1 Week 02: 27 January 2009

2 Systems in one dimension

Let $u : [a, b] \to R^s$ be the solution of the well-posed first-order linear two-point boundary value problem

$$Lu := u' + Qu = f \qquad a < x < b$$

with homogeneous boundary conditions

$$Au(a) - Bu(b) = 0,$$

Thus u is a vector function from $[a, b]$ to s-dimensional Euclidean space R^s, and Q is an $n \times n$ matrix function. (A necessary but not sufficient condition for well-posedness is that the $s \times 2s$ partitioned matrix $C = [A|B]$ has rank s. The example of computing periodic indefinite integrals where $Q = 0$, $A = B = I$, shows that the rank condition is not sufficient in general.)

We begin by converting the equation to a "dual" form. There are several ways to do this, leading to different formulations and numerical methods. Assume the right-hand side f is in the space L^1 of absolutely integrable functions on $[a, b]$:

$$L^1 := \{f : [a, b] \to R^s \mid \|f\|_{L^1} := \int_a^b |f| < \infty\},$$

wher we measure vector lengths in R^s by the 1-norm

$$|f| = \sum_{j=1}^{n} |f_j|.$$

Then seek the solution u in the corresponding space

$$W = W_0^{1,1} = \{w \in L^1 \mid w' \in L^1,\ Aw(a) + Bw(b) = 0\}.$$

The dual viewpoint says that the differential equation is satisfied iff all "test functions" v belonging to the dual space V of L^1 agree: quantitatively, the bilinear form

$$a(u, v) := \int_a^b v^T(u' + Qu) = (f, v) := \int_a^b v^T f$$

agrees with the duality pairing (f, v) for all v. The dual space of L^1 turns out to be the space

$$L^\infty = \{v : [a, b] \to R^s \mid \|v\|_{L^\infty} := \max_{a \le x \le b} |v(x)| < \infty\}$$

of bounded functions, so the test functions v need to vary over all bounded functions. There are no boundary conditions on v, and none would make sense because v need not be continuous. Changing the values of v at any finite set of points doesn't change v as an element of L^∞.

The original differential equation and boundary conditions are said to be well-posed (relative to the specified spaces of data f and solutions u) iff for every $f \in L^1$ there is a unique solution $u \in W$, and a stability estimate is satisfied which says that the solution is bounded by the data in the norms of the specified spaces:

$$\|u\|_W := (b - a)\|u'\|_{L^1} + \|u\|_{L^1} \le C\|f\|_{L^1}$$

for some constant C independent of u and f. In terms of the operator L, well-posedness requires that (a) the range of L is the whole space L^1, (b) the kernel of L is the zero function, and (c) the operator norm of the inverse of L is bounded by C.

Well-posedness is exactly equivalent to the BNB or "inf-sup" condition on the dual formulation, which requires two things. First is the range condition BNB_1, that any test function $v \in V$ which reports 0 on every potential solution $u \in W$ must in fact be the zero function, or

$$v \in V, a(w, v) = 0 \,\text{for all}\, w \in W$$

implies that $v = 0$. Second is the stability condition BNB_2, similar to the equivalent norm condition in the coercive case:

$$\sup_{v \in V} \frac{a(w, v)}{\|v\|_V} \ge \alpha\|w\|_W$$

for some constant α independent of w.

Indeed, suppose the original boundary value problem is well-posed. For the range condition, suppose

$$a(w, v) = \int_a^b v^T(w' + Qw) = 0$$

for all $w \in W$, and construct a potential solution w of the BVP which makes $a(w, v)$ as large as possible. Since the BVP is well-posed, we can find a solution $w \in W$ of the problem

$$(w' + Qw)_j = \text{sign}(v_j) = v_j/|v_j|$$

because the right-hand side $\text{sign}(v)$ is bounded by 1, hence in $L^\infty \subset L^1$. (Bounded functions are absolutely integrable over finite intervals.) That's all we need, because for this $w \in W$ we have

$$0 = a(w, v) = \int_a^b \sum_j |v_j| = \int_a^b |v|$$

which implies $v = 0$ in L^∞ and proves BNB_1. For the stability condition, we choose a potential solution $w \in W$ and arrange a test function v to make $a(w, v)$ as large as possible compared to $\|v\|_{L^\infty}$. The function

$$v_j(x) = \text{sign}(w' + Qw)_j$$

has norm 1 in $V = L^\infty$, and

$$a(w, v) = \int_a^b |w' + Qw| \geq \frac{1}{C}\|w\|_W$$

by the boundedness of L^{-1}. Hence BNB_2 is satisfied with $\alpha = 1/C$.

The converse is also true, and even easier to prove, so the BNB or inf-sup condition on a is equivalent to well-posedness of the original BVP, or bounded invertibility of the operator $L : W \to V$.

