthogonal (or perpendicular) if their inner product is zero.

The concepts of size and angle are as important in function spaces as they are in the Euclidean spaces we are so familiar with, but the definitions of these concepts is not so intuitive in the context of function spaces (and in fact there are many different possible definitions that might prove to be useful). We begin by discussing the bilinear form $\mathcal{L}$ which we defined when discussing the weak form (1.3) of our TPBVP. It turns out that $\mathcal{L}$ functions as an inner product in the space $\mathcal{A}_0$ of admissible functions for the weak form. While one does not generally care if the angle between two functions in $\mathcal{A}_0$ is $\frac{\pi}{3}$ or $\frac{\pi}{4}$, it turns out that the concept of orthogonality in $\mathcal{A}_0$ is very important. That is, two functions $u, v \in \mathcal{A}_0$ are orthogonal if $\mathcal{L}(u, v) = 0$. The size of functions (and the corresponding distance between functions) in $\mathcal{A}_0$ is also important. A consistent way of assigning a size to an object in a space is called a *norm*, and here we shall employ the *energy norm*

$$||u||_{eng} = \sqrt{\mathcal{L}(u, v)}.$$

We note that the energy norm is very intimately connected with the TPBVP that we are considering: it depends on the coefficient a in our problem. At least for the time being, however, this norm will be a convenient way for us to measure the size of a function.

It is also possible to devise other ways to measure the size of a function, some corresonding to an inner product and some not. For example, we can consider the *maximum* (or L_∞) norm

$$||u||_{L_\infty} = \max_{0 \leq x \leq 1} |u(x)|,$$

the L_1 norm

$$||u||_{L_1} = \int_0^1 |u|\, dx,$$

or the L_2 (or *root-mean-square*) norm

$$||u||_{L_2} = \int_0^1 u^2\, dx = \sqrt{(u, u)}.$$

There is no natural concept of angle (i.e., no natural inner product) corresponding to the L_∞ and L_1 norms; the natural inner product corresponding to the L_2 norm is the form $(u, v) = \int_0^1 uv\, dx$ which we previously defined.

1.5 Exercises

(Problems marked with an asterisk (*) are optional, but you should know how to do them. Problems marked with a double asterisk (**) are included for extra interest, and you need not complete them. All others are to be handed in.)

1. Here we consider a problem with mixed boundary conditions: Find a function u such that

$$\begin{aligned} -(a(x)u'(x))' &= f(x) \text{ in } (0,1), \\ u'(0) = 0, &\quad u(1) = 0. \end{aligned} \tag{1.10}$$

a. Show that (1.10) has precisely one solution. (Hint: Integrate to show that a solution exists. To show that only one solution exists, assume that two solutions exist and show that they are equal, which is a standard "trick of the trade".)

b. Find the weak form of (1.10). (Hint: Consider the ways in which essential and natural boundary conditions are handled in the weak form.)

c.* Show that the weak form of (1.10) that you found above and the strong form (1.10) are equivalent.

2. An inner product on a vector space X is a function $(\cdot,\cdot)$ which associates a real number to each pair of vectors $\{x, y\} \in X \times X$ and which satisfies the following five requirements:

$$\begin{aligned} &(a)\ (x, y) = (y, x); \\ &(b)\ (x + y, z) = (x, z) + (y, z); \\ &(c)\ (\alpha x, y) = \alpha(x, y) \text{ if } \alpha \in \mathbb{R}; \\ &(d)\ (x, x) \geq 0 \text{ for all } x \in X; \\ &(e)\ (x, x) = 0 \text{ only if } x = 0. \end{aligned}$$

a.* Verify that the standard dot product on $\mathbb{R}^n$ is an inner product.

b. Verify that $\mathcal{L}(\cdot,\cdot)$ is an inner product on $X = \mathcal{A}_0$.

3. a. Find a function which has finite size when measured in the L_1 norm but not when measured in either the L_2 norm or in the L_∞ norm.

b. Is it possible to find a function which has finite size in the L_∞ norm but not in the L_1 norm?

c.** How about a function in $\mathcal{A}_0$ which has finite size in the norm $\|u\|_{W_1^1} = \int_0^1 |u| + \int_0^1 |u'|$ but not in the L_∞ norm?