Variational Principles for Equilibrium Physical Systems

1. Variational Principles

One way of deriving the governing equations for a physical system is the express the relevant conservation statements and constitutive laws in terms of a set of state variables to obtain a system of differential equations. A second approach is to postulate a variational principle for the physical system. In this approach, one characterizes a scalar variable, one that is often related to the energy of the system, in terms of the system state variables. The variational principle then is the statement that the system will assume the state in which the ”energy” is minimized. Aside from any other reasons for considering the variational approach to the derivation of governing equations, this approach provides a means for weakening the formulation of the equations that govern the behavior of the physical system. We will illustrate the approach with examples. First, we recall some notation and define an alternative norm for the space $H_0^1(U)$.

Equivalent Norms on $H_0^1(U)$

Throughout this chapter we are going to use the notation that for a bounded set U,

$$(u,v)_0 = \int_U u\,v\,dx = L_2 - inner\ product \qquad and \qquad \|u\|_0^2 = (u,u)_0$$

$$(u,v)_1 = \int_U [u\,v + \nabla u \bullet \nabla v] = H^1 - inner\ product \qquad and \quad \|u\|_1^2 = (u,u)_1 = \|u\|_0^2 + \|\nabla u\|_0^2$$

We recall also that the Poincare inequality asserts the existance of a constant C depending only on U such that,

$$\|u\|_0^2 \leq C\,\|\nabla u\|_0^2 \qquad\qquad for\ all\ u \in H_0^1(U),.$$

It follows from this inequality that for all $u \in H_0^1(U)$,

$$\frac{1}{1+C}\,\|u\|_1^2 \leq \ \|\nabla u\|_0^2 \leq \|u\|_1^2.$$

This means that $|u|_1 = \|\nabla u\|_0$ defines a new norm on $H_0^1(U)$ and this new norm is equivalent to the old norm $\|u\|_1$ in the sense that any sequence that is convergent in one of the norms is also convergent in the other. It follows that $H_0^1(U)$ is a Hilbert space for the new inner product and norm given by

$$\langle u,v\rangle_1 = \int_U \nabla u \bullet \nabla v\,dx \qquad and \qquad |u|_1^2 = \langle u,u\rangle_1 \qquad for\ u,v \in H_0^1(U).$$

This is sometimes referred to as the ***Poincare inner product and norm*** on $H_0^1(U)$. Note that since $|C|_1 = 0$ for any constant, this is not a norm on $H^1(U)$.

Transverse Deflection of an Elastic Membrane

Consider an elastic membrane, stretched over a rigid frame lying in the plane. Suppose the frame forms a simple closed curve containing a region U in its interior. Then the curved frame forms the boundary of the domain U containing the membrane.

In its relaxed state the membrane lies in the plane. If we denote the out of plane deflection of the membrane by $u = u(x,y)$ then $u = 0$ corresponds to the relaxed state. Now suppose the membrane is subjected to a transverse loading having force density described by the function $f = f(x,y)$. Then the potential energy stored in the stretched membrane whose deformation state is $u(x,y)$ corresponding to the load $f(x,y)$ can be shown to be given by the expression

$$E[u(x,y)] = \int_U \Big(\tfrac{1}{2} T(x,y) \nabla u \bullet \nabla u - u(x,y) f(x,y) \Big)\, dx\, dy \tag{1.1}$$

Here $T = T(x,y)$ denotes a bounded and strictly positive function representing the tension in the membrane. If we suppose $T(x,y)$ is piecewise continuous, and $f = f(x,y)$ is in $L_2(U)$, then $E[u]$ is a functional whose domain could be taken to be the linear subspace $H_0^1(U)$ in the Sobolev space of order one $H^1(U)$. Clearly $E[u]$ is well defined on this subspace and the condition that the boundary of the membrane is fixed to the rigid frame lying in the plane (i.e., $u(x,y) = 0 \; for \; (x,y) \in \partial U$) is incorporated into the definition of the domain of the functional. The space $H^1(U)$, is often reffered to as the "energy space" since it contains functions $u(x,y)$ for which the energy (1.1) is finite. In general, functions in $L_2(U)$ do not have finite energy and would therefore not be feasible candidates for functions that describe the deformation state of this membrane.

The problem of finding the deformation state $u(x,y)$ corresponding to a given loading $f(x,y) \in L_2(U)$ can be stated as a ***variational principle***

$$\text{Find } u \in H_0^1(U), \quad \text{such that} \quad E[u] \leq E[v] \qquad \text{for all } v \in H_0^1(U) \tag{1.2}$$

Physically, this is equivalent to the assertion that in a state of elastic equilibrium, the membrane will assume the deflection state that minimizes the energy $E[u]$ over all admissible deformation states, u. Without speculating on why this should be true, we accept this statement as a postulate. Our purpose is to consider the mathematical consequences of the problem that ensues.

Note that
$$E[u] = \tfrac{1}{2} a(u,u) - F[u]$$
where
$$a(u,v) = \int_U T(x,y) \nabla u \bullet \nabla v \, dxdy$$
and
$$F[u] = \int_U u(x,y) f(x,y)\, dxdy = (u,f)_0.$$

Clearly $a(u,v)$ is bilinear and $F[u]$ is linear on $H_0^1(U)$. Moreover, if the coefficient $T(x,y)$ satisfies,

$$T_1 \geq T(x,y) \geq T_0 > 0 \;\; for \; (x,y) \in U,$$

then

$$|a(u,v)| \leq \int_U |T(x,y) \nabla u \bullet \nabla v|\, dxdy \leq T_1 \|\nabla u\|_0 \|\nabla v\|_0 \leq T_1 \|u\|_1 \|v\|_1$$

and

$$|F[u]| \leq \|f\|_0 \|u\|_0 \leq \|f\|_0 \|u\|_1$$

from which it is evident that $a(u,v)$ is a bounded bilinear functional and $F[u]$ is a bounded linear functional on $H_0^1(U)$. In addition,

$$a(\phi,\phi) \geq T_0 \|\nabla\phi\|_0^2 = T_0 |\phi|_1^2 \geq \frac{T_0}{2C} \|\phi\|_1^2 \qquad for \; all \; \phi \in D(U)$$

and since the test functions are dense in $H_0^1(U)$, the estimate extends to all $u \in H_0^1(U)$. Then the bounded, symmetric bilinear form $a(u,v)$ is also positive on $H_0^1(U)$ and it follows from lemma 3.1 that the problem of minimizing $E[u]$ over $H_0^1(U)$ is equivalent to the following problem

$$\text{find } u \in H_0^1(U) \text{ such that } a(u,v) = F[v] \text{ for all } v \in H_0^1(U) \tag{1.3}$$

