Variational Method in Hilbert Space

1. A Preview

Boundary-value problems lead us to consider certain function spaces. Consider first the space H of real-valued functions $u(\cdot)$ on the interval (a, b), each of which is continuous except for a finite number of discontinuities and for which the Riemann integral $\int_a^b |u(x)|^2 dx$ is finite. (This could be an improper integral.) Addition and scalar multiplication are defined pointwise. We say that $u(\cdot) = 0$ *almost everywhere* if $u(x) = 0$ for all but a finite number of points $x \in (a, b)$. Then we *identify* any two functions $u(\cdot)$, $v(\cdot)$ if $u(\cdot) = v(\cdot)$ almost everywhere, that is, if $u(\cdot) - v(\cdot) = 0$ almost everywhere. The set of functions $v(\cdot)$ for which $u(\cdot) = v(\cdot)$ almost everywhere is called the *equivalence class* of $u(\cdot)$; note that the integral $\int_a^b v(x) dx$ has the same value for all such $v(\cdot)$. We identity the entire equivalence class with $u(\cdot)$. It is easy to check that H is a linear space and that

$$(u, v)_H \equiv \int_a^b u(x)v(x)\, dx$$

is a scalar product on H. Note that the corresponding norm is

$$\|u\|_H = \Big(\int_a^b |u(x)|^2\, dx\Big)^{\frac{1}{2}}.$$

It is necessary to extend this space somewhat, in the same way that the set of rational numbers is extended to obtain the real number system. The appropriate technical modification of the above is to use the more general *Lebesgue integral* and to let *almost everywhere* mean *except on a set of measure zero*. All finite sets have measure zero as before, but there are also infinite sets of measure zero. Then we denote the corresponding space by $L^2(a, b)$. Hereafter, we set $H = L^2(a, b)$. Further discussion of these topics would require a discussion of Lebesgue integration.

We need a notion of *derivative*. By $u' \in H$ we mean that u is an anti-derivative of a function in H, hence, it is absolutely continuous and the classical derivative $u'(x)$ exists at a.e. point x in (a, b). We recall two of the boundary-value problems from above. Let $c \in \mathbb{R}$ and $F \in H$ be given. The *Dirichlet problem* is to find

$$u \in H : -u'' + cu = F \text{ in } H, \qquad u(a) = u(b) = 0 \ ,$$

and the *Neumann problem* is to find

$$u \in H : -u'' + cu = F \text{ in } H, \qquad u'(a) = u'(b) = 0 \ .$$

An implicit requirement of each of these classical formulations is that $u'' \in H$. This smoothness condition can be relaxed in the following way: multiply the equation by $v \in H$ and integrate. If also $v' \in H$ we obtain the following by an integration-by-parts. Let $V_0 \equiv \{v \in H : v' \in H$ and $v(a) = v(b) = 0\}$; a solution of the Dirichlet problem is characterized by

$$u \in V_0 \text{ and } \int_a^b (u'v' + cuv)\, dx = \int_a^b Fv\, dx \ , \qquad v \in V_0 \ . \tag{1}$$

Similarly, a solution of the Neumann problem satisfies

$$u \in V_1 \text{ and } \int_a^b (u'v' + cuv)\, dx = \int_a^b Fv\, dx \ , \qquad v \in V_1 \ , \tag{2}$$

where $V_1 \equiv \{v \in H : v' \in H\}$. These are the corresponding *weak formulations* of the respective problems. We shall see below that they are actually equivalent to their respective classical formulations. Moreover we see already the primary ingredients of the variational theory:

(1) *Functionals*. Each function, e.g., $F \in H$, is identified with a functional, $\tilde{F} : H \to \mathbb{R}$, defined by $\tilde{F}(v) \equiv \int_a^b Fv\, dx$, $v \in H$. This identification is achieved by way of the L^2 scalar product. For a pair $u \in V_1$, $v \in V_0$ an integration by parts shows $\tilde{u}'(v) = -\tilde{u}(v')$. Thus, for this identification of functions with functionals to be consistent with the usual differentiation of functions, it is *necessary* to define the *generalized derivative* of a functional f by $\partial f(v) \equiv -f(v')$, $v \in V_0$.

(2) *Function Spaces*. From $L^2(a, b)$ and the generalized derivative ∂ we construct the Sobolev space $H^1(a, b) \equiv \{v \in L^2(a, b) : \partial v \in$

$L^2(a,b)\}$. We see that $H^1(a,b)$ is a linear space with the scalar product

$$(u,v)_{H^1} \equiv \int_a^b (\partial u \partial v + uv)\, dx$$

and corresponding norm $\|u\|_{H^1} = (u,u)_{H^1}^{1/2}$, and each of its members is absolutely continuous, hence, $u(x) - u(y) = \int_y^x \partial u$ for $u \in H^1(a,b)$, $a < y < x < b$. This space arises naturally above in the Neumann problem; we denote by $H_0^1(a,b)$ the subspace $\{v \in H^1(a,b) : v(a) = v(b) = 0\}$ which occurs in the Dirichlet problem.

(3) *Forms.* Each of our weak formulations is phrased as

$$u \in V : a(u,v) = f(v)\ , \qquad v \in V\ ,$$

where V is the appropriate linear space (either H_0^1 or H^1), $f = \tilde{F}$ is a continuous linear functional on V, and $a(\cdot,\cdot)$ is the bilinear form on V defined by

$$a(u,v) = \int_a^b (\partial u \partial v + cuv)\, dx\ , \qquad u, v \in V\ .$$

This form is bounded or *continuous* on V: there is a $C > 0$ such that

$$|a(u,v)| \le C\|u\|_V \|v\|_V\ , \qquad u, v \in V\ . \tag{3}$$

Moreover, it is V*-coercive*, i.e., there is a $c_0 > 0$ for which

$$|a(v,v)| \ge c_0 \|v\|_V^2\ , \qquad v \in V$$

in the case of $V = H^1(a,b)$ if (and only if!) $c > 0$ and in the case of $V = H_0^1(a,b)$ for any $c > -2/(b-a)^2$. (This last inequality follows from (6) below, but it is not the optimal constant.) We shall see that the weak formulation constitutes a well-posed problem whenever the bilinear form is bounded and coercive.

1.1. Functionals. But first we consider the notion of a generalized derivative of functions and, even more generally, of *functionals.* Let $-\infty \le a < b \le +\infty$. The *support* of a function $\varphi : (a,b) \to \mathbb{R}$ is the closure in (a,b) of the set $\{x \in (a,b) : \varphi(x) \ne 0\}$. We define $C_0^\infty(a,b)$ to be the linear space of those infinitely differentiable functions $\varphi : (a, b) \to \mathbb{R}$ each of which has compact support in (a, b). An example is given by

$$\varphi(x) = \begin{cases} \exp[-1/(1 - |x|^2)] \ , & |x| < 1 \\ 0 \ , & |x| \geq 1 \ . \end{cases}$$

We shall refer to $C_0^\infty(a, b)$ as the space of *test functions* on (a, b). A linear functional, $T : C_0^\infty(a, b) \to \mathbb{R}$, is called a *distribution* on (a, b). Thus, the linear space of all distributions is the algebraic dual $C_0^\infty(a, b)^*$ of $C_0^\infty(a, b)$.

Next we show that the space of distributions contains essentially all functions. A (measurable) function $u : (a, b) \to \mathbb{R}$ is *locally integrable* on (a, b) if for every compact set $K \subset (a, b)$, we have $\int_K |u| \, dx < \infty$. The space of all such (equivalence classes of) functions is denoted by $L^1_{loc}(a, b)$. Suppose u is (a representative of) an element of $L^1_{loc}(a, b)$. Then we define a corresponding distribution $\tilde{u}$ by

$$\tilde{u}(\varphi) = \int_a^b u(x)\varphi(x) \, dx \ , \qquad \varphi \in C_0^\infty(a, b) \ .$$

Note that $\tilde{u}$ is independent of the representative and that the function $u \mapsto \tilde{u}$ is linear from L^1_{loc} to $C_0^{\infty *}$.

Lemma 1. *If $\tilde{u} = 0$ then $u = 0$.*

This is a technical result which means that if $\int u\varphi = 0$ for all $\varphi \in C_0^\infty$, then $u(\cdot) = 0$ almost everywhere in (a, b). A consequence of it is the following.

Proposition 1. *The mapping $u \mapsto \tilde{u}$ of $L^1_{loc}(a, b)$ into $C_0^\infty(a, b)^*$ is linear and one-to-one.*

We call $\{\tilde{u} : u \in L^1_{loc}(G)\}$ the *regular distributions*. Two examples in $C_0^\infty(\mathbb{R})^*$ are the *Heaviside functional*

$$\tilde{H}(\phi) = \int_0^\infty \phi \ , \qquad \phi \in C_0^\infty(\mathbb{R}) \ ,$$

obtained from the *Heaviside function*: $H(x) = 1$ if $x > 0$ and $H(x) = 0$ for $x < 0$, and the constant functional

$$T(\phi) = \int_{\mathbb{R}} \phi \ , \qquad \phi \in C_0^\infty(\mathbb{R})$$

given by $T = \tilde{1}$. An example of a non-regular distribution is the *Dirac functional* given by

$$\delta(\phi) = \phi(0) \ , \qquad \phi \in C_0^\infty(\mathbb{R}) \ .$$

According to Proposition 1 the space of distributions is so large that it contains all functions with which we shall be concerned, *i.e.*, it contains L^1_{loc}. Such a large space was constructed by taking the dual of the "small" space C_0^∞.

1.2. Derivative. Next we shall take advantage of the linear differentiation operator on C_0^∞ to construct a corresponding generalized differentiation operator on the dual space of distributions. Moreover, we shall define the derivative of a distribution in such a way that it is consistent with the classical derivative on functions. Let D denote the classical derivative, $D\varphi = \varphi'$, when it is defined at a.e. point of the domain of φ. As we observed above, if we want to define a generalized derivative ∂T of a distribution T so that for each $u \in C^\infty(a,b)$ we have $\partial\tilde{u} = (\tilde{Du})$, that is,

$$\partial\tilde{u}(\varphi) = -\int_a^b u \cdot D\varphi = -\tilde{u}(D\varphi) \ , \qquad \varphi \in C_0^\infty(a,b) \ ,$$

then we must define ∂ as follows.

DEFINITION 1. *For each distribution $T \in C_0^\infty(a,b)^*$ the derivative $\partial T \in C_0^\infty(a,b)^*$ is defined by*

$$\partial T(\varphi) = -T(D\varphi) \ , \qquad \varphi \in C_0^\infty(a,b) \ .$$

Note that $D : C_0^\infty(a,b) \to C_0^\infty(a,b)$ and $T : C_0^\infty(a,b) \to \mathbb{R}$ are both linear, so $\partial T : C_0^\infty(a,b) \to \mathbb{R}$ is clearly linear. Since ∂ is defined on all distributions, it follows that every distribution has derivatives of all orders. Specifically, every $u \in L^1_{loc}$ has derivatives in $C_0^\infty(a,b)^*$ of all orders.

EXAMPLE 1. *Let f be continuously differentiable on $\mathbb{R}$. Then we have*

$$\partial\tilde{f}(\varphi) = -\tilde{f}(D\varphi) = -\int f D\varphi \, dx = \int Df\varphi = (\tilde{Df})(\varphi)$$

for $\varphi \in C_0^\infty(\mathbb{R})$. The third equality follows from integration-by-parts and all other equalities are definitions.

This shows that the generalized derivative coincides with the classical derivative on smooth functions. Of course the definition was rigged to make this occur!

EXAMPLE 2. *Let $r(x) = xH(x)$ where $H(x)$ is given above. For this piecewise-differentiable function we have*

$$\partial\tilde{r}(\varphi) = -\int_0^\infty xD\varphi(x)\,dx = \int_0^\infty \varphi(x)\,dx = \tilde{H}(\varphi)\ , \qquad \varphi \in C_0^\infty(\mathbb{R})\ ,$$

so $\partial\tilde{r} = \tilde{H}$ even though $Dr(0)$ does not exist.