The finite element method discretizes the dual formulation with the aid of solution and test spaces $W_h \subset W$ and $V_h \subset V$, defining the approximate solution $u_h \in W_h$ by the requirement

$$a(u_h, v_h) = (f, v_h) \qquad \text{for all}\, v_h \in V_h.$$

Clearly the first requirement on W_h and V_h is that their dimensions should match, so that u_h can be found by solving a square linear system. The simplest choice makes u_h a continuous piecewise-linear function satisfying the boundary conditions:

$$W_h = \{u_h = \sum_{j=0}^{N} u_j \phi_j(x) \,|\, u_j \in R^s,\ Au_0 - Bu_N = 0\}.$$

Here the vectors u_j are the values of u_h at the grid points x_j, and ϕ_j is the scalar hat function

$$\phi_j(x) = \max(0, 1 - |x - x_j|/h).$$

The dimension of this space is $(N+1)s$ for the vectors u_j, minus s for the s linearly independent boundary conditons, for a total of Ns. The natural test function space has one derivative less since $a(u, v)$ involves one derivative of u and none of v, and no continuity requirements or boundary conditions. Thus V_h should be piecewise constant functions

$$v_h(x) = v_j \in R^s, \qquad x_{j-1} \leq x < x_j, \qquad 1 \leq j \leq N.$$

The dimensions match since $\dim(V_h) = Ns = \dim(W_h)$. Thus the requirement

$$a(u_h, v_h) = (f, v_h)$$

for all $v_h \in V_h$, with the boundary conditions $Au_0 - Bu_N = 0$, constitutes a square linear system

$$L_h u = f$$

for the coefficient vector $u \in R^{(N+1)s}$. If the matrix coefficient Q is constant, then the coefficient integrals involved can be evaluated analytically and we get the familiar midpoint rule

$$(I + \frac{hQ}{2})u_{j+1} - (I - \frac{hQ}{2})u_j = f_j$$

for $0 \leq j \leq N-1$, coupled with the boundary conditions $Au_0 - Bu_N = 0$. Here the right-hand side vectors f_j are cell averages of the right-hand side f. From previous experience, we expect the midpoint rule to give second-order accuracy with centered data such as these cell averages. However, we may measure the error in different norms and get different rates of convergence in the finite element context.

For numerical purposes, this square linear system should be well-posed independent of the mesh size h: the matrix L_h should be invertible, with a bounded inverse $\|L_h^{-1}\| \leq S$ for some stability constant S independent of h. In the dual formulation, this translates to a discrete inf-sup condition: whenever $a(w_h, v_h) = 0$ for all $w_h \in W_h$ then $v_h = 0$, and

$$\sup_{v_h \in V_h} \frac{a(w_h, v_h)}{\|v_h\|_V} \geq \alpha \|w\|_W$$

for some constant α independent of h. We can easily prove these conditions for the easy case when $Q = 0$, because then the exact operator transforms continuous piecewise-linear functions in W_h into piecewise constants in V_h.

Indeed, let v_h be such that $a(w_h, v_h) = 0$ for all $w_h \in W_h$. Let w_h be the exact solution of the well-posed BVP $w_h' + Qw_h = \text{sign}(v_h)$. Then $w_h \in W_h$, and

$$0 = a(w_h, v_h) = \int_a^b |v_h|$$

so $v_h = 0$ and the range condition is satisfied. Thus the range of L_h is the whole space V_h, and since $\dim(V_h) = \dim(W_h) < \infty$ this is enough to guarantee invertibility for each fixed h. Stability requires a little more, because we need a bound on the inverse $\|L_h^{-1}\| \leq S$ independent of h, and that is provided by the inf-sup condition.

For the stability condition, fix $w_h \in W_h$ and seek $v_h \in V_h$ to make $a(w_h, v_h)$ as large as possible. For example, take $v_h = \text{sign}(w_h')$ piecewise constant with norm $\|v_h\| = 1$, so

$$a(w_h, v_h) = \int_a^b |w_h'| \geq \frac{1}{C}\|w\|_W$$

by well-posedness (the continuous inf-sup condition).

However, our final goal is not just well-posedness or stability of the discrete problem. We actually want to get the right answer, which requires convergence: $\|u - u_h\|_W \to 0$ as $h \to 0$. Thus we need two more tools: a bound on the optimality of the finite element solution in terms of the approximation power of W_h, and an estimate of approximation power.

It is very easy to estimate the finite element solution error in terms of any approximating function w_h from W_h. First the triangle inequality implies

$$\|u - u_h\|_W \leq \|u - w_h\|_W + \|w_h - u_h\|_W.$$

Use the discrete inf-sup inequality on the second contribution to the error above:

$$\|w_h - u_h\|_W \leq \frac{1}{\alpha} \sup_{v_h \in V_h} \frac{a(w_h - u_h, v_h)}{\|v_h\|}.$$

Estimate this error by a-orthogonality as follows. The exact solution satisfies

$$a(u, v) = (f, v)$$

for all $v \in V$. Since $V_h \subset V$,

$$a(u, v_h) = (f, v_h)$$

for all $v_h \in V_h$. On the other hand, the finite element solution u_h satisfies

$$a(u_h, v_h) = (f, v_h)$$

for all $v_h \in V_h$, so subtraction gives a useful analogue of orthogonality:

$$a(u - u_h, v_h) = 0$$

for all $v_h \in V_h$. Thus the finite element error is a-perpendicular to V_h, which makes u_h analogous to the a-orthogonal projection of u on V_h. Thus u_h is the best thing available in the a-norm (which is not of course a norm). There may be better things available in W_h, but they are not computable the way u_h is.

