

# Chapter 1 Foundations of Elliptic Boundary Value Problems

## 1.1 Euler equations of variational problems

Elliptic boundary value problems often occur as the Euler equations of variational problems the latter representing the optimality conditions of minimization problems.

As a simple example let us consider the computation of the **stationary equilibrium of a clamped membrane** (cf. Figure 1).

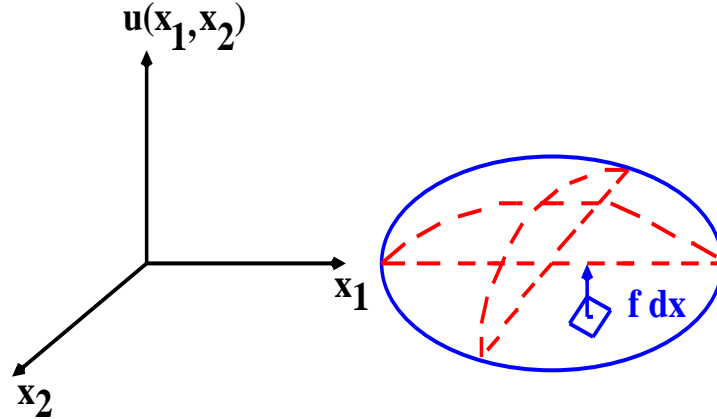


Figure 1: Deflection of a clamped membrane

A membrane is a surfacic, in its ground state plane, elastic body, whose potential energy is directly proportional to the change of the surface area. Thus, the ground state can be described by a bounded domain  $\Omega \subset \mathbb{R}^2$  with boundary  $\Gamma = \partial\Omega$  which we assume to be piecewise smooth.

Under the influence of a force with density  $f = f(x), x \in \Omega$ , which acts perpendicular to the  $(x_1, x_2)$ -plane, the membrane is deflected in the  $x_3$  direction. The deflection will be described by a function  $u = u(x_1, x_2)$ . Since the membrane is clamped at its boundary  $\Gamma$ , there is no deflection for  $x \in \Gamma$ , i.e., we have  $u(x_1, x_2) = 0, x = (x_1, x_2) \in \Gamma$ .

The **equilibrium state** is characterized as the physical state, where the **total energy** of the membrane attains its minimum. The total energy  $J = J(u)$  consists of the potential energy  $J = J_p$  and the

energy  $J = J_f$  associated with the exterior force  $f$  according to

$$(1.1) \quad J = J_p - J_f .$$

As said before, the potential energy is proportional to the change in surface area

$$\int_{\Omega} (1 + |\nabla u|^2)^{1/2} dx - \int_{\Omega} dx ,$$

where  $\nabla u := (\partial u / \partial x_1, \partial u / \partial x_2)$ .

If we restrict ourselves to small deflections, i.e.,  $|\nabla u| \ll 1$ , we obtain

$$\int_{\Omega} (1 + |\nabla u|^2)^{1/2} dx = \int_{\Omega} dx + \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx + o(|\nabla u|^2) .$$

This results in

$$(1.2) \quad J_p(u) = \frac{\mu}{2} \int_{\Omega} |\nabla u|^2 dx ,$$

where the quantity  $\mu > 0$  is a material constant that reflects the **elastic response** of the membrane.

On the other hand, we have

$$(1.3) \quad J_f(u) = \int_{\Omega} fu dx .$$

Consequently, using (1.2) and (1.3) in (1.1) yields

$$(1.4) \quad J(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx - \int_{\Omega} fu dx .$$

We refer to  $L^2(\Omega)$  as the Hilbert space of square integrable functions with the inner product

$$(u, v)_{0, \Omega} := \int_{\Omega} uv dx .$$

We equip the function space  $C_0^1(\Omega)$  with the inner product

$$(u, v)_{1, \Omega} := \int_{\Omega} uv dx + \int_{\Omega} \nabla u \cdot \nabla v dx .$$

However, the space  $V := (C_0^1(\Omega) | \|\cdot\|_{1, \Omega})$  is not complete. We denote its completion with respect to the  $\|\cdot\|_{1, \Omega}$ -norm by  $H_0^1(\Omega)$ . (A systematic introduction to Sobolev spaces will be provided in Chapter 2).