EXAMPLE 3. *For the piecewise-continuous function $H(\cdot)$ we have*

$$\partial\tilde{H}(\varphi) = -\int_0^\infty D\varphi(x)\,dx = \varphi(0) = \delta(\varphi)\ , \qquad \varphi \in C_0^\infty(\mathbb{R})\ ,$$

so $\partial\tilde{H} = \delta$, the non-regular Dirac functional.

More generally, let $f : \mathbb{R} \to \mathbb{R}$ be absolutely continuous in a neighborhood of each $x \neq 0$ and have one-sided limits $f(0^+)$ and $f(0^-)$ from the right and left, respectively, at 0. Then we obtain

$$\begin{aligned}\partial\tilde{f}(\varphi) &= -\int_0^\infty fD\varphi - \int_{-\infty}^0 fD\varphi = \int_0^\infty (Df)\varphi + f(0^+)\varphi(0)\\ &\quad + \int_{-\infty}^0 (Df)\varphi - f(0^-)\varphi(0) = (\tilde{Df})(\varphi) + \sigma_0(f)\varphi(0), \qquad \varphi \in C_0^\infty\ ,\end{aligned}$$

where $\sigma_0(f) = f(0^+) - f(0^-)$ is the jump in f at 0. That is, $\partial\tilde{f} = \tilde{Df} + \sigma_0(f)\delta$, and this formula can be repeated if Df satisfies the preceding conditions on f:

$$\partial^2\tilde{f} = (\tilde{D^2 f}) + \sigma_0(Df)\delta + \sigma_0(f)\partial\delta\ .$$

For example we have

$$\partial(H \cdot \sin) = H \cdot \cos\ , \quad \partial(H \cdot \cos) = -H \cdot \sin + \delta\ .$$

Before discussing further the interplay between ∂ and D we note that a distribution T on $\mathbb{R}$ is constant if and only if $T = \tilde{c}$ for some $c \in \mathbb{R}$, *i.e.*,

$$T(\varphi) = c\int_{\mathbb{R}} \varphi\ , \qquad \varphi \in C_0^\infty\ .$$

This occurs exactly when T depends only on the *mean value* of each φ. This observation is the key to the description of primitives or antiderivatives of a given distribution. Suppose we are given a distribution S on $\mathbb{R}$; does there exist a *primitive*, a distribution T such that $\partial T = S$? That is, do we have a distribution T for which

$$T(D\psi) = -S(\psi)\ , \qquad \psi \in C_0^\infty(\mathbb{R})\ ?$$

LEMMA 2.

(a) $\{D\psi : \psi \in C_0^\infty(\mathbb{R})\} = \{\zeta \in C_0^\infty(\mathbb{R}) : \int \zeta = 0\}$ *and the correspondence is given by* $\psi(x) = \int_{-\infty}^x \zeta$. *Denote this space by* $\mathcal{H}$.

(b) *Let* $\varphi_0 \in C_0^\infty(\mathbb{R})$ *with* $\int \varphi_0 = 1$. *Then each* $\varphi \in C_0^\infty(\mathbb{R})$ *can be uniquely written as* $\varphi = \zeta + c\varphi_0$ *with* $\zeta \in \mathcal{H}$, *and this occurs when* $c = \int \varphi$.

PROPOSITION 2.

(a) *For each distribution* S *there is a distribution* T *with* $\partial T = S$.

(b) *If* T_1, T_2 *are distributions with* $\partial T_1 = \partial T_2$, *then* $T_1 = T_2 +$ *constant.*

PROOF. (a) Define T on $\mathcal{H}$ by $T(\zeta) = -S(\psi)$, $\zeta \in \mathcal{H}$, $\psi(x) = \int_{-\infty}^x \zeta$, and extend to all of $C_0^\infty(\mathbb{R})$ by $T(\varphi_0) = 0$.

(b) If $\partial T = 0$, then $T(\varphi) = T(\zeta + c\varphi_0) = T(\varphi_0)\int \varphi$, so $T = T(\varphi_0)\tilde{1}$ is a constant.

$\square$

COROLLARY 1. *If* T *is a distribution on* $\mathbb{R}$ *with* $\partial T \in L^1_{loc}(\mathbb{R})$, *then* $T = \tilde{f}$ *for some absolutely continuous* f, *and* $\partial T = \tilde{Df}$.

PROOF. Note first that if f is absolutely continuous then $Df(x)$ is defined for a.e. $x \in \mathbb{R}$ and $Df \in L^1_{loc}(\mathbb{R})$ with $\tilde{Df} = \partial\tilde{f}$ as before. In the converse situation of Corollary 1, let $g \in L^1_{loc}$ with $\tilde{g} = \partial T$ and define $h(x) = \int_0^x g$, $x \in \mathbb{R}$. Then h is absolutely continuous, $\partial(T - \tilde{h}) = 0$, so Proposition 2 shows $T = \tilde{h} + \tilde{c}$ for some $c \in \mathbb{R}$. Thus $T = \tilde{f}$ with $f(x) = h(x) + c$, $x \in \mathbb{R}$. $\square$

COROLLARY 2. *The weak formulations of the Dirichlet and Neumann problems are equivalent to the original formulations.*

Exercise 1. *Define the function* $k(\cdot)$ *by*

$$k(x) = k_1,\ 0 < x < 1, \quad k(x) = k_2,\ 1 < x < 2\,.$$

Find the solution of the problem

$$-\partial(k(\cdot)\partial u(\cdot)) = 0 \ \textit{in}\ L^2(0,2), \quad u(0) = 1, u(2) = 0.$$

Exercise 2. *Let* $V \equiv \{v \in L^2(0,2) : \partial v \in L^2(0,2) \text{ and } v(0) = 0\}$*; the function* $k(\cdot)$ *is defined above. Let* $F(\cdot) \in L^2(0,2)$ *and* $\lambda \in \mathbb{R}$ *be given. Show that the function* $u(\cdot)$ *satisfies*

$$u \in V \ \text{and} \int_0^2 (k(x)\partial u(x)\partial v(x) + \lambda u(x)v(x))\,dx = \int_a^b F(x)v(x)\,dx\,, \qquad v \in V\,,$$

if and only if it satisfies the interface problem

$$u(0) = 0, \quad -\partial(k_1\partial u(x)) + \lambda u(x) = F(x), \quad 0 < x < 1,$$
$$u(1^-) = u(1^+), \quad k_1\partial u(1^-) = k_2\partial u(1^+),$$
$$-\partial(k_2\partial u(x)) + \lambda u(x) = F(x), \quad 1 < x < 2, \quad k_2\partial u(2) = 0\,.$$

2. Hilbert Space

Let V be a linear space over the reals $\mathbb{R}$ and the function $x, y \mapsto (x, y)$ from $V \times V$ to $\mathbb{R}$ be a *scalar product*. The corresponding norm on V is given by $\|x\| = (x,x)^{\frac{1}{2}}$, and we have shown that

$$|(x,y)| \le \|x\|\,\|y\|, \qquad x,\, y \in V. \tag{4}$$

A sequence $\{x_n\}$ *converges to* x in V if $\lim_{n\to\infty} \|x_n - x\| = 0$. This is denoted by $\lim_{n\to\infty} x_n = x$. A convergent sequence is always *Cauchy*: $\lim_{m,n\to\infty} \|x_m - x_n\| = 0$. The space V with norm $\|\cdot\|$ is *complete* if each Cauchy sequence is convergent in V. A complete normed linear space is a *Banach space*, and a complete scalar product space is a *Hilbert space.*

Example 4. *Consider the space* $C[0,1]$ *of (uniformly) continuous functions on the interval* $[0,1]$ *with the norm* $\|x\|_{L^2} = \left(\int_0^1 |x(t)|^2\,dt\right)^{\frac{1}{2}}$*. Let*

$0 < c < 1$ and for each n with $0 < c - 1/n$ define $x_n \in C[0,1]$ by

$$x_n(t) = \begin{cases} 1 \;, & c \le t \le 1, \\ n(t-c)+1 \;, & c - 1/n < t < c, \\ 0 \;, & 0 \le t \le c - 1/n \,. \end{cases}$$

For $m \ge n$ we have $\|x_m - x_n\|_{L^2}^2 \le 1/n$, so $\{x_m\}$ is Cauchy. If $x \in C[0,1]$, then

$$\|x_n - x\|_{L^2}^2 \ge \int_0^{c-1/n} |x(t)|^2\, dt + \int_c^1 |1 - x(t)|^2\, dt \;.$$

This shows that if $\{x_n\}$ converges to x then $x(t) = 0$ for $0 \le t < c$ and $x(t) = 1$ for $c \le t \le 1$, a contradiction. Hence, $C[0,1]$ is not *complete with the norm $\|\cdot\|_{L^2}$.*

EXERCISE 3. *Show that $C[0,1]$* is *complete with the norm*

$$\|x\|_{C[0,1]} = sup\{|x(t)| : \; 0 \le t \le 1\}\,.$$

2.1. Examples. Some familiar examples of Hilbert spaces include Euclidean space $\mathbb{R}^m = \{\vec{x} = (x_1, x_2, \dots, x_m) : x_j \in \mathbb{R}\}$ with $(\vec{x}, \vec{y}) = \sum_{j=1}^m x_j y_j$ and the sequence space $\ell^2 = \{\vec{x} = \{x_1, x_2, x_3, \dots\} : \sum_{j=1}^\infty |x_j|^2 < \infty\}$ with $(\vec{x}, \vec{y}) = \sum_{j=1}^\infty x_j y_j$.

EXERCISE 4. *Show that $\mathbb{R}^m$ and ℓ^2 are Hilbert spaces.*

The Lebesgue space $L^2(a,b) = \{$equivalence classes of measurable functions $f : (a,b) \to \mathbb{R} : \int_a^b |f(x)|^2\, dx < \infty\}$ has the scalar product $(f,g) = \int_a^b f(x) g(x)\, dx$. From the theory of the Lebesgue integral, we find that this space is complete, hence, it is a Hilbert space.

Finally we describe the spaces that naturally arise in the consideration of boundary-value problems. The Sobolev space $H^1(a,b)$ is given by

$$H^1(a,b) = \{u \in L^2(a,b) : \partial u \in L^2(a,b)\}$$

where we have identified $u \cong \tilde{u}$. Thus each $u \in H^1(a,b)$ is absolutely continuous with

$$u(x) - u(y) = \int_y^x \partial u \;, \qquad a \le x \;, y \le b \;.$$

This gives the Hölder continuity estimate

$$|u(x) - u(y)| \le |x-y|^{1/2} \|\partial u\|_{L^2(a,b)} \;, \quad u \in H^1(a,b), a \le x, y \le b \;.$$

(We use the inequality (4) in L^2 to prove this.) If also we have $u(a) = 0$ then there follow

$$|u(x)| \leq (b-a)^{1/2} \|\partial u\|_{L^2(a,b)} , \qquad a \leq x \leq b , \tag{5}$$

$$\|u\|_{L^2(a,b)} \leq \big((b-a)/\sqrt{2}\big) \|\partial u\|_{L^2(a,b)} , \tag{6}$$

and such estimates also hold for those $u \in H^1(a,b)$ with $u(b) = 0$. Let $\lambda(x) = (x-a)(b-a)^{-1}$ and $u \in H^1(a,b)$. Then $\lambda u \in H^1(a,b)$ and $\partial(\lambda u) = \lambda \partial u + (b-a)^{-1} u$, so $\|\partial(\lambda u)\|_{L^2} \leq \|\partial u\|_{L^2} + (b-a)^{-1}\|u\|_{L^2}$. The same holds for $\partial((1-\lambda)u)$ so by writing $u = \lambda u + (1-\lambda)u$ we obtain

$$\begin{aligned} \max\{|u(x)| : a \leq x \leq b\} &\leq 2(b-a)^{1/2}\|\partial u\|_{L^2} \\ &\quad + 2(b-a)^{-1/2}\|u\|_{L^2} , \quad u \in H^1(a,b) . \end{aligned} \tag{7}$$

This simple estimate will be very useful.