Using this analogue to substitute u for u_h, we find

$$\|u_h - w_h\|_W \leq \frac{1}{\alpha} \sup_{v_h \in V_h} \frac{a(w_h - u, v_h)}{\|v_h\|}.$$

Since a is a bounded bilinear form, in the sense that $|a(w, v)| \leq K\|w\|_W\|v\|_V$, this can be bounded by

$$\|u_h - w_h\|_W \leq \frac{K}{\alpha}\|u - w_h\|_W$$

and added back to the triangle inequality. Thus

$$\|u - u_h\|_W \leq (1 + K/\alpha)\|u - w_h\|_W$$

for any $w_h \in W_h$, so the finite element solution is within a non-numerical constant factor of the best possible approximation from W_h.

The next question is how close does W_h come to the exact solution u? This leads to interpolation estimates, covered exhaustively in Chapter 1 of [EG04].

The simplest estimate for continuous piecewise-linear interpolation in $W^{1,1}$ suffices for the present analysis. First observe that any function u in $W^{1,1}$ satisfies the basic inequality

$$|u(x) - u(y)| = |\int_x^y u'(t)dt| \leq \int_a^b |u'| = \|u'\|_{L^1}.$$

If y is a point where $|u(y)|$ achieves its global minimum, then

$$\int_a^b |u(t)|dt \geq (b-a)|u(y)|$$

so at any point x,

$$|u(x)| \leq |u(x)-u(y)| + |u(y)| \leq \frac{1}{b-a}\|u\|_{L^1} + \|u'\|_{L^1} = \frac{1}{b-a}\|u\|_{W^{1,1}}.$$

Hence all potential solutions $w \in W$ are bounded and

$$\|u\|_{L^\infty} \leq \frac{1}{b-a}\|u\|_{W^{1,1}}.$$

The continuous piecewise-linear interpolant p to equidistant grid values u_j of $u \in W$ is given by

$$p(x) = \theta u_{j+1} + (1-\theta)u_j = \frac{x-x_j}{h}u_{j+1} + \frac{x_{j+1}-x}{h}u_j$$

on the interval $x_j \leq x = \theta x_{j+1} + (1-\theta)x_j < x_{j+1}$. We want to show p is bounded by a constant times u in W, so the interpolation operator is bounded, and then that the error $u-p$ goes to zero in the norm of W as $h \to 0$. For the first, we have

$$\int_a^b |p(x)|dx = h\sum_{j=1}^{N}\int_{x_{j-1}}^{x_j} |\theta u_j + (1-\theta)u_{j+1}|d\theta \leq h\sum_{j=1}^{N}\|u\|_{L^\infty} = \|u\|_{L^\infty},$$

and

$$\int_a^b |p'(x)|dx = \sum_{j=1}^{N}|u_j - u_{j-1}| \leq \sum_{j=1}^{N}\int_{x_{j-1}}^{x_j} |u'(x)|dx = \|u'\|_{L^1}.$$

Summing up,

$$\|p\|_{W^{1,1}} \leq \|u\|_{W^{1,1}}$$

so the interpolation operator $u \to p$ is bounded on $W^{1,1}$.

The error estimate requires a higher smoothness assumption

$$u \in W^{2,1} = \{w \in W^{1,1} \,|\, w' \in W^{1,1}\}.$$

Such estimates hold if the right-hand side $f \in W^{1,1}$, because then $u'' = f' - Q'u - Qu'$ is a sum of L^1 functions.

The estimate is derived by starting with the derivative errors

$$e(x) = u'(x) - p'(x),$$

which vanishes at some point z_j on each interval $[x_{j-1}, x_j]$ because $u = p$ at grid points. Thus on each interval,

$$|e(x)| = |e(x) - e(z_j)| \leq \int_{z_j}^{x} |e'(t)|dt \leq \int_{x_{j-1}}^{x_j} |u''(t)|dt$$

since $p'' = 0$. Integrating,

$$\|u' - p'\|_{L^1} \leq \sum_{j=1}^{N} \int_{x_{j-1}}^{x_j} \int_{x_{j-1}}^{x_j} |u''(t)|dtdx = h\|u''\|_{L^1}.$$

Since the lower-order error $u - p$ vanishes at grid points, the same argument shows that

$$|u(x) - p(x)| \leq \int_{x_{j-1}}^{x_j} |e(x)|dx$$

and integration gives

$$\|u - p\|_{L^1} \leq h \int_a^b |e| \leq h^2 \int_a^b |u''|.$$

Adding up the various tems in the W-norm gives

$$\|u - p\|_W \leq Kh\|u\|_{W^{2,1}}$$

completing the proof of first-order accuracy for the continuous piecewise-linear finite element method we derived above.

References

[EG04] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements.* Springer-Verlag, New York, 2004.