This equation asserts that $u = u(x,y)$ satisfies

$$a(u,\phi) = \int_U T(x,y)\nabla u \bullet \nabla\phi\, dxdy = F[v] = \int_U \phi(x,y)f(x,y)\, dxdy \qquad \forall\phi \in D(U) \subset H_0^1(U).$$

Formal integration by parts leads to

$$\int_U T(x,y)\nabla u \bullet \nabla\phi\, dxdy = \int_{\partial U} \phi(T\nabla u) \bullet n\, dS - \int_U \phi\nabla(T(x,y)\nabla u)\, dxdy, \quad \forall\phi \in D(U) \subset H_0^1(U)$$

Since $\phi = 0$ on ∂U for all test functions ϕ, the boundary integral vanishes and we see that $u(x,y)$ satisfies

$$\int_U \phi\,[\nabla(T(x,y)\nabla u) + f(x,y)]\, dxdy = 0 \qquad \forall\phi \in D(U) \subset H_0^1(U) \tag{1.4}$$

This is just the assertion that

$$\nabla(T(x,y)\nabla u) + f(x,y) = 0 \quad \text{in the sense of distributions on U.}$$

Since the test functions are dense in $H_0^1(U)$, it is evident that (1.4) holds not just for all test functions but for all functions in $H_0^1(U)$ as well and then (1.4) asserts that $\nabla(T(x,y)\nabla u) + f(x,y) = 0$ in a somewhat stronger sense than the distributional sense. In addition, the fact that $u \in H_0^1(U)$ means that u is the limit in the $H^1(U) - norm$ of a sequence of test functions (which all vanish on the boundary of U) so, in some sense u can be said to vanish on the boundary of U. If it were known in addition that $u \in H_0^1(U) \cap C^0(\bar{U})$ then it would follow that $u(x) = 0$ at each point $x \in \partial U$. Without such additional information we have to be content to say that u satisfies the boundary condition in some generalized sense.Thus we would say that $u \in H_0^1(U)$ satisfying (1.2) or (1.3) is a distributional solution of the problem

$$\begin{aligned} -\nabla(T(x,y)\nabla u) &= f(x,y) \qquad in\ U \text{TCItag} \\ u &= 0 \qquad on\ \partial U \end{aligned}$$

In more advanced courses on PDE's it is shown how to infer additional smoothness for the solution u from smoothness assertions about the data f. If the solution has additional smoothness then the solution of (1.2)-(1.3) may satisfy the abstract boundary value problem (1.5) in some sense stronger than the distributional sense. We refer to (1.2), (1.3) and (1.5) respectively as the ***variational formulation***, ***the weak formulation and strong formulation of the Dirichlet boundary value problem*** for the elliptic operator $L[u] = -\nabla(T(x,y)\nabla u)$

Next consider the variational problem

$$Find\ u \in H^1(U), \quad such\ that \quad E[u] \le E[v] \qquad for\ all\ v \in H^1(U) \tag{1.6}$$

where we minimize the energy over the space $H^1(U)$ instead of the closed subspace, $H_0^1(U)$. Physically, this corresponds to releasing the condition that the membrane is attached to the rigid frame on the boundary of U. Here the solution space, $H^1(U)$, contains no mention of the boundary conditions and it remains to be seen what conditions, if any, apply on the boundary.

We continue to have $E[u] = \frac{1}{2}a(u,u) - F[u]$ with $a(u,v)$ and $F[u]$ both bounded on the new domain, $H^1(U)$. Note, however, that

$$a(u,u) \geq T_0\|\nabla u\|_0^2 \qquad \textit{for} \quad u \in H^1(U)$$

does not imply that $a(u,v)$ is positive on $H^1(U),$ since constants belong to $H^1(U)$ and $a(C_1,C_1) = 0$ for any constant, C_1. We remove this difficulty by letting $\hat{H}^1(U)$ denote the quotient space of equivalence classes of functions in $H^1(U)$ where two functions belong to the same equivalence class if they differ by a constant. Then the Poincare norm is a norm on this quotient space and we have

$$a(u,u) \geq T_0\|\nabla u\|_0^2 = T_0|u|_1^2 > 0 \qquad \textit{for} \quad u \in \hat{H}^1(U), \quad u \neq 0.$$

Now it follows that problem (1.6) (with $H^1(U)$ replaced by $\hat{H}^1(U)$) is equivalent to the problem

$$\text{find} \quad u \in \hat{H}^1(U) \quad \text{such that} \quad a(u,v) = F[v] \quad \textit{for} \quad \textit{all} \quad v \in \hat{H}^1(U) \tag{1.7}$$

To see what the strong version of this problem is, write

$$\int_U T(x,y)\nabla u \bullet \nabla v\, dxdy = \int_U v(x,y)f(x,y)\, dxdy \qquad \forall v \in \hat{H}^1(U).$$

Formal integration by parts (i.e., Green's second identity) leads to

$$\begin{aligned}\int_U T(x,y)\nabla u \bullet \nabla v\, dxdy &= \int_{\partial U} v(T\nabla u) \bullet n\, dS \\ &\quad - \int_U v\nabla(T(x,y)\nabla u)\, dxdy, \quad \forall v \in \hat{H}^1(U) \qquad (1.8)\end{aligned}$$

First choose $v \in \hat{H}^1(U)$ to be an arbitrary test function. The test functions are not dense in $\hat{H}^1(U)$ but they are contained in the space so this is perfectly legal. For $v = v(x,y)$ a test function, the boundary integral vanishes and we are left with

$$\int_U T(x,y)\nabla u \bullet \nabla v\, dxdy = -\int_U v\nabla(T(x,y)\nabla u)\, dxdy, \quad \forall v \in C_c^\infty(U),$$

and for u solving (1.7) this implies

$$\int_U v[\nabla(T(x,y)\nabla u) + f(x,y)]\, dxdy = 0, \quad \forall v \in C_c^\infty(U).$$

This is just the statement that

$$-\nabla(T(x,y)\nabla u) = f(x,y) \qquad \text{in the sense of distributions on U.}$$