Exercise 5. *Show that convergence in $H^1(a,b)$ implies* uniform convergence *on the interval $[a,b]$.*

To verify that $H^1(a,b)$ is complete, let $\{u_n\}$ be a Cauchy sequence, so that both $\{u_n\}$ and $\{\partial u_n\}$ are Cauchy sequences in $L^2(a,b)$. Since $L^2(a,b)$ is complete there are $u, v \in L^2(a,b)$ for which $\lim u_n = u$ and $\lim \partial u_n = v$ in $L^2(a,b)$. For each $\varphi \in C_0^\infty(a,b)$ we have

$$-\int_a^b u_n \cdot D\varphi = \int_a^b \partial u_n \varphi , \qquad n \geq 1 ,$$

so letting $n \to \infty$ shows $v = \partial u$. Thus $u \in H^1(a,b)$ and $\lim u_n = u$ in $H^1(a,b)$.

More generally, we define for each integer $k \geq 1$ the *Sobolev space*

$$H^k(a,b) = \{u \in L^2(a,b) : \partial^j u \in L^2(a,b)\} , \qquad 1 \leq j \leq k .$$

Estimates analogous to those above can be easily obtained in appropriate subspaces.

2.2. Continuity. Let V_1 and V_2 be normed linear spaces with corresponding norms $\|\cdot\|_1$, $\|\cdot\|_2$. A function $T : V_1 \to V_2$ is *continuous at* $x \in V_1$ if $\{T(x_n)\}$ converges to $T(x)$ in V_2 whenever $\{x_n\}$ converges to x in V_1. It is *continuous* if it is continuous at every x. For example, the norm is continuous from V_1 into $\mathbb{R}$. If T is linear, we shall also denote its value at x by Tx instead of $T(x)$.

PROPOSITION 3. *If $T : V_1 \to V_2$ is linear, the following are equivalent:*
(a) *T is continuous at 0,*
(b) *T is continuous at every $x \in V_1$,*
(c) *there is a constant $K \geq 0$ such that $\|Tx\|_2 \leq K\|x\|_1$ for all $x \in V_1$.*

PROOF. Clearly (c) implies (b) by linearity and (b) implies (a). If (c) were false there would be a sequence $\{x_n\}$ in V_1 with $\|Tx_n\|_2 > n\|x_n\|_1$, but then $y_n \equiv \|Tx_n\|_2^{-1}x_n$ is a sequence which contradicts (a). $\square$

EXERCISE 6. *Show that the identity operator $H^1(a,b) \to C[a,b]$ is continuous.*

We shall denote by $\mathcal{L}(V_1, V_2)$ the set of all continuous linear functions from V_1 to V_2; these are called the *bounded linear functions* because of (c) above. Additional structure on this set is given as follows.

PROPOSITION 4. *For each $T \in \mathcal{L}(V_1, V_2)$ we have*

$$\begin{aligned} \|T\| &\equiv \sup\{\|Tx\|_2 : x \in V_1, \|x\|_1 \leq 1\} = \sup\{\|Tx\|_2 : \|x\|_1 = 1\} \\ &= \inf\{K > 0 : \|Tx\|_2 \leq K\|x\|_1\ ,\ x \in V_1\}\ , \end{aligned}$$

and this gives a norm on $\mathcal{L}(V_1, V_2)$. If V_2 is complete, then $\mathcal{L}(V_1, V_2)$ is complete.

PROOF. Consider the two numbers

$$\lambda = \sup\{\|Tx\|_2 : \|x\|_1 \leq 1\}, \quad \mu = \inf\{K > 0 : \|Tx\|_2 \leq K\|x\|_1\ ,\ x \in V_1\}.$$

If K is in the set defining μ, then for each $x \in V_1$ with $\|x\|_1 \leq 1$ we have $\|Tx\|_2 \leq K$, so $\lambda \leq K$. This holds for all such K so $\lambda \leq \mu$. If $x \in V_1$ with $\|x\|_1 > 0$ then $x/\|x\|_1$ is a unit vector and so $\|T(x/\|x\|_1)\|_2 \leq \lambda$. Thus $\|Tx\|_2 \leq \lambda\|x\|_1$ for all $x \neq 0$, and it clearly holds if $x = 0$, so we have $\mu \leq \lambda$. This establishes the equality of the three expressions for $\|T\|$.

EXERCISE 7. *Verify that $\|T\|$ defines a norm on $\mathcal{L}(V_1, V_2)$.*

Suppose V_2 is complete and let $\{T_n\}$ be a Cauchy sequence in $\mathcal{L}(V_1, V_2)$. For each $x \in V_1$,

$$\|T_m x - T_n x\|_2 \leq \|T_m - T_n\|\ \|x\|_1$$

so $\{T_n x\}$ is Cauchy in V_2, hence, convergent to a unique Tx in V_2. This defines $T : V_1 \to V_2$ and it follows by continuity of addition and scalar multiplication that T is linear. Also

$$\|T_n x\|_2 \leq \|T_n\| \ \|x\|_1 \leq \sup\{\|T_m\|\}\|x\|_1$$

so letting $n \to \infty$ shows T is continuous with $\|T\| \leq \sup\{\|T_m\|\}$. Finally, to show $\lim T_n = T$, let $\varepsilon > 0$ and choose N so large that $\|T_m - T_n\| < \varepsilon$ for $m, n \geq N$. Then for each $x \in V_1$ $\|T_m x - T_n x\|_2 \leq \varepsilon\|x\|_1$, and letting $m \to \infty$ gives

$$\|Tx - T_n x\|_2 \leq \varepsilon\|x\|_1 \ , \qquad x \in V_1 \ .$$

Thus, $\|T - T_n\| \leq \varepsilon$ for $n \geq N$. □

As a consequence it follows that the *dual* $V' \equiv \mathcal{L}(V, \mathbb{R})$ of any normed linear space V is complete with the dual norm

$$\|f\|_{V'} \equiv \sup\{|f(x)| : x \in V, \|x\|_V \leq 1\}$$

for $f \in V'$.

2.3. The Minimization Principle. Hereafter we let V denote a Hilbert space with norm $\|\cdot\|$, scalar product $(\cdot,\cdot)$, and dual space V'. A subset K of V is called *closed* if each $x_n \in K$ and $\lim x_n = x$ imply $x \in K$. The subset K is *convex* if $x, y \in K$ and $0 \leq t \leq 1$ imply $tx + (1-t)y \in K$. The following *minimization principle* is fundamental.

THEOREM 1. *Let K be a closed, convex, non-empty subset of the Hilbert space V, and let $f \in V'$. Define $\phi(x) \equiv (1/2)\|x\|^2 - f(x)$, $x \in V$. Then there exists a unique*

$$x \in K : \phi(x) \leq \phi(y) \ , \qquad y \in K \ . \tag{8}$$

PROOF. Set $d \equiv \inf\{\phi(y) : y \in K\}$ and choose $x_n \in K$ such that $\lim_{n\to\infty} \phi(x_n) = d$. Then we obtain successively

$$d \leq \phi\big(1/2(x_m + x_n)\big) = (1/2)\big(\phi(x_m) + \phi(x_n)\big) - (1/8)\|x_n - x_m\|^2 \ ,$$
$$(1/4)\|x_n - x_m\|^2 \leq \phi(x_m) + \phi(x_n) - 2d \ ,$$

and this last expression converges to zero. Thus $\{x_n\}$ is Cauchy, it converges to some $x \in V$ by completeness, and $x \in K$ since it is closed. Since ϕ is continuous, $\phi(x) = d$ and x is a solution of (8). If x_1 and x_2 are both solutions of (8), the last inequality shows $(1/4)\|x_1 - x_2\| \leq d + d - 2d = 0$, so $x_1 = x_2$. □

The solution of the minimization problem (8) can be characterized by a *variational inequality*. For $x, y \in V$ and $t > 0$ we have

$$(1/t)\Big(\phi\big(x+t(y-x)\big)-\phi(x)\Big)=(x,y-x)-f(y-x)+(1/2)t\|y-x\|^2,$$

so the *derivative* of ϕ at x in the direction $y-x$ is given by

$$\begin{aligned}\phi'(x)(y-x)&=\lim_{t\to 0}(1/t)\Big(\phi\big(x+t(y-x)\big)-\phi(x)\Big)\\&=(x,y-x)-f(y-x)\ .\end{aligned}\tag{9}$$

From the definition of $\phi(\cdot)$, we find that the above equals $\phi(y)-\phi(x)+(x,y)-(1/2)\|x\|^2-(1/2)\|y\|^2$, and (4) gives

$$\phi'(x)(y-x)\le\phi(y)-\phi(x)\ ,\qquad x,y\in V\ .\tag{10}$$

Suppose x is a solution of (8). Since for each $y \in K$ we have $x+t(y-x) \in K$ for small $t > 0$, it follows from (9) that

$$x\in K:\varphi'(x)(y-x)\ge 0\ ,\qquad y\in K\ .$$

Conversely, for any such x it follows from (10) that it satisfies (8). Thus, we have shown that (8) is equivalent to

$$x\in K:(x,y-x)\ge f(y-x)\ ,\qquad y\in K\ .\tag{11}$$

The equivalence of (8) and (11) is merely the fact that the point where a quadratic function takes its minimum is characterized by having a non-negative derivative in each direction into the set.

2.4. Consequences of the Principle. As an example, let $x_0 \in V$ and define $f \in V'$ by $f(y) = (x_0, y)$ for $y \in V$. Then $\phi(x) = (1/2)(\|x - x_0\|^2 - \|x_0\|^2)$ so (8) means that x is that point of K which is closest to x_0. Recalling that the angle θ between $x - x_0$ and $y - x$ is determined by

$$(x-x_0,y-x)=\cos(\theta)\|x-x_0\|\ \|y-x\|\ ,$$

we see (11) means x is that point of K for which $-\pi/2 \le \theta \le \pi/2$ for every $y \in K$. We define x to be the *projection* of x_0 on K and denote it by $P_K(x_0)$.

Corollary 3. *For each closed convex non-empty* subset K *of* V *there is a projection operator* $P_K : V \to K$ *for which* $P_K(x_0)$ *is that point of* K *closest to* $x_0 \in V$*; it is characterized by*

$$P_K(x_0) \in K : \big(P_K(x_0) - x_0 \ , y - P_K(x_0)\big) \geq 0 \ , \qquad y \in K \ .$$

It follows from this characterization that the function P_K satisfies

$$\|P_K(x_0) - P_K(y_0)\|^2 \leq (P_K(x_0) - P_K(y_0), x_0 - y_0) \ , \qquad x_0, y_0 \in V \ .$$

From this we see that P_K is a *contraction*, i.e.,

$$\|P_K(x_0) - P_K(y_0)\| \leq \|x_0 - y_0\| \ , \qquad x_0, y_0 \in V \ ,$$

and that P_K satisfies the *angle condition*

$$(P_K(x_0) - P_K(y_0), x_0 - y_0) \geq 0 \ , \qquad x_0, y_0 \in V \ .$$

Corollary 4. *For each closed* subspace K *of* V *and each* $x_0 \in V$ *there is a unique*

$$x \in K : (x - x_0, y) = 0 \ , \qquad y \in K \ .$$

Two vectors $x, y \in V$ are called *orthogonal* if $(x, y) = 0$, and the *orthogonal complement* of the set S is $S^{\perp} \equiv \{x \in V : (x, y) = 0$ for $y \in S\}$. Corollary 4 says each $x_0 \in V$ can be uniquely written in the form $x_0 = x_1 + x_2$ with $x_1 \in K$ and $x_2 \in K^{\perp}$ whenever K is a closed subspace. We denote this orthogonal decomposition by $V = K \oplus K^{\perp}$.

Exercise 8. *Show that* $S^{\perp\perp} = \overline{S}$*, the* closure *of* S.

The *Riesz map* $\mathcal{R}$ of V into V' is defined by $\mathcal{R}(x) = f$ if $f(y) = (x, y)$ for $y \in V$. It is easy to check that $\|\mathcal{R}x\|_{V'} = \|x\|_V$; Theorem 1 with $K = V$ shows the following by way of (11).

Corollary 5. *For each linear functional* $f \in V'$ *there is a unique vector*

$$x \in V : (x, y) = f(y) \ , \qquad y \in V \ . \tag{12}$$

Thus, the linear map $\mathcal{R}$ is *onto* V', so $\mathcal{R}$ is an isometric isomorphism of the Hilbert space V onto its dual V'.

We recognize (12) as the weak formulation of certain boundary value problems. Specifically, when $V = H_0^1$ or H^1, (12) is the Dirichlet problem (1) or the Neumann problem (2), respectively, with $c = 1$. An easy but useful generalization is obtained as follows. Let $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ be bilinear (linear in each variable separately), continuous (see (3)), symmetric ($a(x,y) = a(y,x), x, y \in V$) and V-*elliptic*: there is a $c_0 > 0$ such that

$$a(x,x) \geq c_0 \|x\|_V^2 \ , \qquad x \in V \ . \tag{13}$$

Thus, $a(\cdot,\cdot)$ determines an *equivalent scalar product* on V: a sequence converges in V with $\|\cdot\|_V$ if and only if it converges with $a(\cdot,\cdot)^{1/2}$. Thus we may replace $(\cdot,\cdot)_V$ by $a(\cdot,\cdot)$ above.

THEOREM 2. *Let $a(\cdot,\cdot)$ be a bilinear, symmetric, continuous and V-elliptic form on the Hilbert space V, let K be a closed, convex and non-empty subset of V, and let $f \in V'$. Set $\phi(x) = (\frac{1}{2})a(x,x) - f(x)$, $x \in V$. Then there is a unique*

$$x \in K : \phi(x) \leq \phi(y) \ , \qquad y \in K \ . \tag{14}$$

The solution of (14) *is characterized by*

$$x \in K : a(x, y-x) \geq f(y-x) \ , \qquad y \in K \ . \tag{15}$$

If, in addition, K is a subspace of V, then (15) *is equivalent to*

$$x \in K : a(x,y) = f(y) \ , \qquad y \in K \ . \tag{16}$$

Now (16) is precisely our weak formulation, and we see it is the special case of a *variational inequality* (15) which is the characterization of the solution of the minimization problem (14).

2.5. Examples. Consider the vertical displacement $u(x)$ within a fixed plane of a string of length $\ell > 0$ whose initial position ($u = 0$) is the interval $0 \leq x \leq \ell$. The string is stretched with tension $T > 0$ and it is flexible and elastic, so this tension acts in the direction of the tangent. A vertical load or force $F(x)$ per unit length is applied and this results in the displacement $u(x)$ at each point x. For each segment $[x_1, x_2]$ the balance of vertical components of force gives

$$-T \sin\theta_{x_2} + T \sin\theta_{x_1} = \int_{x_1}^{x_2} F(x)\, dx$$

where $\sin\theta_x = u'(x)/\sqrt{1 + (u'(x))^2}$ is the vertical component of the unit tangent at $(x, u(x))$. We assume displacements are small, so $1 + (u')^2 \cong 1$

and we obtain

$$-T\big(u'(x_2) - u'(x_1)\big) = \int_{x_1}^{x_2} F(x)\,dx\ , \qquad 0 \le x_1 < x_1 \le \ell\ .$$

If F is locally integrable on $(0,\ell)$ it follows that $u' = \partial u$ is locally absolutely continuous and the equation

$$-T\partial^2 u = F \qquad \text{in } L^1_{loc}(0,\ell)$$

describes the displacement of the string in the interior of the interval $(0,\ell)$. At the end-points we need to separately prescribe the behavior. For example, at $x = 0$ we could specify either the position, $u(0) = c$, the vertical force, $T\partial u(0) = f_0$, or some combination of these such as an elastic restoring force of the form $T\partial u(0) = h(u(0) - c)$. Such conditions will be prescribed at each of the two boundary points of the interval.

To calculate the energy that is added to the string to move it to the position u, we take the product of the forces and displacements. These tangential and vertical changes are given respectively by

$$T\int_0^\ell \big(\sqrt{1+(u')^2} - 1\big)\,dx - \int_0^\ell F(x)u(x)\,dx$$

where the first term depends on the change in length of the string, and for small displacements this gives us the approximate *potential energy* functional

$$\varphi(u) = \int_0^\ell \big((T/2)(\partial u)^2 - Fu\big)\,dx\ .$$

Here we have used the expansion for small values of r

$$\sqrt{1+r^2} = 1 + \frac{1}{2}r^2 + \cdots\ .$$

We shall see the displacement u corresponding to the external load F can be obtained by minimizing (3.2) over the appropriate set of admissable displacements. Moreover, this applies to more general loads. For example, a "point load" of magnitude $F_0 > 0$ applied at the point c, $0 < c < \ell$, leads to the energy functional

$$\varphi(u) = T/2\int_0^\ell (\partial u)^2\,dx - F_0 u(c)$$

in which the second term is just a Dirac functional concentrated at c.

Next we give a set of boundary-value problems on the interval (a, b). Each is guaranteed to have a unique solution by Theorem ??. Each example will be related to a stretched-string problem, and for certain special cases we shall compute the displacement u to see if it appears to be consistent with the physical problem. In both of these examples, we rescale the load so that we may assume $T = 1$. Thus, the load becomes the ratio of the actual load to the tension.

EXAMPLE 5. *The displacement of a string fixed at both ends is given by the solution of*

$$u \in H_0^1(0,\ell) : -\partial^2 u = f$$

where $f \in H_0^1(0,\ell)'$. This problem is well-posed by Theorem ??; note that the corresponding bilinear form is H_0^1-elliptic by the estimate (1.4). If we apply a load $F(x) = sgn(x - \ell/2)$, where the sign function is given by $sgn(x) = x/|x|$, $x \neq 0$, the resulting displacement is

$$u(x) = \begin{cases} -\frac{1}{2}x(\ell/2 - x) \ , & 0 < x < \ell/2 \\ \frac{1}{2}(x - \ell/2)(\ell - x) \ , & \ell/2 < x < \ell \end{cases}$$

If we apply a point load, δ_c concentrated at $x = c$, the displacement is

$$u(x) = 1/2(c - |x - c|) + (1/2 - c/\ell)x$$

with maximum value $u(c) = c(1-c/\ell)$. Both of these solutions can be computed directly from the ordinary differential equation by using Proposition 1.2.

EXAMPLE 6. *Non-homogeneous boundary conditions arise when the displacements at the end-points are fixed at non-zero levels. For example, the solution to*

$$u \in H^1(0,\ell) : u(0) = f_1 \ , \quad u(\ell) = f_2 \ , \quad -\partial^2 u = F$$

is obtained by minimizing (3.2) over the set of admissable displacements

$$K = \{v \in H^1(0,\ell) : v(0) = f_1, v(\ell) = f_2\} \ .$$

This minimum u satisfies (2.9) where

$$a(u,v) = \int_0^\ell \partial u \partial v \, dx \ , \qquad f(v) = \int_0^\ell F v \, dx \qquad u, v \in H^1(0,\ell) \ .$$

Since the set K is the translate of the subspace $H_0^1(0,\ell)$ by the function

$$u_0(x) = (\ell - x)f_1/\ell + xf_2/\ell \ ,$$

this variational inequality is equivalent to

$$u \in K : a(u,\varphi) = f(\varphi) \ , \qquad \varphi \in H_0^1(0,\ell) \ .$$

Moreover, this problem is actually a "linear" problem for the unkown $w \equiv u - u_0$ in the form

$$w \in H_0^1(0,\ell) : a(w,\varphi) = f(\varphi) - a(u_0,\varphi) \ , \qquad \varphi \in H_0^1(0,\ell) \ ,$$

and thus it is well-posed by any one of the Theorems of Section 2.

Exercise 9. *Let the function $F(\cdot) \in L^2(0,\ell)$ and the numbers $\lambda \geq 0$, u_o, $a \geq 0$ and $g \in \mathbb{R}$ be given.*

(a). Show that the function $u(\cdot) \in H^1(0,\ell)$ is a solution of the problem

$$-\partial^2 u(x) + \lambda u(x) = F(x), \quad 0 < x < \ell,$$
$$u(0) = u_0, \quad \partial u(\ell) + au(\ell) = g \, .$$

if and only if it satisfies

$$u \in H^1(0,\ell) : \ u(0) = u_0, \ \text{and}$$
$$\int_0^\ell (\partial u(x)\partial v(x) + \lambda u(x)v(x))\,dx + au(\ell)v(\ell)$$
$$= \int_0^\ell F(x)v(x)\,dx + gv(\ell) \ \text{for all} \ v \in H^1(0,\ell) : \ v(0) = 0 \, .$$

(b). Show that Theorem 2 applies, and specify V, K, $a(\cdot,\cdot)$, $f(\cdot)$.

2.6. Adjoint Operator. Let V and W be Hilbert spaces and $T \in \mathcal{L}(V,W)$. We define the *adjoint* of T as follows: if $u \in W$, then the functional $v \mapsto (u,Tv)_W$ belongs to V', so Theorem 5 shows that there is a unique $T^*u \in V$ such that

$$(T^*u, v)_V = (u, Tv)_W \ , \qquad u \in W \ , \ v \in V \ .$$

Exercise 10. *If V and W are Hilbert spaces and $T \in \mathcal{L}(V,W)$, then $T^* \in \mathcal{L}(W,V)$, $Rg(T)^\perp = K(T^*)$ and $Rg(T^*)^\perp = K(T)$. If T is an isomorphism with $T^{-1} \in \mathcal{L}(W,V)$, then T^* is an isomorphism and $(T^*)^{-1} = (T^{-1})^*$.*

3. Approximation of Solutions

The weak formulation of boundary-value problems is given by (16), and this is precisely a special case of the *variational inequality* (15) which characterizes the solution of the minimization problem (14). Here we shall discuss the Rayleigh-Ritz-Galerkin procedure for approximating the solution by means of a finite-dimensional problem that can be computed very efficiently. This provides an example of a *finite element method.*

Suppose that we are in the situation of Theorem 2, and let u be the solution of the variational equation

$$u \in V : a(u, v) = f(v) \ , \qquad v \in V \ . \tag{17}$$

Let S be a subspace of V. Then Theorem 2 asserts likewise the existence of a unique solution of the problem

$$u_S \in S : a(u_S, v) = f(v) \ , \qquad v \in S \ . \tag{18}$$

We shall first show that the size of the *error* $u - u_S$ depends on how well the subspace S approximates V.

Theorem 3. *Let $a(\cdot,\cdot)$ be a bilinear, symmetric, continuous and V-elliptic form on the Hilbert space V, let S be a closed subspace of V, and let $f \in V'$. Then* (17) *has a unique solution u and* (18) *has a unique solution u_S, and these satisfy the estimate*

$$\|u - u_S\| \leq \frac{C}{c_0} \inf\{\|u - v\| : \ v \in S\}, \tag{19}$$

where C is the continuity constant of $a(\cdot,\cdot)$ and c_0 is the V-elliptic constant.