Now returning to $(1.7),$ we have

$$a(u,v) - F[v] = \int_{\partial U} v(T\nabla u) \bullet n\, dS + \int_U v[\nabla(T(x,y)\nabla u) + f] \;\; dxdy = 0 \qquad \forall v \in \hat{H}^1(U).$$

Now the integral over U is zero if $v \in \hat{H}^1(U)$ is chosen to be a test function. Then this integral must also vanish if $v \in \hat{H}^1(U)$ is chosen to be an arbitrary function in $C^\infty(U),$ since the boundary is a set of measure zero and cannot alter the value of the integral. Then this last equation becomes

$$a(u,v) - F[v] = \int_{\partial U} v(T\nabla u) \bullet n\, dS = 0 \quad \forall v \in C^\infty(U) \subset \hat{H}^1(U).$$

Now the subspace $C^\infty(U)$ is dense in $\hat{H}^1(U)$ so this last equation implies that $(T\nabla u) \bullet n = 0$ on $\partial U,$ at least in some generalized sense we will not be able to precisely define. At any rate, it appears that the equivalent problems, (1.6)-(1.7), are the variational and weak formulations of the following strong problem

$$\begin{aligned} -\nabla(T(x,y)\nabla u) &= f(x,y) \qquad \textit{for}\ (x,y) \in U\,\text{TCItag} \\ (T\nabla u) \bullet n &= 0 \qquad\quad\;\; \textit{on}\ \partial U \end{aligned}$$

which we recognize as the Neumann problem for the elliptic operator, $L[u] = -\nabla(T(x,y)\nabla u)$. Note that the Neumann boundary conditions were not built into the definition of the solution space but appeared only later as a necessary consequence of being a weak or variational solution of the problem. Boundary conditions that must be incorporated into the definition of the solution space, like the Dirichlet boundary condition of the first example, are called ***stable boundary conditions***. Conditions that are not built into the definition of the solution space but appear naturally, like the Neumann boundary conditions, are called ***natural boundary conditions***. In the context of the elastic membrane, the Neumann boundary condition means that the edges of the membrane are free to move in the out of plane direction since there is no constraining force exerted at the boundary.

We have considered the problem of minimizing the quadratic functional $E[u]$ over $H_0^1(U)$ and over the larger space, $H^1(U)$. In the first case the variational problem is equivalent to the weak form of the Dirichlet boundary value problem, while in the second case it is the weak form of the Neumann boundary value problem that is equivalent to the variational problem. Now consider a space, V, that is "between" $H^1(U)$ *and* $H_0^1(U)$, in the sense that $H_0^1(U) \subset V \subset H^1(U)$. We have in mind the space obtained by completing, in the norm of $H^1(U)$, the subspace of infinitely differentiable functions which vanish on a part of ∂U. More precisely, let ∂U be comprised of two disjoint pieces ∂U_1 *and* ∂U_2; i.e., $\partial U = \partial U_1 \cup \partial U_2$. Then let V denote the $H^1(U) - closure$ of the infinitely differentiable functions which are zero on ∂U_1. When $\partial U_1 = \partial U$, $\partial U_2 = empty$, we have $V = H_0^1(U)$ and, when $\partial U_1 = empty$, $\partial U_2 = \partial U$, we have $V = H^1(U)$. We will suppose that neither ∂U_1 *nor* ∂U_2 is empty so that V is strictly between $H_0^1(U)$ *and* $H^1(U)$. Then we consider the problem

$$\text{Find } u \in V, \quad \text{such that} \quad E[u] \leq E[v] \qquad for\ all\ v \in V. \tag{1.10}$$

Now it is not hard to show that the Poincare norm defines a norm on V since V does not contain any nonzero constants. Then $a(u,v)$ is symmetric bounded and positive on V and (1.10) is equivalent to the weak problem,

$$\text{Find } u \in V \text{ such that} \quad a(u,v) = F[v] \quad for \quad all \quad v \in V \tag{1.11}$$

Since V contains the test functions $C_c^\infty(U)$ it follows that a solution of (1.11) satisfies the equation $-\nabla(T(x,y)\nabla u) = f$, in the sense of distributions on U. Then it also follows that

$$a(u,v) - F[v] = \int_{\partial U} v(T\nabla u) \bullet n\,dS = 0 \quad \forall v \in C^\infty(U) \cap \{v = 0 \ on\ \partial U_1\} \subset V.$$

But for $v \in C^\infty(U) \cap \{v = 0\ on\ \partial U_1\}$ the integral over ∂U_1 vanishes,

$$\int_{\partial U} v(T\nabla u) \bullet n\,dS = \int_{\partial U_1} v(T\nabla u) \bullet n\,dS + \int_{\partial U_2} v(T\nabla u) \bullet n\,dS = \int_{\partial U_2} v(T\nabla u) \bullet n\,dS = 0,$$

and since the behavior of the smooth function v *on* ∂U_2 is completely free, we can conclude that $(T\nabla u) \bullet n$ must equal zero on ∂U_2 in some generalized sense. Then the problems (1.10),(1.11) are the variational and weak formulations of the following strong boundary value problem,

$$\begin{aligned} -\nabla(T(x,y)\nabla u) &= f(x,y) \qquad in \quad U \qquad \text{TCItag} \\ u &= 0 \qquad on \quad \partial U_1 \\ (T\nabla u) \bullet n &= 0 \qquad on\ \partial U_2 \end{aligned}$$

Note that the Dirichlet condition is incorporated into the definition of the solution space V, while the Neumann boundary condition is a natural boundary condition and is automatically satisfied by solutions of the weak and variational problems.