Proof. We need only to verify the estimate (19). By subtracting the identities (17) and (18) we obtain

$$a(u_S - u, v) = 0 \ , \qquad v \in S \ . \tag{20}$$

Then for any $w \in S$ we have

$$a(u_S - u, u_S - u) = a(u_S - u, w - u) + a(u_S - u, u_S - w), \tag{21}$$

and $u_S - w \in S$, so the last term is zero and we obtain

$$c_0\|u_S - u\|^2 \leq C\|u_s - u\|\|w - u\|, \qquad w \in S,$$

and this gives the estimate (19). $\square$

The right side of equation (19) is determined by the *best approximation* of V by S, namely, $\inf\{\|u-v\|: \ v \in S\} = \|u - P_S(u)\|$ where P_S is the *projection* onto S. In terms of the norm determined by the scalar-product $a(\cdot,\cdot)$ on V, $\|\cdot\|_a = (a(\cdot,\cdot))^{1/2}$, we have precisely

$$\|u - u_S\|_a = \inf\{\|u-v\|_a : \ v \in S\}.$$

Consider the case of a *separable* space V. That is, there is a sequence $\{v_1, v_2, v_3, \dots\}$ in V which is a basis for V. For each integer $m \geq 1$, the set $\{v_1, v_2, \dots, v_m\}$ is linearly independent and its linear span will be denoted by V_m. If P_m is the projection of V onto V_m, then $lim_{m\to\infty} P_m v = v$ for each $v \in V$. The approximate problem (18) with $S = V_m$ is equivalent to

$$u_m \in V_m : a(u_m, v_j) = f(v_j)\ , \qquad 1 \leq j \leq m\ . \tag{22}$$

For each integer m there is exactly one such u_m, and we have the estimate $\|u_m - u\| \leq \frac{C}{c_0}\|P_m u - u\|$, so $lim_{m\to\infty} u_m = u$ in V. Moreover, we can write u_m as a linear combination of basis elements, $u_m = \sum_{i=1}^m x_i v_i$ and then (22) is equivalent to the $m \times m$ algebraic system

$$\sum_{i=1}^m a(v_i, v_j) x_i = f(v_j)\ , \qquad 1 \leq j \leq m\ . \tag{23}$$

This linear system with the matrix $a(v_i, v_j)$ is invertible and can be solved to obtain the coefficients, x_i, and thereby the approximate solution u_m.

Now this linear system (23) will be large if we want our approximation u_m to be close to the solution u, and so we need to keep the solution procedure as easy as possible. In particular, we may want to permit m to be of the order of 10^2. Two fundamental approaches include the *eigenfunction expansion* method, which we consider in the following section, and the *finite element method* which we discuss now.

Consider the Dirichlet problem of Example 6,

$$u \in H_0^1(0,\ell) : \quad \int_0^\ell \partial u \partial v \, dx = \int_0^\ell F v \, dx\ , \quad v \in H_0^1(0,\ell)\ ,$$

where we have set $V = H_0^1(0,\ell)$ and

$$a(u,v) = \int_0^\ell \partial u \partial v \, dx\ , \qquad f(v) = \int_0^\ell F v \, dx \qquad u, v \in H_0^1(0,\ell)\ .$$

The Approximation. We construct a finite-dimensional subspace of V as follows. Let $m \geq 1$ be an integer and define $h > 0$ by $h(m+1) = \ell$. Partition the interval $(0, \ell)$ with nodes $x_j = jh,\ 1 \leq j \leq m$, and set $x_0 = 0,\ x_{m+1} = \ell$. Now we define the space V_h to be those continuous functions on the interval $[0, \ell]$ which are affine on each subinterval $[x_{j-1}, x_j], 1 \leq j \leq m+1$, and which vanish at the end points, $x_0 = 0,\ x_{m+1} = \ell$. These are *piecewise linear*, and since their derivatives are piecewise constant, V_h is a subspace of $H_0^1(0, \ell)$.

In order to parametrize the elements of V_h, we introduce a *basis*. For each integer j, $1 \leq j \leq m$, define $l_j(\cdot) \in V_h$ by

$$l_j(x_i) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

That is, $l_j(\cdot)$ is the continuous and piecewise affine function on $(0, \ell)$ which takes on the value 1 at the node x_j and the value 0 at all other nodes. Each function $v_h(\cdot) \in V_h$ has the unique representation

$$v_h(x) = \sum_{j=1}^{m} v_j l_j(x), \quad 0 < x < \ell,$$

where the coefficients are given by the nodal values $v_j = v_h(x_j),\ 1 \leq j \leq m$. In particular, the subspace V_h has dimension m. The corresponding *finite element method* for the Dirichlet boundary-value problem consists of the approximating problem (18) with $S = V_h$, and this is equivalent to finding

$$\tilde{u_h} \in V_h : a(\tilde{u_h}, l_j) = f(l_j)\ , \qquad 1 \leq j \leq m\ .$$

In terms of the basis, the solution is given by

$$\tilde{u_h}(\cdot) = \sum_{i=1}^{m} u_i l_i(\cdot) \in V_h$$

where the coefficients satisfy the algebraic system

$$\sum_{j=1}^{m} a(l_i, l_j) u_i = f(l_j)\ , \qquad 1 \leq j \leq m, \tag{24}$$

consisting of m equations in m unknowns. This is a matrix equation of the form

$$\vec{u} \in \mathbb{R}^m : A\vec{u} = \vec{f} \in \mathbb{R}^m,$$

where the matrix has elements $a_{ij} = a(l_i, l_j)$ and the right side has components $f_j = f(l_j)$.

The elements of the matrix A are easily computed. First, note that since the support of $l_j(\cdot)$ is the interval $[x_{j-1}, x_{j+1}]$ we have $a(l_i, l_j) = 0$ whenever $|i - j| \geq 2$. That is, the matrix is *tridiagonal*. Furthermore, we have $a(l_i, l_i) = \frac{2}{h}$ and $a(l_j, l_{j-1}) = -\frac{1}{h}$. Note that this is precisely the matrix and corresponding algebraic problem obtained in the *discrete models* with Dirichlet boundary conditions. This shows that those discrete models are essentially projection approximations onto appropriate subspaces of the corresponding continuum models.

Interpolation. Now we investigate the size of the *best approximation* constants for the subspace V_h of V. For each function $v \in V = H_0^1(0, \ell)$ we choose the *piecewise linear interpolant* of v to be that $v_h \in V_h$ for which $v_h(x_j) = v(x_j)$ for each node x_j, $1 \leq j \leq m$. By considering the difference, $v(\cdot) - v_h(\cdot)$, we are led to the subspace of V given by

$$V_0 = \{v \in V : \ v(x_j) = 0, \ 1 \leq j \leq m\}.$$

From the estimate (6) applied to each subinterval, $[x_{j-1}, x_j]$, we obtain

$$\|v_0\|_{L^2(0,\ell)} \leq \big(h/\sqrt{2}\big) \|\partial v_0\|_{L^2(0,\ell)}, \quad v_0 \in V_0 . \tag{25}$$

Let $v \in V = H_0^1(0, \ell)$ be given. Then we can use Theorem 2 to show there exists a unique

$$v_0 \in V_0 : \ (\partial(v_0 - v), \partial\phi)_{L^2(0,\ell)} = 0, \ \text{for all } \phi \in V_0 .$$

Then the difference $v_h \equiv v - v_0$ satisfies

$$v - v_h \in V_0 : \ (\partial v_h, \partial\phi)_{L^2(0,\ell)} = 0, \ \text{for all } \phi \in V_0 . \tag{26}$$

By choosing $\phi \in C_0^\infty(x_{j-1}, x_j)$, we see that $-\partial^2 v_h = 0$ on each subinterval (x_{j-1}, x_j), and then with $v - v_h \in V_0$ it follows that v_h is precisely the piecewise linear interpolant of v. This computation is reversible, so we find that (26) characterizes this interpolant. Thus, we have established the orthogonal decomposition $V = V_0 \oplus V_h$ with respect to the scalar product $a(\cdot,\cdot)$, in which each $v \in V$ is written as $v = v_0 + v_h$ as above. In particular, we have

$$\|\partial v_h\|^2_{L^2(0,\ell)} + \|\partial(v - v_h)\|^2_{L^2(0,\ell)} = \|\partial v\|^2_{L^2(0,\ell)}$$

because of the orthogonality of v_h and $v - v_h$. This leads to

$$\|\partial(v - v_h)\|_{L^2(0,\ell)} \leq \|\partial v\|_{L^2(0,\ell)}\,, \tag{27}$$

and as a consequence from (25) with $v_0 = v - v_h$ we obtain

$$\|v - v_h\|_{L^2(0,\ell)} \leq \big(h/\sqrt{2}\big)\|\partial v\|_{L^2(0,\ell)}\,, \quad v \in V\,. \tag{28}$$

That is, for a function $v \in V$, the piecewise linear interpolation gives approximation with error estimate that is *stable* for the derivative and of first order in h for the value. This implies that the approximate solution $\tilde{u_h}$ of the Dirichlet problem converges to the solution u, and it even shows the *rate* of convergence with respect to the mesh size, h, of the partition. In particular, it justifies the expectation that the computed $\tilde{u_h}$ really is close to the desired solution, u. Moreover, it shows that the discrete models are *close* to the corresponding continuum models.

We have shown that the *interpolation error* $v_h - v$ is of order h when $v \in V$. Next we obtain a better estimate when v is smoother. Specifically, let's assume that $v \in V \cap H^2(0,\ell)$. That is, $v \in V$ has *two* derivatives in $L^2(0,\ell)$. Then we compute explicitly

$$\|\partial(v - v_h)\|^2_{L^2(0,\ell)} = \sum_{j=1}^{m+1} \int_{x_{j-1}}^{x_j} \partial(v - v_h)\partial(v - v_h)\,dx =$$
$$-\sum_{j=1}^{m+1} \int_{x_{j-1}}^{x_j} \partial^2(v - v_h)(v - v_h)\,dx \leq \|\partial^2 v\|_{L^2(0,\ell)}\|v - v_h\|_{L^2(0,\ell)}.$$

(We have use the fact that $\partial^2 v_h = 0$ on each subinterval (x_{j-1}, x_j).) Together with (25), this gives

$$\|\partial(v - v_h)\|^2_{L^2(0,\ell)} \leq \|\partial^2 v\|_{L^2(0,\ell)}\big(h/\sqrt{2}\big)\|\partial(v - v_h)\|_{L^2(0,\ell)}\,,$$

so we obtain

$$\|\partial(v - v_h)\|_{L^2(0,\ell)} \leq \|\partial^2 v\|^2_{L^2(0,\ell)}\big(h/\sqrt{2}\big)\,. \tag{29}$$

Finally, combining this with (25) again, we get

$$\|v - v_h\|_{L^2(0,\ell)} \leq \|\partial^2 v\|^2_{L^2(0,\ell)} \big(h^2/2\big)\,. \tag{30}$$

Thus, for a smoother function $v \in V \cap H^2(0,\ell)$, the piecewise linear interpolation gives an approximation error estimate of first order for the derivative and of second order for the value.

Second-order estimates. Let u be the solution of our Dirichlet problem and let $\tilde{u_h}$ be its V_h-approximation above. Let $g = \|u - \tilde{u_h}\|^{-1}_{L^2(0,\ell)}(u - \tilde{u_h})$ be the indicated *normalized* function, and then use Theorem 2 to find

$$w \in V :\ a(w,v) = (g,v)_{L^2}, \quad v \in V\,.$$

Then we set $v = u - \tilde{u_h}$ to obtain $a(u - \tilde{u_h}, w) = \|u - \tilde{u_h}\|_{L^2}$, and so for each $v \in V_h$ we obtain

$$\|u - \tilde{u_h}\|_{L^2} = a(u - \tilde{u_h}, w - v) \leq \|u - \tilde{u_h}\|_a \|w - v\|_a\,.$$

But $u \in H^2(0,\ell)$ and $w \in H^2(0,\ell)$, so if we choose v to be the interpolant of w, that is, $v = w_h$, then we obtain

$$\begin{aligned}\|u - \tilde{u_h}\|_{L^2} &\leq \|\partial(u - \tilde{u_h})\|_{L^2} \|\partial(w - w_h)\|_{L^2} \\ &\qquad \leq \|\partial(u - \tilde{u_h})\|_{L^2} \| \|\partial^2 w\|^2_{L^2(0,\ell)}\big(h/\sqrt{2}\big) \\ &\leq \|\partial^2 u\|^2_{L^2(0,\ell)}\big(h/\sqrt{2}\big)\|g\|^2_{L^2(0,\ell)}\big(h/\sqrt{2}\big) = \|F\|^2_{L^2(0,\ell)}\big(h^2/2\big)\end{aligned}$$

That is, the approximation error $\|u - \tilde{u_h}\|_{L^2}$ in this finite element method is of the same order as the best approximation error of the corresponding subspace, V_h. This calculation depended on the fact that the solution of the boundary-value problem is in $H^2(0,\ell)$ when the data F is in $L^2(0,\ell)$. This is a typical *regularizing property* for elliptic problems.