2. Weak Formulations

We have just seen that the following problems are alternative formulations of the same problem:

1) Find $u \in H_0^1(U)$, *such that* $E[u] \leq E[v]$ *for all* $v \in H_0^1(U)$
2) Find $u \in H_0^1(U)$ such that $a(u,v) = F[v]$ *for all* $v \in H_0^1(U)$
3) $-\nabla(T(x,y)\nabla u) = f(x,y)$ *in* U and $u = 0$ *on* ∂U

Similarly, the following are alternative formulations of one problem:

4) Find $u \in \hat{H}^1(U)$, *such that* $E[u] \leq E[v]$ *for all* $v \in \hat{H}^1(U)$
5) find $u \in \hat{H}^1(U)$ such that $a(u,v) = F[v]$ *for all* $v \in \hat{H}^1(U)$
6) $-\nabla(T(x,y)\nabla u) = f(x,y)$ *in* U *and* $(T\nabla u) \bullet n = 0$ on ∂U,

as are:

7) Find $u \in V$, *such that* $E[u] \leq E[v]$ *for all* $v \in V$
8) find $u \in V$ such that $a(u,v) = F[v]$ *for all* $v \in V$
9) $-\nabla(T(x,y)\nabla u) = f(x,y)$ *in* U, $u = 0$ *on* ∂U_1 *and* $(T\nabla u) \bullet n = 0$ on ∂U_2,

Recall that in problems 3),6) and 9), we are interpreting the partial differential equation in the distributional sense and the boundary condition in some unspecified sense that is weaker than the pointwise sense. Of course, it may be the case that both the equation and the boundary condition are valid in some stronger sense but we have not done the analysis needed to establish this.

Problems 1,2,3 are the variational, weak and strong formulations for the Dirichlet problem, while 4,5,6 are the variational, weak and strong formulations for the Neumann problem. Problems 7,8,9 are the variational, weak and strong formulations for a mixed BVP with Dirichlet conditions on part of the boundary and Neumann conditions on the remainder of the boundary. The weak and variational problems are equivalent in all three examples and the strong form of the problem is related to the other two in that if $u = u(x,y)$ is a strong solution then it is also a weak and variational solution but the converse does not necessarily hold. Lemma 3.1 implies the existence of a unique solution to the variational/weak problems in these examples. We have no information in any of the examples about whether a strong solution exists or, what is the same thing, if the weak solution is also a strong solution. That a weak solution is, in fact, also a strong solution can be proved under appropriate hypotheses on the data but it requires techniques that are beyond the scope of this class.

In order to see whether it is always the case that a problem has all three formulations, consider the following strong formulation:

Find $u = u(x,y)$ such that $L[u(x,y)] = f(x,y)$ *in* U
and $u = 0$ *on* ∂U.

where

$$L[u(x,y)] = -\nabla(T(x,y)\nabla u) + \vec{c}(x,y) \bullet \nabla u + b(x,y)\,u$$

and

$$\vec{c}(x,y) \bullet \nabla u = c_1(x,y)\partial_x u + c_2(x,y)\partial_y u.$$

This strong problem has a corresponding weak formulation but it does not have a variational formulation. To obtain the weak formulation, multiply both sides of the PDE by an arbitrary function $v = v(x,y) \in H_0^1(U)$, and integrate over U. Then

$$\int_U v(x,y)[-\nabla(T(x,y)\nabla u) + \vec{c}(x,y) \bullet \nabla u + b\,u]\,dx = \int_U f v\,dx,$$

and

$$-\int_{\partial U} v(T(x,y)\nabla u)\bullet\vec{n}\,dS+\int_U[T\nabla u\bullet\nabla v+v\,\vec{c}\bullet\nabla u+b\,u\,v]\,dx=\int_U fv\,dx.$$

Now the boundary integral vanishes for $v\in H_0^1(U)$, and then we have

$$a(u,v)=F[v]\qquad for\ v\in H_0^1(U),$$

where

$$a(u,v)=\int_U[T\nabla u\bullet\nabla v+v\,\vec{c}\bullet\nabla u+b\,u\,v]\,dx,\ \ for\ \ u,v\in H_0^1(U).$$

Note that, because of the term, $v\,\vec{c}\bullet\nabla u,$ the bilinear form $a(u,v)$ is not symmetric; i.e., $a(u,v)\neq a(v,u)$. For a nonsymmetric bilinear form there can be no quadratic functional whose gradient equals $a(u,v)-F[v]$. To see why this happens, recall that the quadratic functional

$$\Phi(u)=\tfrac{1}{2}a(u,u)-F[u]+C$$

satisfies

$$\begin{aligned}\Phi(u+tv)&=\Phi(u)+t[a(u,v)-F[v]]+\tfrac{1}{2}t^2a(u,u)\\&=\Phi(u)+t\,d\Phi(u,v)+\tfrac{1}{2}t^2d^2\Phi(u,u).\end{aligned}$$

Evidently, when the quadratic functional is defined from a bilinear form, then $d\Phi(u,v)$, the "gradient of the functional at u in the direction v" is equal to $a(u,v)-F[v]$. On the other hand, any bilinear form can be written as the sum of a symmetric and an antisymmetric part as follows,

$$a(u,v)=\tfrac{1}{2}(a(u,v)+a(v,u))+\tfrac{1}{2}(a(u,v)-a(v,u))=a_S(u,v)+a_A(u,v).$$

But then $a(u,u)=a_S(u,u)+0$ so the antisymmetric part of the bilinear form cannot contribute to the quadratic functional. Now u minimizes the functional $\Phi(u)\ over\ H_0^1(U)$ if and only if u satisfies

$$a_S(u,v)=F[v]\ for\ all\ v\in H_0^1(U),$$

and this is not the same as

$$a(u,v)=F[v]\text{ for all }v\in H_0^1(U)$$

unless $a(u,v)=a_S(u,v);\ i.e.,$ unless $a(u,v)$ is symmetric.

Even when the bilinear form is not symmetric so there is no variational formulation for the problem, the weak problem

$$Find\ \ u\in H_0^1(U)\ such\ that\ \ a(u,v)=F[v]\qquad for\ v\in H_0^1(U)\qquad(2.2)$$

is still uniquely solvable. We assume that the coefficients satisfy,

$$\begin{aligned}&0<T_0\le T(x,y)\le T_1,\\&0<b_0\le b(x,y)\le b_1,\qquad\qquad (x,y)\in\bar{U},\quad(2.3)\\&|\vec{c}(x,y)|\le c_0\end{aligned}$$

We will show first that the bilinear form $a(u,v)$ is bounded. Write

$$\begin{aligned}|a(u,v)|&\le\int_U|T\nabla u\bullet\nabla v+v\,\vec{c}\bullet\nabla u+b\,u\,v|\,dx\\&\le T_1\|\nabla u\|_0\|\nabla v\|_0+|\vec{c}|\,\|v\|_0\|\nabla u\|_0+|b|\,\|v\|_0\|u\|_0\\&\le(T_1+c_0+b_1)\,\|u\|_1\|v\|_1\qquad\qquad for\ u,v\in H_0^1(U).\end{aligned}$$