4. Expansion in Eigenfunctions

4.1. Fourier Series. We consider the Fourier series of a vector in the scalar product space H with respect to a given set of orthogonal vectors. The sequence $\{v_j\}$ of vectors in H is called *orthogonal* if $(v_i, v_j)_H = 0$ for each pair i, j with $i \neq j$. Let $\{v_j\}$ be such a sequence of non-zero vectors and let $u \in H$. For each j we define the *Fourier coefficient* of u with respect to v_j by $c_j = (u, v_j)_H/(v_j, v_j)_H$. For each $n \geq 1$ it follows that $\sum_{j=1}^n c_j v_j$ is the projection of u on the subspace M_n spanned by $\{v_1, v_2, \ldots, v_n\}$. This follows from Corollary 4 by noting that $u - \sum_{j=1}^n c_j v_j$

is orthogonal to each v_i, $1 \le j \le n$, hence belongs to $M_n^{\perp}$. We call the sequence of vectors *orthonormal* if they are orthogonal and if $(v_j, v_j)_H = 1$ for each $j \ge 1$.

PROPOSITION 5. *Let $\{v_j\}$ be an orthonormal sequence in the scalar product space H and let $u \in H$. The Fourier coefficients of u are given by $c_j = (u, v_j)_H$ and satisfy*

$$\sum_{j=1}^{\infty} |c_j|^2 \le \|u\|^2 \ . \tag{31}$$

Also we have $u = \sum_{j=1}^{\infty} c_j v_j$ if and only if equality holds in (31).

PROOF. Let $u_n \equiv \sum_{j=1}^{n} c_j v_j$, $n \ge 1$. Then $u - u_n \perp u_n$ so we obtain

$$\|u\|^2 = \|u - u_n\|^2 + \|u_n\|^2 \ , \qquad n \ge 1 \ . \tag{32}$$

But $\|u_n\|^2 = \sum_{j=1}^{n} |c_j|^2$ follows since the set $\{v_i, \ldots, v_n\}$ is orthonormal, so we obtain $\sum_{j=1}^{n} |c_j|^2 \le \|u\|^2$ for all n, hence (31) holds. It follows from (32) that $\lim_{n\to\infty} \|u - u_n\| - 0$ if and only if equality holds in (31). □

The inequality (31) is *Bessel's inequality* and the corresponding equality is called *Parseval's equation*. The series $\sum_{j=1}^{\infty} c_j v_j$ above is the *Fourier series* of u with respect to the orthonormal sequence $\{v_j\}$.

PROPOSITION 6. *Let $\{v_j\}$ be an orthonormal sequence in the scalar product space H. Then every element of H equals the sum of its Fourier series if and only if $\{v_j\}$ is a* basis *for H, that is, its linear span is dense in H.*

PROOF. Suppose $\{v_j\}$ is a basis and let $u \in H$ be given. For any $\varepsilon > 0$, there is an $n \ge 1$ for which the linear span M_n of the set $\{v_1, v_2, \ldots, v_n\}$ contains an element which approximates u within ε. That is, $\inf\{\|u-w\| : w \in M_n\} < \varepsilon$. If u_n is given as in the proof of Proposition 5, then we have $u - u_n \in M_n^{\perp}$. Hence, for any $w \in M_n$ we have

$$\|u - u_n\|^2 = (u - u_n, u - w)_H \le \|u - u_n\| \, \|u - w\| \ ,$$

since $u_n - w \in M_n$. Taking the infimum over all $w \in M_n$ then gives

$$\|u - u_n\| \le \inf\{\|u - w\| : w \in M_n\} < \varepsilon \ . \tag{33}$$

Thus, $\lim_{n\to\infty} u_n = u$. The converse is clear. □

4.2. Eigenvalue Problem. Let $T \in \mathcal{L}(H)$. A non-zero vector $v \in H$ is called an *eigenvector* of T if $T(v) = \lambda v$ for some $\lambda \in \mathbb{R}$. The number λ is the *eigenvalue* of T corresponding to v. We shall show that certain operators possess a rich supply of eigenvectors. These eigenvectors form an orthonormal basis to which we can apply the preceding Fourier series expansion techniques.

Definition 2. *An operator $T \in \mathcal{L}(H)$ is called self-adjoint if $(Tu, v)_H = (u, Tv)_H$ for all $u, v \in H$. A self-adjoint T is called non-negative if $(Tu, u)_H \geq 0$ for all $u \in H$.*

Lemma 3. *If $T \in \mathcal{L}(H)$ is non-negative self-adjoint, then $\|Tu\| \leq \|T\|^{1/2}(Tu, u)_H^{1/2}$, $u \in H$.*

Proof. The bilinear form $[u, v] \equiv (Tu, v)_H$ satisfies the first two scalar-product axioms and this is sufficient to obtain

$$|[u, v]|^2 \leq [u, u][v, v]\ , \qquad u, v \in H\ . \tag{34}$$

(If either factor on the right side is strictly positive, this follows from the proof of (3). Otherwise, $0 \leq [u + tv, u + tv] = 2t[u, v]$ for all $t \in \mathbb{R}$, hence, both sides of (34) are zero.) The desired result follows by setting $v = T(u)$ in (34). ☐

The operators we shall consider are the *compact* operators.

Definition 3. *If V, W are normed linear spaces, then $T \in \mathcal{L}(V, W)$ is called* compact *if for any bounded sequence $\{u_n\}$ in V its image $\{Tu_n\}$ has a subsequence which converges in W.*

Exercise 11. *The composition of a continuous operator with a compact operator is compact.*

The essential fact we need is the following.

Lemma 4. *If $T \in \mathcal{L}(H)$ is self-adjoint and compact, then there exists a vector v with $\|v\| = 1$ and $T(v) = \mu v$, where $|\mu| = \|T\|_{\mathcal{L}(H)} > 0$.*

Proof. If λ is defined to be $\|T\|_{\mathcal{L}(H)}$, it follows from Proposition 4 that there is a sequence u_n in H with $\|u_n\| = 1$ and $\lim_{n\to\infty} \|Tu_n\| = \lambda$. Then $((\lambda^2 - T^2)u_n, u_n)_H = \lambda^2 - \|Tu_n\|^2$ converges to zero. The operator $\lambda^2 - T^2$ is non-negative self-adjoint so Lemma 3 implies $\{(\lambda^2 - T^2)u_n\}$

converges to zero. Since T is compact we may replace $\{u_n\}$ by an appropriate subsequence for which $\{Tu_n\}$ converges to some vector $w \in H$. Since T is continuous there follows $\lim_{n\to\infty}(\lambda^2 u_n) = \lim_{n\to\infty} T^2 u_n = Tw$, so $w = \lim_{n\to\infty} Tu_n = \lambda^{-2}T^2(w)$. Note that $\|w\| = \lambda$ and $T^2(w) = \lambda^2 w$. Thus, either $(\lambda + T)w \neq 0$ and we can choose $v = (\lambda + T)w/\|(\lambda + T)w\|$, or $(\lambda + T)w = 0$, and we can then choose $v = w/\|w\|$. Either way, the desired result follows. □

THEOREM 4. *Let H be a scalar product space and let $T \in \mathcal{L}(H)$ be self-adjoint and compact. Then there is an orthonormal sequence $\{v_j\}$ of eigenvectors of T for which the corresponding sequence of eigenvalues $\{\mu_j\}$ converges to zero and the eigenvectors are a basis for the* closure *of $Rg(T)$, $\overline{Rg(T)}$, that is, all the limit points of $Rg(T)$.*

PROOF. By Lemma 4 it follows that there is a vector v_1 with $\|v_1\| = 1$ and $T(v_1) = \mu_1 v_1$ with $|\mu_1| = \|T\|_{\mathcal{L}(H)}$. Set $H_1 = \{v_1\}^\perp$ and note $T\{H_1\} \subset H_1$. Thus, the restriction $T|_{H_1}$ is self-adjoint and compact so Lemma 4 implies the existence of an eigenvector v_2 of T of unit length in H_1 with eigenvalue μ_2 satisfying $|\mu_2| = \|T\|_{\mathcal{L}(H_1)} \leq |\mu_1|$. Set $H_2 = \{v_1, v_2\}^\perp$ and continue this procedure to obtain an orthonormal sequence $\{v_j\}$ in H and sequence $\{\mu_j\}$ in $\mathbb{R}$ such that $T(v_j) = \mu_j v_j$ and $|\mu_{j+1}| \leq |\mu_j|$ for $j \geq 1$.

We claim that $\lim_{j\to\infty}(\mu_j) = 0$. Otherwise, since $|\mu_j|$ is decreasing we would have all $|\mu_j| \geq \varepsilon$ for some $\varepsilon > 0$. But then

$$\|T(v_i) - T(v_j)\|^2 = \|\mu_i v_i - \mu_j v_j\|^2 = \|\mu_i v_i\|^2 + \|\mu_j v_j\|^2 \geq 2\varepsilon^2$$

for all $i \neq j$, so $\{T(v_j)\}$ has no convergent subsequence, a contradiction. We shall show $\{v_j\}$ is a basis for $Rg(T)$. Let $w \in Rg(T)$ and $\sum b_j v_j$ the Fourier series of w. Then there is a $u \in H$ with $T(u) = w$ and we let $\sum c_j v_j$ be the Fourier series of u. The coefficients are related by

$$b_j = (w, v_j)_H = (Tu, v_j)_H = (u, Tv_j)_H = \mu_j c_j \ ,$$

so there follows $T(c_j v_j) = b_j v_j$, hence,

$$w - \sum_{j=1}^{n} b_j v_j = T\left(u - \sum_{j=1}^{n} c_j v_j\right) , \qquad n \geq 1 \ . \tag{35}$$

Since T is bounded by $|\mu_{n+1}|$ on H_n, and since $\|u - \sum_{j=1}^{n} c_j v_j\| \le \|u\|$ by (32), we obtain from (35) the estimate

$$\left\| w - \sum_{j=1}^{n} b_j v_j \right\| \le |\mu_{n+1}| \cdot \|u\| \ , \qquad n \ge 1 \ . \tag{36}$$

Since $\lim_{j\to\infty} \mu_j = 0$, we have $w = \sum_{j=1}^{\infty} b_j v_j$ as desired.

It remains to show the eigenvectors form a basis for the *closure* of $Rg(T)$. So let w belong to this closure, and set $M_n = \langle v_1,\, v_2, \ldots v_n \rangle$, the linear span of the first n eigenvectors. As before, we let $w_n = \sum_{j=1}^{n} (w, v_j)_H v_j$ be the partial Fourier expansion of w in H For any $w_0 \in H$ we have $\|w - w_n\| = dist(w, M_n) \le \|w - w_0\| + dist(w_0, M_n)$. Now choose $w_0 \in Rg(T)$ and close to w, then choose n so large that the second term is small by Proposition 6, and we see that $w = \lim_{n\to\infty} w_n$ as desired. $\square$

Corollary 6. *Assume that H is complete, that is, a Hilbert space. If $w = T(u) \in Rg(T)$ and $u_n = \sum_{j=1}^{n} c_j v_j$ is the partial Fourier expansion of u as above, then $u_n \to u^*$ and $T(u_n) \to w$ for a vector $u^* \in H$ with $T(u^*) = w$.*

This is a *stronger* statement: the sequence $[u_n, w_n]$ converges to $[u, w]$ in the *graph* of T in the product space $H \times H$.