This establishes that the bilinear form is bounded on $H_0^1(U)\times H_0^1(U)$. Next, write

$$|a(u,u)| \geq |\int_U [T\,\nabla u \bullet \nabla u + u\vec{c} \bullet \nabla u + b\,u^2]\,dx|.$$

$$|a(u,u)| \geq T_0\|\nabla u\|_0^2 - c_0 \int_U |u|\,|\nabla u|\,dx + b_0\,\|u\|_0^2.$$

Now for all real numbers A and B and $\varepsilon > 0$,

$$\left(\sqrt{\varepsilon}A - \frac{B}{2\sqrt{\varepsilon}}\right)^2 = \varepsilon A^2 - AB + \frac{B^2}{4\varepsilon} \geq 0.$$

It follows that

$$\int_U |u|\,|\nabla u|\,dx \leq \varepsilon\|\nabla u\|_0^2 + \frac{1}{4\varepsilon}\|u\|_0^2,$$

and then we have

$$|a(u,u)| \geq (T_0 - c_0\varepsilon)\,\|\nabla u\|_0^2 + \left(b_0 - \frac{c_0}{4\varepsilon}\right)\|u\|_0^2.$$

If we choose $\varepsilon < T_0/c_0$, then under certain additional assumptions on the coefficients in the problem, there exists a positive constant a_0 such that

$$|a(u,u)| \geq a_0\|u\|_1^2 \qquad \forall u \in H_0^1(U).$$

This shows that the nonsymmetric bilinear form is positive or coercive.

Problem 1 Show that if the coefficients T, b, and c are such that

$$4T_0b_0 > c_0^2 \qquad (2.4)$$

then ε can be chosen such that for some constant a_0,

$$T_0 - c_0\varepsilon \geq a_0 > 0 \qquad \textit{and} \qquad b_0 - \frac{c_0}{4\varepsilon} \geq a_0 > 0. \qquad (2.5)$$

It now follows from (2.5) that for coefficients satisfying (2.4) the nonsymmetric bilinear form $a = a(u,v)$ satisfies the hypotheses of the Lax-Milgram lemma. In addition,

$$F[v] = (f,v)_0 \;\; \textit{for} \;\; v \in H_0^1(U)$$

is a bounded linear functional on $H_0^1(U)$. Then the Lax-Milgram lemma implies that for each $f \in L_2(U)$, there exists a unique weak solution for (2.1). That is, there exists a unique $u \in H_0^1(U)$ satisfying (2.2).

3. Additional Variational Problems

It becomes evident at some point that the fundamental ingredient in the variational formulation of boundary value problems is the Green's identity. This identity makes it possible to reduce certain partial differential equations and associated boundary conditions to a corresponding variational statement. Of course not every boundary value problem can be treated in this way but the class of problems to which it applies includes a large number of problems of practical importance.

Interface Problems

Consider a domain $U \subset R^n$ composed of complementary subdomains, $U_1 \cup U_2 = U$, and let Γ denote the interface between U_1 *and* U_2. Let $\Sigma_1 = \partial U_1 \backslash \Gamma$ *and* $\Sigma_2 = \partial U_2 \backslash \Gamma$ denote the ”non-interfacial” portions of the boundary of U_1 *and* U_2 respectively. Now consider the problems

For $k = 1,2$ $\quad -\nabla(A_k(x)\nabla u_k) = f_k(x) \quad x \in U_k \qquad (3.1)$

$$u_k = 0 \quad on \quad \Sigma_k$$

and

$$u_1 = u_2 \qquad on \quad \Gamma.$$

$$[A_1(x)\nabla u_1 - A_2(x)\nabla u_2]{\bullet}n = 0 \quad on \quad \Gamma \qquad (3.2)$$

Here $A_k(x)$ denotes an n by n matrix whose entries are in $H_0^1(U_k)$ *and* $f_k \in H^0(U_k)$. This type of problem corresponds to an elliptic problem in which the domain is composed of two parts having distinctly different physical properties.

Define spaces $H = H^0(U_1) \times H^0(U_2),$

$$V = H^1(U_1) \times H^1(U_2),$$

$$V_0 = H_0^1(U_1) \times H_0^1(U_2),$$

and, for $\vec{u} = [u_1,u_2],\ \vec{v} = [v_1,v_2]$ in V, define

$$a[\vec{u},\vec{v}] = a_1(u_1,v_1) + a_2(u_2,v_2) = \int_{U_1} \nabla v_1 \bullet A_1 \nabla u_1\, dx + \int_{U_2} \nabla v_2 \bullet A_2 \nabla u_2\, dx.$$

Then, for $\vec{u} = [u_1,u_2],\ \vec{v} = [v_1,v_2]$ in $C^\infty(U) \times C^\infty(U)$, the Green's identity leads to

$$a[\vec{u},\vec{v}] = -\int_{U_1} v_1 \nabla(A\nabla u_1)\, dx + \int_{\partial U_1 \backslash \Gamma} v_1 \vec{n}_1 \bullet A_1 \nabla u_1 dS + \int_\Gamma v_1 \vec{n}_1 \bullet A_1 \nabla u_1 dS$$

$$-\int_{U_2} v_2 \nabla(A\nabla u_2)\, dx + \int_{\partial U_2 \backslash \Gamma} v_2 \vec{n}_2 \bullet A_2 \nabla u_2\, dS + \int_\Gamma v_2 \vec{n}_2 \bullet A_2 \nabla u_2\, dS.$$

Now let

$$W = \left\{ \vec{u} = [u_1,u_2] \in V :\ u_k = 0 \ on\ \partial U_k \backslash \Gamma,\ k = 1,2 \ \ and\ u_1 = u_2\ on\ \Gamma \right\}.$$

Then

$$V_0 \subset W \subset V \subset H,$$

and, for any $\vec{u},\ \vec{v} \in W \cap C^\infty(U),$

$$a[\vec{u},\vec{v}] = -\int_{U_1} v_1 \nabla(A\nabla u_1)\, dx - \int_{U_2} v_2 \nabla(A\nabla u_2)\, dx + \int_\Gamma v_2 \vec{n}_1 \bullet [A_1 \nabla u_1 - A_2 \nabla u_2]\, dS,$$

because $\quad v_1 = 0 \ \ on\ \partial U_1 \backslash \Gamma, \qquad v_2 = 0 \ \ on\ \partial U_2 \backslash \Gamma, \qquad \vec{n}_1 = -\vec{n}_2\ on\ \Gamma.$