When $Rg(T)$ is finite-dimensional, we get additional structure. Suppose the sequence $\{\mu_j\}$ is eventually zero; let n be the first integer for which $\mu_n = 0$. Then $H_{n-1} \subset K(T)$, since $T(v_j) = 0$ for $j \ge n$. Also we see $v_j \in Rg(T)$ for $j < n$, so $Rg(T)^\perp \subset \{v_1, v_2, \ldots, v_{n-1}\}^\perp = H_{n-1}$ and from Exercise 10 follows $K(T) = Rg(T)^\perp \subset H_{n-1}$. Therefore $K(T) = H_{n-1}$ and $Rg(T)$ equals the linear span of $\{v_1, v_2, \ldots, v_{n-1}\}$.

4.3. Eigenvalues for Boundary-Value Problems.

4.3.1. *The boundary-value problems.* We recall three examples of the weak formulation of boundary-value problems.

Example 7. *Let $F(\cdot) \in L^2(0, \ell)$ and $c \in \mathbb{R}$ be given; define the bilinear form*

$$a(u, v) = \int_0^\ell (\partial u\, \partial v + c\, u\, v)\, dx, \quad u,\ v \in V\ ,$$

and the linear functional

$$f(v) = \int_0^\ell F\,v\,dx, \quad v \in V\,,$$

on the Hilbert space $V = H^1(0,\ell)$*. Then the variational equation*

$$u \in V: \quad a(u,v) = f(v) \text{ for all } v \in V, \tag{37}$$

is equivalent to the Neumann problem

$$u \in H^2(0,\ell): \ -\partial^2 u + c\,u = F\,, \quad \partial u(0) = \partial u(\ell) = 0\,.$$

If the constant satisfies $c > 0$*, then the bilinear form* $a(\cdot,\cdot)$ *is* V*-elliptic, and Theorem 2 shows that there is exactly one solution of the problem.*

Example 8. *With the same bilinear form and functional as above, take the space* $V = H_0^1(0,\ell)$*. Then the variational statement* (37) *is equivalent to the* Dirichlet problem

$$u \in H^2(0,\ell): \ -\partial^2 u + c\,u = F\,, \quad u(0) = u(\ell) = 0\,.$$

From Theorem 2 it follows that this problem has exactly one solution if $c \geq 0$*.*

Example 9. *For our last example, we choose the space* $V = \{v \in H^1(0,\ell): \ v(0) = v(\ell)\}$*. Then the variational statement* (37) *is equivalent to the* periodic problem

$$u \in H^2(0,\ell): \ -\partial^2 u + c\,u = F\,, \quad u(0) = u(\ell), \ \partial u(0) = \partial u(\ell)\,.$$

If the constant satisfies $c > 0$*, then there is exactly one solution of the problem.*

In each of these examples we have constructed on the Hilbert space $H = L^2(0,\ell)$ an operator which represents the corresponding boundary-value problem.

4.3.2. *The Compactness.* Consider now the identity map $H^1(0,\ell) \to L^2(0,\ell)$. This map is certainly continuous, that is, the H^1 norm is stronger than the L^2 norm. But we can say much more. In fact we have the Hölder continuity estimate

$$|u(x) - u(y)| \leq |x-y|^{1/2} \|\partial u\|_{L^2(0,\ell)}\,, \quad u \in H^1(0,\ell), 0 \leq x, y \leq \ell\,,$$

and the *uniform estimate*

$$\max\{|u(x)| : 0 \le x \le \ell\} \le 2(\ell)^{1/2}\|\partial u\|_{L^2} + 2(\ell)^{-1/2}\|u\|_{L^2}\ , \quad u \in H^1(0,\ell)\ .$$

From these it follows that the unit ball in $H^1(0,\ell)$ is a set of functions that is *equicontinuous* and *uniformly bounded*. From the Ascoli-Arzela Theorem, we see that any sequence from this set has a uniformly convergent subsequence. That is, the identity map of $H^1(0,\ell) \to C[0,\ell]$ is *compact*, and from this it follows that the map $H^1(0,\ell) \to L^2(0,\ell)$ is likewise compact.

We describe the general situation as follows. Let V and H be Hilbert spaces with respective norms $\|\cdot\|_V$ and $|\cdot|_H$, and assume V is dense and continuously imbedded in H: $\|v\|_V \ge |v|_V$ for all $v \in V$. We identify H with its dual through the scalar product of H, so we also obtain the identifications $H = H' \subset V'$. Since V is dense in H, each $f \in H'$ is determined by its *restriction* to the dense subspace V.

Assume the bilinear form $a(\cdot,\cdot)$ is continuous on V, symmetric, and for some $c \in \mathbb{R}$, $a(\cdot,\cdot) + c(\cdot,\cdot)_H$ is V-elliptic: there is a $c_0 > 0$ for which

$$a(v,v) + c\,(v,v)_H \ge c_0\|v\|_V^2, \quad v \in V\ .$$

By taking $K = V$ in the situation of Theorem 2, we see that for each $F \in H$ there is a unique

$$u \in V : \quad a(u,v) + c\,(u,v)_H = (F,v)_H, \qquad v \in V\ .$$

4.3.3. *The strong operator.* We define a subspace of V by

$$D(A) = \{u \in V : \text{ for some } F \in H,\ a(u,v) = (F,v)_H,\ v \in V\}$$

and then define the operator $A : D(A) \to H$ by $Au = F$. Since V is dense in H, this holds for at most one $F \in H$. Thus, $Au = F$ means that $u \in V$, $F \in H$, and

$$a(u,v) = (F,v)_H, \qquad v \in V\ . \tag{38}$$

Furthermore, we have noted above that $A + cI : D(A) \to H$ is a one-to-one map *onto* H, and this is precisely the map which characterizes the boundary-value problems in the preceding examples. Note that the domain of A is a proper subspace of H and the symmetry of $a(\cdot,\cdot)$ gives

$$(Au, v)_H = (u, Av)_H, \quad u,\ v \in D(A),$$

so A is a symmetric. The operator A is *not* bounded on H. However, from the V-ellipticity condition we obtain

$$c_0\|v\|_V^2 \leq ((A + cI)v, v)_H \leq |(A + cI)v|_H |v|_H \,,$$

and since $|v|_H \leq \|v\|_V$ we can delete this factor to obtain

$$c_0\|v\|_V \leq |(A + cI)v|_H, \quad v \in D(A).$$

Thus, the inverse $(A + cI)^{-1}$ is bounded from H into V.

Finally, we note that $D(A)$ is dense in H. To see this, let $w \in D(A)^\perp$. Set $w = (A + cI)u$ to obtain

$$0 = (w, u)_H = ((A + cI)u, u)_H \geq c_0\|u\|_V^2,$$

so $u = 0$ and then $w = 0$. Thus, $D(A)^\perp = \{0\}$ and $D(A)$ is dense in H.

4.4. The Expansion Theorem. Each of the examples above is an instance of the following situation.

THEOREM 5. *Let the Hilbert spaces H and V be given with the dense and* compact *embedding $V \to H$ and the identification $H = H'$. Let $a(\cdot,\cdot)$ be a bilinear, symmetric and continuous form on the Hilbert space V and $c \in \mathbb{R}$ a number for which $a(\cdot,\cdot) + c\,(\cdot,\cdot)_H$ is V-elliptic: there is a $c_0 > 0$ such that*

$$a(v, v) + c\,(v, v)_H \geq c_0\|v\|_V^2 \,, \qquad v \in V \,. \tag{39}$$

Let the operator A be given by (38). *Then there is an H-orthonormal sequence $\{v_j\}$ of eigenvectors of A for which the corresponding sequence of eigenvalues is monotone and satisfies $\{\lambda_j \to \infty\}$, and the eigenvectors are a basis for H.*

PROOF. The operator $A + cI$ is one-to-one onto H. Define $T \equiv (A + cI)^{-1}$. Then T is continous from H into V. Since $V \to H$ is compact, T is a compact operator on H. Furthermore, we check that it is self-adjoint and non-negative: $(T(F), F)_H \geq 0$ for all $F \in H$. Let $\{v_j\}$ be the orthonormal sequence of eigenvectors of T for which the corresponding sequence of (necessarily positive) eigenvalues $\{\mu_j\}$ converges downward to zero. Then from $Tv_j = \mu_j v_j$ we obtain

$$Av_j = \lambda_j v_j, \qquad j \geq 1$$

where $\lambda_j \equiv \frac{1}{\mu_j} - c$. It remains to check that the sequence $\{v_j\}$ is a basis for all of H. But $D(A) = D(A + cI)$ is dense in H, and $Rg(T) = D(A + cI)$, so this is immediate. $\square$

4.4.1. *The eigenfunctions.* The eigenfunctions for the *Neumann problem* are the non-zero solutions of

$$u \in H^2(0, \ell) : \; -\partial^2 u = \lambda\, u\,, \quad \partial u(0) = \partial u(\ell) = 0\,.$$

We compute that these are given by

$$v_n(x) = \begin{cases} \sqrt{\frac{2}{\ell}}\; cos(\frac{n\pi x}{\ell}), & n \geq 1\,, \\ \frac{1}{\sqrt{\ell}}, \quad n = 0\,, \end{cases}$$

with the corresponding eigenvalues $\lambda_n = \left(\frac{n\pi}{\ell}\right)^2$, $n \geq 0$. It follows from above that these are are a basis for $L^2(0, \ell)$, so every such function is equal to the sum of its *cosine series.*

Similarly, the eigenfunctions for the *Dirichlet problem* are the non-zero solutions of

$$u \in H^2(0, \ell) : \; -\partial^2 u = \lambda\, u\,, \quad u(0) = u(\ell) = 0\,,$$

and these are the functions

$$v_n(x) = \sqrt{\frac{2}{\ell}}\; sin(\frac{n\pi x}{\ell}), \quad n \geq 1\,,$$

with the corresponding eigenvalues $\lambda_n = \left(\frac{n\pi}{\ell}\right)^2$. These are likewise a basis for $L^2(0, \ell)$, so every such function is equal to the sum of its *sine series.*

Finally, the eigenvalue problem for the *periodic problem* is

$$u \in H^2(0, \ell) : \; -\partial^2 u = \lambda\, u\,, \quad u(0) = u(\ell),\; \partial u(0) = \partial u(\ell)\,,$$

and the corresponding eigenfunctions consist of the functions

$$v_n(x) = \begin{cases} \sqrt{\frac{2}{\ell}}\; cos(\frac{2n\pi x}{\ell}) \text{ and } \sqrt{\frac{2}{\ell}}\; sin(\frac{2n\pi x}{\ell}),\; n \geq 1, \\ \frac{1}{\sqrt{\ell}},\; n = 0, \end{cases}$$

with the eigenvalues $\lambda_n = \left(\frac{2n\pi}{\ell}\right)^2$. Note that for $n \geq 1$ the multiplicity of each eigenvalue is *two*. That is, there are two corresponding eigenfunctions.

4.4.2. *Eigenvalue characterization of the subspaces.* We have obtained an orthonormal basis of eigenvectors of A in H:

$$Av_j = \lambda_j v_j, \qquad j \geq 1,$$

where $\lambda_j \to +\infty$ as $j \to +\infty$. For each $w \in H$ we have $\sum_{j=1}^{\infty}(w, v_j)_H^2 < +\infty$ and $w = \lim_{n\to\infty} w_n$ where $w_n = \sum_{j=1}^{n}(w, v_j)_H v_j$ is the partial Fourier expansion of w in H.