Now it follows that $\vec{u} = [u_1,u_2] \in W$ is a weak solution of the transmission problem, (3.1),(3.2) if

$$a[\vec{u},\vec{v}] = -\int_{U_1} v_1 f_1\, dx - \int_{U_2} v_2 f_2\, dx = \left(\vec{f},\vec{v}\right)_0 \qquad \forall \vec{v} \in W.$$

Then, since $[u_1,u_2] \in W$, it follows that in some generalized sense which we have not clearly defined,

$$u_k = 0 \; on \; \partial U_k \backslash \Gamma, \; k = 1,2 \quad and \quad u_1 = u_2 \; on \; \Gamma.$$

In addition, based on the previous calculations, it is clear that we have,

$$-\nabla(A_k(x)\nabla u_k) = f_k \quad in \quad U_k, \qquad k = 1,2 \; ,$$

where the equality here is in the sense of distributions. Finally, we can show in the usual way that the weak solutions satisfy the natural boundary condition,

$$\vec{n}_1 \bullet [A_1\nabla u_1 - A_2\nabla u_2] = 0 \; on \; \Gamma.$$

The precise sense in which this equality holds has not been defined.

Problem 2 Show that W is a Hilbert space for the $H^1(U) - norm.$
Use the Poincare inequality to show that the Poincare norm is equivalent to the V norm on W.
Is W a Hilbert space for the Poincare norm?

Problem 3 Show that $a[\vec{u},\vec{v}]$ is a bounded bilinear form on W (which norm must you use?)

Problem 4 Show that $a[\vec{u},\vec{u}] \geq C|\vec{u}|^2 \;\; for \; all \; u \in W$ (which norm must you use?)

Problem 5 Use the results of problems 3 and 4 to show that the weak form of the transmission problem is uniquely solvable for all $\vec{f} = [f_1,f_2] \in H$

A Higher Order Equation

Consider the following boundary value problem for the so called ***biharmonic equation***

$$\begin{aligned} -\nabla^2\nabla^2 u(x) &= f(x) \qquad in \quad U \subset R^n, \\ u = 0 \quad and \quad \vec{n} \bullet \nabla u &= 0 \qquad on \quad \partial U. \end{aligned} \qquad (3.3)$$

In R^2, $\qquad \nabla^2\nabla^2 u(x,y) = \partial_{xxxx}u(x,y) + 2\partial_{xxyy}u(x,y) + \partial_{yyyy}u(x,y)$

Problem (3.3) is the ***Dirichlet problem for the biharmonic equation***. We will now consider the weak formulation of this problem.
The weak/variational solution of this fourth order problem will reside in the Hilbert space

$$H^2(U) = \left\{u \in H^0(U) : \partial^\alpha u(x) \in H^0(U), \; |\alpha| \leq 2\right\};$$

It is the space of all functions in $H^0(U)$ whose distributional derivatives of order less than or equal to two are also in $H^0(U)$. This linear space is a Hilbert space for the inner product

$$(u,v)_2 = \sum_{|\alpha|\leq 2}(\partial^\alpha u, \partial^\alpha v)_0 \qquad and \qquad \|u\|_2^2 = \sum_{|\alpha|\leq 2}\|\partial^\alpha u\|_0^2.$$

In the case $n = 1$ this becomes

$$(u,v)_2 = (u,v)_0 + (u',v')_0 + (u'',v'')_0 \quad and \quad \|u\|_2^2 = \|u\|_0^2 + \|u'\|_0^2 + \|u''\|_0^2.$$

The proof that $H^2(U)$ is complete is a slight generalization of the proof that showed $H^1(U)$ is complete.

Let $H_0^2(U)$ denote the completion of the test functions in the norm $\|\bullet\|_2$. A Poincare inequality holds for $H_0^2(U)$

$$\sum_{|\alpha|=2}\|\partial^\alpha u\|_0^2 \geq C\|u\|_2 \quad \forall u \in C_0^\infty(U). \qquad (3.4)$$

This inequality can be proved by a slight extension of the proof used for the Sobolev space of order one. Since we didn't give this proof, we can, without loss of generality, skip this proof as well. The inequality implies that the Poincare norm,

$$|u|_2^2 = \sum_{|\alpha|=2}\|\partial^\alpha u\|_0^2, \quad u \in H_0^2(U),$$

defines a norm on $H_0^2(U)$. In 2 dimensions this becomes

$$|u|_2^2 = \|\partial_{xx}u\|_0^2 + \|\partial_{xy}u\|_0^2 + \|\partial_{yy}u\|_0^2 \qquad u \in H_0^2(U).$$

The functions in $H_0^2(U)$ are those functions in $H^2(U)$ that satisfy (in some sense, not specified here), $u = \partial_N u = 0 \ \ on\ \partial U.$

In addition to the Poincare inequality, (3.4), we also have a Green's identity for the biharmonic operator,

$$\int_U v\,\nabla^2\nabla^2 u\;dx = \int_{\partial U}\{v\,\partial_N(\nabla^2 u) + \partial_N v\,\nabla^2 u\}dS + \int_U \nabla^2 v\,\nabla^2 u\;dx \quad \forall u,v \in C^\infty(U). \qquad (3.5)$$

Now let

$$H = H^0(U), \quad V = H^2(U)$$

and $\quad V_0 = H_0^2(U) =$ completion of test functions in the $H^2(U) - norm.$

Then

$$C_0^\infty(U) \subset V_0 \subset V \subset H,$$

and we can define a weak solution of (3.3) to be a function $u \in V_0$ such that

$$a(u,v) = \int_U \nabla^2 v\,\nabla^2 u\;dx = (f,v)_0 \qquad \forall v \in V_0 \qquad (3.5)$$

The associated variational problem is to minimize

$$E[u] = \tfrac{1}{2}a(u,u) - F(u) = \int_U [\tfrac{1}{2}|\nabla^2 u|^2 - fu]\ dx$$

over V_0. In order to apply the Hilbert space lemma to prove existence of a unique solution, we have to show that the bilinear form is bounded and positive.

Problem 6 Show that for $u \in H_0^2(U)$,

a) $\|\partial_{xy}u\|_0^2 \le \|\partial_{xx}u\|_0 \|\partial_{yy}u\|_0 \le \frac{1}{2}\left(\|\partial_{xx}u\|_0^2 + \|\partial_{yy}u\|_0^2\right)$

b) $|(\partial_{xx}u, \partial_{yy}u)_0| = |(\partial_{xy}u, \partial_{xy}u)_0|$

Problem 7 Show that for $u,v \in H_0^2(U)$, $\quad |a(u,v)| \le C\,|u|_2\,|v|_2$

Problem 8 Use the result of problem 6 b) to show that for $u \in H_0^2(U)$,

$$a(u,u) \ge \|\partial_{xx}u\|_0^2 + 2\|\partial_{xy}u\|_0^2 + \|\partial_{yy}u\|_0^2 \ge C\,|u|_2^2$$

Problem 9 What strong boundary value problem is solved by the solution of the variational problem obtained by minimizing $E[u]$ over the space $H^2(U) \cap H_0^1(U)$?

Problem 10 What strong boundary value problem is solved by the solution of the variational problem obtained by minimizing $E[u]$ over the space $H^2(U)$?

4. Approximation Methods

The conditions under which it is possible to construct an analytical (exact) solution for a BVP are rather special and it is often the case that no analytical solution is possible. In such cases the only option is then to try to construct an approximate solution to the BVP (after first ascertaining by abstract methods that the problem has a unique solution in a specific solution class).

Finite difference methods are one way of approximating the solution to a BVP for which analytic methods fail. These methods replace derivatives by difference quotients, thereby changing a differential equation plus boundary conditions into a system of algebraic equations. However, this is not the only way in which a BVP can be transformed into a system of algebraic equations. We will consider three apparently different but ultimately equivalent ways for approximating the solution of a BVP.

The Rayleigh-Ritz Method

The Rayleigh-Ritz method applies only to problems with a variational formulation. Therefore, consider the variational problem,

$$\textit{Find } u \in V, \quad \textit{such that} \quad E[u] \le E[v] \qquad \textit{for all } v \in V \tag{4.1}$$

where $H_0^1(U) \subset V \subset H^1(U)$. Let $\{\phi_1, \ldots, \phi_N\}$ denote N linearly independent functions in V, and let $M = span[\phi_1, \ldots, \phi_N]$. Then consider the following approximation to the problem (4.1)

$$\textit{Find } u_M \in M, \quad \textit{such that} \quad E[u_M] \leq E[v] \qquad \textit{for all } v \in M \subset V \tag{4.2}$$

We refer to M as the approximating subspace and the functions ϕ_k are called "trial functions" or "shape functions". The approximate solution, u_M, belongs to M, and since the functions $\phi_k(x)$ span M, we can write

$$u_M(x) = \sum_{k=1}^{N} C_k \phi_k(x).$$

Now we can define a function of N real variables, $C_1,...,C_N$, by $F(C_1,...,C_N) = E[u_M]$ and seek to minimize the function F over all of R^N. This will have the effect of minimizing $E[u_M]$ over M as required by (4.2). Since we are minimizing F over all of R^N, all minima for F must be interior minima hence we must have

$$\frac{\partial F}{\partial C_k}\left(\hat{C}_1,...,\hat{C}_N\right) = 0 \qquad for \qquad k = 1,...,N \tag{4.3}$$

This is a system of N linear equations for the N unknowns $\hat{C}_1,...,\hat{C}_N$. To see what these equations might look like, recall that quadratic functional, $E[u]$, in the variational problem has the form,

$$E[u] = \tfrac{1}{2}a(u,u) - (f,u)_2.$$

Then

$$\begin{aligned} E[u_M] &= \tfrac{1}{2}a(u_M,u_M) - (f,u_M)_2 \\ &= \tfrac{1}{2}a\left(\sum_{j=1}^{N} C_j \phi_j(x), \sum_{k=1}^{N} C_k \phi_k(x)\right) - \left(f, \sum_{k=1}^{N} C_k \phi_k(x)\right)_2 \\ &= \tfrac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} C_j C_k a(\phi_j(x),\phi_k(x)) - \sum_{k=1}^{V} C_k (f, \phi_k(x))_2 \end{aligned}$$

i.e.,
$$F(C_1,...,C_M) = \tfrac{1}{2}\sum_{j=1}^{N}\sum_{k=1}^{N} C_j C_k A_{jk} - \sum_{k=1}^{N} C_k F_k,$$

where

$$A_{jk} = a(\phi_j(x),\phi_k(x)) = A_{kj} \quad and \quad F_k = (f, \phi_k(x))_2.$$

Then (1.3) is equivalent to the system

$$\begin{aligned} \frac{\partial F}{\partial C_k} &= \tfrac{1}{2}\sum_{k=1}^{N} A_{jk}C_k + \tfrac{1}{2}\sum_{j=1}^{N} A_{jk}C_j - F_k \\ &= \sum_{k=1}^{N} A_{jk}C_k - F_k = 0, \quad k = 1,...,N. \end{aligned} \tag{4.3$'$}$$

If the bilinear form $a(u,v)$ is positive then the matrix A is positive definite and the system (4.3') has a unique solution. The corresponding u_M is then the Rayleigh-Ritz approximation to the solution of the variational problem, (4.1).

Problem 11 Show that if the bilinear form $a(u,v)$ is positive and symmetric, then the matrix $[A_{jk}] = [a(\phi_j,\phi_k)]$ is symmetric and positive definite.

The Galerkine Procedure

Since not every BVP has a variational formulation, the Rayleigh-Ritz procedure can not always be applied. Consider then the following weak formulation of a BVP

$$\text{Find } u \in V \text{ such that } \quad a(u,v) = (f,v)_2 = F[v] \quad for \quad all \quad v \in V \qquad (4.4)$$

Let $\{\phi_1, \ldots, \phi_N\}$ and M be as they were defined in the previous example and let u_M denote the solution of the problem,

$$\text{Find } u_M \in M \text{ such that } \quad a(u_M,v) = F[v] \quad for \quad all \quad v \in M. \qquad (4.5)$$

Obviously, (4.5) is equivalent to

$$\text{Find } u_M \in M \text{ such that } \quad a(u_M,\phi_k) = F[\phi_k] \quad for\ k = 1,...,N. \qquad (4.5')$$

Now $\qquad u_M(x) = \sum_{k=1}^{N} B_k\,\phi_k(x)$

satisfies $\qquad a\left(\sum_{j=1}^{N} B_j\,\phi_j(x), \phi_k\right) = F[\phi_k],$

i.e.,

$$\sum_{j=1}^{N} B_j \;\; a(\phi_j,\phi_k) = \sum_{j=1}^{Mn} A_{jk} B_j \; = F_k, \quad for\ k = 1,...,N.$$

Clearly these are the same equations as the system (4.3’). The Rayleigh-Ritz procedure only applies to problems having a variational formulation, while the Galerkine procedure applies to the weak formulation whether or not there is a variational formulation. In the case that there is no variational formulation, the matrix $[A_{jk}]$ is not symmetric. However, if the original bilinear form $a(u,v)$ is positive, then the matrix will be positive definite and the approximate problem is uniquely solvable.