Recall that the bilinear form $[u, v] = a(u, v) + c\,(u, v)_H$ is a scalar product on V that is equivalent to the original scalar product on V. From the computation $[v_i, v_j] = (\lambda_j + c)(v_i, v_j)_H$ we find that the sequence $\{(\lambda_j + c)^{-\frac{1}{2}} v_j\}$ is orthonormal in V with the scalar product $[\cdot, \cdot]$. More generally, we find that if $w \in V$ then $[w, v_j] = (\lambda_j + c)(w, v_j)_H$, so $w_n = \sum_{j=1}^{n}[w, (\lambda_j + c)^{-\frac{1}{2}} v_j](\lambda_j + c)^{-\frac{1}{2}} v_j$ is also the partial Fourier expansion of w in V. From the orthogonality of the eigenvectors, we find that

$$\begin{aligned}[w_m - w_n, w_m - w_n] &= ((A + cI)(w_m - w_n), w_m - w_n)_H \\ &= \sum_{j=n+1}^{m} (\lambda_j + c)(w, v_j)_H^2\,,\end{aligned}$$

so the sequence $\{w_n\}$ is Cauchy in V if and only if $\sum_{j=1}^{+\infty}(\lambda_j + c)(w, v_j)_H^2 < \infty$. It follows from these remarks that $w \in V$ if and only if $\sum_{j=1}^{+\infty}(\lambda_j + c)(w, v_j)_H^2 < \infty$, and in this case we have $w = \lim_{n\to\infty} w_n$ in V.

It is easy to see that $((A + cI)u, (A + cI)v)_H$ is a scalar product on $D(A)$ with a norm that is stronger than that of V. As above, we find that the sequence $\{(\lambda_j + c)^{-1} v_j\}$ is orthonormal in $D(A)$ with respect to this scalar product, and that $w \in D(A)$ if and only if $\sum_{j=1}^{+\infty}(\lambda_j + c)^2(w, v_j)_H^2 < \infty$, and in this case we have $w = \lim_{n\to\infty} w_n$ in $D(A)$. We summarize these results in the following.

Corollary 7. *For each $w \in H$ we have $\sum_{j=1}^{\infty}(w, v_j)_H^2 < +\infty$ and $w = \lim_{n\to\infty} w_n$, where $w_n = \sum_{j=1}^{n}(w, v_j)_H v_j$ is the partial Fourier expansion of w in H.*

$w \in V$ if and only if $\sum_{j=1}^{+\infty}(\lambda_j + c)(w, v_j)_H^2 < \infty$, and in this case we have $w = \lim_{n\to\infty} w_n$ in V.

$w \in D(A)$ if and only if $\sum_{j=1}^{+\infty}(\lambda_j + c)^2(w, v_j)_H^2 < \infty$, and in this case we have $w = \lim_{n\to\infty} w_n$ in $D(A)$ and

$$Aw = \sum_{j=1}^{\infty} \lambda_j (w, v_j)_H \, v_j \,.$$

Similar results hold for functions with values in the various spaces. For example, for each $w(\cdot) \in C([0,T], H)$ we have $\sum_{j=1}^{\infty}(w(t), v_j)_H^2 \leq |w(t)|_H^2 \leq C < +\infty$ and $w(t) = \lim_{n\to\infty} w_n(t)$ for each $t \in [0,T]$, where $w_n(t) = \sum_{j=1}^{n}(w(t), v_j)_H v_j$ is the partial Fourier expansion of $w(t)$ in H. Furthermore, the sequence of *continuous* real-valued funcions $\|w(t) - w_n(t)\|^2 = \sum_{j=n+1}^{+\infty} |(w(t), v_j)_H|^2$ converges *monotonically* to zero, so by Dini's theorem, the convergence is *uniform* on $[0,T]$. That is, we have convergence $w_n(\cdot) \to w(\cdot)$ in $C([0,T], H)$ as $n \to +\infty$. If $w(\cdot) \in C^1([0,T], H)$, we find similarly that $w_n(\cdot) \to w(\cdot)$ in $C^1([0,T], H)$. Similar statements follow for functions taking values in the spaces V and $D(A)$.

Such results are fundamental for developing the expansion theory for our problems below. In particular, recall that in our examples above, convergence in V implies (uniform) convergence in $C[0,\ell]$. Similarly, we check that convergence in $D(A)$ implies convergence in $C^1[0,\ell]$.

4.5. Applications of the Expansion Theorem.

4.5.1. *The elliptic problem.* For given $\lambda \in \mathbb{R}$ and $F \in H$, describe the solutions of

$$Au = \lambda u + F \,. \tag{40}$$

The vector $u \in D(A)$ is a solution of (40) if and only if

$$\sum_{j=1}^{\infty} \lambda_j (u, v_j)_H \, v_j = \lambda \sum_{j=1}^{\infty} (u, v_j)_H \, v_j + \sum_{j=1}^{\infty} (F, v_j)_H \, v_j \,,$$

and this holds exactly when $(\lambda_j - \lambda)(u, v_j)_H = (F, v_j)_H$ for all $j \geq 1$. This observation yields the following.

COROLLARY 8. *If $\lambda \neq \lambda_j$ for all $j \geq 1$, then for every $F \in H$ there is exactly one solution of* (40), *and it is given by the series*

$$u = \sum_{j=1}^{\infty} \frac{(F, v_j)_H}{\lambda_j - \lambda} v_j$$

which converges in $D(A)$.

If $\lambda = \lambda_J$ for some $J \geq 1$, then a solution exists only if $(F, v_j)_H = 0$ for all j with $\lambda_j = \lambda_J$, and then a solution is given by

$$u = \sum_{\{j:\lambda_j \neq \lambda_J\}} \frac{(F, v_j)_H}{\lambda_j - \lambda} v_j + \sum_{\{j:\lambda_j = \lambda_J\}} c_j v_j$$

where the constants $\{c_j\}$ are arbitrary for each index in the finite set $\{j : \lambda_j = \lambda_J\}$.

EXERCISE 12. *Discuss the convergence of this series representation of the solution u.*

4.5.2. *The diffusion equation.* For given $u_0 \in H$, describe the solutions of the basic initial-value problem

$$\dot{u}(t) + Au(t) = 0\,, \quad u(0) = u_0\,. \tag{41}$$

If $u(\cdot) \in C([0, T], H)$ is a solution of (41) with $u(\cdot) \in C^1((0, T], D(A))$, then each coefficient $u_j(t) = (u(t), v_j)_H$ satisfies

$$\dot{u}_j(t) + \lambda_j u_j(t) = 0\,, \quad u_j(0) = (u_0, v_j)_H\,.$$

These are then given by

$$u_j(t) = e^{-\lambda_j t}(u_0, v_j)_H, \quad j \geq 1\,,$$

so the Fourier series for $u(t)$ converges in $C([0, T], H)$. But the convergence is much stronger because of the exponential coefficients. In fact, for any polynomial $P(\lambda)$ and $\varepsilon > 0$, the product $P(\lambda)e^{-\lambda t}$ is uniformly bounded on $\lambda \geq 0$ and converges to 0 for $\lambda \to +\infty$, uniformly for all $t \geq \varepsilon$.

From this it follows that the series for the derivative $\dot{u}(t)$ as well as that for $Au(t)$ converges in H for every $t > 0$. Moreover, much more is true and we summarize it as follows.

Corollary 9. *If* $u_0 \in H$*, then there is exactly one solution* $u(\cdot) \in C([0,T],H)$ *of* (41) *with* $\dot{u}(\cdot),\ Au(\cdot) \in C((0,T],H)$*, and it is given by the series*

$$u(t) = \sum_{j=1}^{\infty} e^{-\lambda_j t}(u_0, v_j)_H\, v_j\, .$$

Moreover, for any $\varepsilon > 0$*, every derivative of any power of* A *of this series converges uniformly in* $C([\varepsilon,T],H)$*, so* $u(t)$ *is an infinitely differentiable function of* $x \in (0,\ell),\ t > 0$.

Exercise 13. *Let* A *be as above, and discuss the solvability of the initial-value problem for the* pseudo-parabolic *equation,*

$$\dot{u}(t) + \varepsilon A\dot{u}(t) + Au(t) = 0\,, \quad u(0) = u_0\,,$$

in which $\varepsilon > 0$.

4.5.3. *The wave equation.* For given $g \in H$, describe the solutions of the basic initial-value problem

$$\ddot{w}(t) + Aw(t) = 0\,, \quad w(0) = 0, \quad \dot{w}(0) = g\,. \tag{42}$$

We have seen that this is the fundamental problem whose solution $w(t) = S(t)g$ defines the operators $S(\cdot)$ which yield the general solution of the initial-value problem for the corresponding non-homogeneous

$$\ddot{u}(t) + Au(t) = F(t)\,, \quad u(0) = u_0, \quad \dot{u}(0) = v_0\,.$$

The solution is then represented by

$$u(t) = S'(t)u_0 + S(t)v_0 + \int_0^t S(t-s)F(s)\,ds\,.$$

If $w(\cdot) \in C([0,T],D(A))$ is a solution of (42), then each coefficient $w_j(t) = (w(t),v_j)_H$ satisfies the equation

$$\ddot{w}_j(t) + \lambda_j w_j(t) = 0\,, \quad w_j(0) = 0, \quad \dot{w}_j(0) = (g,v_j)_H\,.$$

These are given by

$$w_j(t) = \lambda_j^{-\frac{1}{2}} sin(\lambda_j^{\frac{1}{2}} t)(g,v_j)_H$$

if $\lambda_j > 0$, by $w_j(t) = (g,v_j)_H\, t$ if $\lambda_j = 0$, and by the hyperbolic function

$$w_j(t) = (-\lambda_j)^{-\frac{1}{2}} sinh((-\lambda_j)^{\frac{1}{2}} t)(g,v_j)_H$$

if $\lambda_j < 0$. Note that all but a finite number of λ_j's will be strictly positive, so all convergence issues are determined by the terms with positive eigenvalues. Thus, with the coefficients given by the above, the Fourier series for $w(t)$ converges in $C([0,T],V)$, and the series for the derivative $\dot{w}(t)$ converges in $C([0,T],H)$.

If we require further that $g \in V$, so that its expansion has an extra factor of $\lambda_j^{-\frac{1}{2}}$, then it follows that the series for $w(t)$ converges in $C([0,T],D(A))$, the series for $\dot{w}(t)$ converges in $C([0,T],V)$, and the series for the second derivative $\ddot{w}(t)$ converges in $C([0,T],H)$ to give a solution as desired. Note that the solution is exactly as smooth as the initial condition allows, *i.e.*, the smoothness is preserved. We summarize this in the following.

Corollary 10. *If $g \in V$, then there is exactly one solution of (42) with $\ddot{w}(\cdot)$, $Aw(\cdot) \in C([0,T],H)$, and it is given by the series*

$$w(t) = \sum_{j=1}^{\infty} \lambda_j^{-\frac{1}{2}} sin(\lambda_j^{\frac{1}{2}} t)(g, v_j)_H \, v_j = S(t)g$$

if all eigenvalues are positive, and it is modified accordingly in a finite number of terms otherwise.

We obtain the solution of the wave equation with given initial value u_0,

$$\ddot{w}(t) + Aw(t) = 0\,, \quad w(0) = u_0, \quad \dot{w}(0) = 0\,. \tag{43}$$

by solving (42) with $g = u_0$ and then computing its time derivative to obtain

$$w(t) = \sum_{j=1}^{\infty} cos(\lambda_j^{\frac{1}{2}} t)(u_0, v_j)_H \, v_j = S'(t)u_0\,.$$

In order for this to have the same smoothness as above, we need to require that $u_0 \in D(A)$.

Exercise 14. *Find conditions on the function $F(t)$ for which the initial-value problem*

$$\ddot{u}(t) + Au(t) = F(t)\,, \quad u(0) = 0\,, \quad \dot{u}(0) = 0\,,$$

has a solution.

EXERCISE 15. *Let A be as above, and discuss the solvability of the initial-value problem for the wave equation with transverse inertia*

$$\ddot{u}(t) + \varepsilon A\ddot{u}(t) + Au(t) = 0\,, \quad u(0) = u_0\,, \quad \dot{u}(0) = v_0\,.$$

EXERCISE 16. *Let A be as above, and discuss the solvability of the initial-value problem for the* strongly-damped *wave equation*

$$\ddot{u}(t) + \varepsilon A\dot{u}(t) + Au(t) = 0\,, \quad u(0) = u_0, \quad \dot{u}(0) = v_0\,.$$