The Method of Weighted Residuals

Consider the following strong formulation of a BVP

$$\begin{aligned} -\nabla(T(x,y)\nabla u) + \vec{c}(x,y) \bullet \nabla u + b(x,y)\,u &= f(x,y) && in \quad U \\ u &= 0 && on \quad \partial U_1 \\ (T\nabla u) \bullet n &= 0 && on \quad \partial U_2 \end{aligned}$$

We can write this as

$$\text{Find } u \in V \text{ such that } \quad L[u(x,y)] = f(x,y) \qquad (4.6)$$

Where ∂U is comprised of two disjoint pieces $\partial U_1 \ and \ \partial U_2$; i.e., $\partial U = \partial U_1 \cup \partial U_2$, and V denotes the $H^1(U) - closure$ of the infinitely differentiable functions which are zero on ∂U_1. We will suppose that neither $\partial U_1 \ nor \ \partial U_2$ is empty so that V is strictly between $H_0^1(U) \ and \ H^1(U)$. To construct an approximate solution by the method of weighted residuals, the trial functions $\{\phi_1, \ldots, \phi_N\}$ are chosen to be N linearly independent functions in V with the additional constraint that $L[\phi_k] \in L_2(U)$ for each k. Then let S denote the N-dimensional subspace spanned by the trial functions. In the previous two methods the trial functions were not required to have this additional smoothness.

With these trial functions, let $\qquad u_S(x) = \sum_{k=1}^{N} c_k\,\phi_k(x)$

be define by

$$(L[u_S],\phi_j)_2 = (f,\phi_j)_2 \qquad for \quad j = 1,...,N \qquad (4.7)$$

Note that $$(L[u_S],\phi_k)_2 = \left(L\left[\sum_{k=1}^{N} c_k\phi_k(x)\right],\phi_j\right)_2 = \sum_{k=1}^{N} c_k(L[\phi_k],\phi_j)_2,$$

and
$$(L[\phi_k],\phi_j)_2 = a(\phi_k,\phi_j) \qquad for \;\; \phi_k,\phi_j \in S \subset V.$$

Then (4.7) is equivalent to
$$\sum_{k=1}^{N} A_{kj}c_k = F_j, \qquad for \;\; j = 1,...,N \qquad (4.7')$$

Evidently, the method of weighted residuals leads to the same equations as the other two methods although the trial functions carry an extra smoothness constraint that is not required of the Galerkine and Rayleigh-Ritz trial functions.

The success of these approximation methods depends on making a good choice of trial functions. If the trial functions are chosen from families of piecewise polynomials called finite element spaces then several advantages arise:

- it is possible to deal systematically with irregular regions, even ones having curved boundaries
- the accuracy of the approximation can be estimated in a systematic way in terms of the adjustable parameters characterizing the finite element family
- the ingredients of the approximation problem, including the coefficient matrix $[A_{jk}]$ and the data vector $[F_k]$ in the system of algebraic equations and even the mesh decomposition of the region can be efficiently generated by packaged software.

Employing one of these three approximation techniques in concert with a finite element family of trial functions is referred to as the ”finite element method”.

We should note that the Hilbert space $H^1(U)$ is the ”best” solution space in which to look for solutions to second order elliptic boundary value problems for several reasons:

- an elliptic BVP of order 2 may fail to have any solution in a smaller space (e.g. $H^2(U)$). For example, this could be due to irregularities in ∂U
- solutions in a space larger than $H^1(U)$ may not make sense physically. For example, the functions may fail to have finite energy.
- it is easy to build finite dimensional subspaces of $H^1(U)$ which are convenient for constructing approximate solutions to the BVP. This is less easy for spaces like $H^2(U)$ or $C(\bar{U})$.

We should also be aware that there are various possible formulations we could imagine for a weak solution to the problem,
$$-\nabla^2 u = F \qquad in \;\; U, \qquad u = 0 \;\; on \;\; \partial U.$$

(a) **Ultra-regular Weak solution** Find $u \in H^2(U) \cap H_0^1(U)$ such that
$$\int_U (\nabla^2 u + F)\, v\, dx = 0 \quad \text{for all } v \in L_2(U)$$

(b) **Weak solution** Find $u \in H_0^1(U)$ such that
$$\int_U \nabla u \bullet \nabla v\, dx = \int_U F v\, dx \quad \text{for all } v \in H_0^1(U)$$

(c) **Ultra-Weak solution** Find $u \in L_2(U)$ such that

$$\int_U (\nabla^2 v + F) u \, dx = 0 \quad \text{for all } v \in H^2(U) \cap H_0^1(U).$$

In formulation (a) we are obliged to construct a family $\{V_n\}$ of approximate solution spaces which approach $H^2(U) \cap H_0^1(U)$ with increasing n. Then the family $\{W_n\}$ of test function spaces is just $L_2(U)$ for every n. In formulation (c) the situation is reversed with the space V_n of approximate solutions equal to $L_2(U)$ for all n and the family $\{W_n\}$ of test function spaces approaching $H^2(U) \cap H_0^1(U)$ with increasing n. In case (b) we make the choice $\{V_v\} = \{W_n\}$ and these spaces must approach $H_0^1(U)$ with increasing n. This compromise leads to an approximate problem with a symmetric and positive definite matrix as an approximation to the operator in the BVP. This does not happen in the other two cases. This fact, coupled with the difficulty in building a sequence of finite dimensional spaces tending to $H^2(U) \cap H_0^1(U),$ suggests that (b) is the optimal weak formulation of the BVP.