Assuming  $f \in L^2(\Omega)$ , the determination of the stationary state of the deflected membrane amounts to the solution of the **unconstrained**

**minimization problem:**

Find  $u \in H_0^1(\Omega)$  such that

$$(1.5) \quad J(u) = \inf_{v \in H_0^1(\Omega)} J(v) .$$

Minimization problems such as (1.5) can be shown to admit a unique solution under more general assumptions. We assume  $V$  to be a Hilbert space with inner product  $(\cdot, \cdot)_V$  and consider a functional  $J : V \rightarrow \overline{\mathbb{R}} := [-\infty, +\infty]$ .

**Definition 1.1 Lower semicontinuous functionals**

A functional  $J : V \rightarrow \overline{\mathbb{R}}$  is called **lower semicontinuous** [**weakly lower semicontinuous**] at  $u \in V$ , if for any sequence  $(u_n)_{n \in \mathbb{N}}$  such that

$$u_n \rightharpoonup u \quad (n \rightarrow \infty) \quad [u_n \rightharpoonup u \quad (n \in \mathbb{N})]$$

there holds

$$J(u) \leq \liminf_{n \rightarrow \infty} J(u_n) .$$

**Example 1.2. Examples of lower semicontinuous functionals**

(i) Let  $V := \mathbb{R}$  and

$$J(v) := \begin{cases} +1 & , \quad v < 0 \\ -1 & , \quad v \geq 0 \end{cases} .$$

Then  $J$  is lower semicontinuous on  $\mathbb{R}$ .

(ii) Let  $K \subset V$  be a closed, convex set. Then, the indicator function of  $K$  as given by

$$I_K(v) := \begin{cases} 0 & , \quad v \in K \\ +\infty & , \quad v \notin K \end{cases}$$

is lower semicontinuous on  $V$ .

**Lemma 1.3** [[1]] **Properties of convex sets**

Assume that  $K \subset V$  is a closed, convex set. Then  $K$  is weakly closed.

**Definition 1.4 Proper convex functionals**

We recall that a functional  $J : V \rightarrow \overline{\mathbb{R}}$  is called convex, if for all  $u, v \in V$  and  $\lambda \in [0, 1]$  there holds

$$J(\lambda u + (1 - \lambda)v) \leq \lambda J(u) + (1 - \lambda)J(v) ,$$

provided the right-hand side in the above inequality is well-defined.

A convex functional  $J : V \rightarrow \overline{\mathbb{R}}$  is said to be **proper convex**, if

$J(v) > -\infty, v \in V$ , and  $J \neq +\infty$ .

A convex functional  $J : V \rightarrow \overline{\mathbb{R}}$  is called **strictly convex**, if

$$J(\lambda u + (1 - \lambda)v) < \lambda J(u) + (1 - \lambda)J(v)$$

for all  $u, v \in V, u \neq v$ , and  $\lambda \in (0, 1)$ .

**Lemma 1.5** [[1]] **Properties of convex functionals**

Assume that  $J : V \rightarrow \overline{\mathbb{R}}$  is a lower semicontinuous, convex functional. Then  $J$  is weakly lower semicontinuous.

**Definition 1.6** **Coercive functionals**

A functional  $J : V \rightarrow \overline{\mathbb{R}}$  is said to be **coercive**, if

$$J(v) \rightarrow +\infty \quad \text{for } \|v\|_V \rightarrow +\infty .$$

**Theorem 1.7** **Solvability of minimization problems**

Suppose that  $J : V \rightarrow (-\infty, +\infty], J \neq +\infty$ , is a weakly semicontinuous, coercive functional. Then, the unconstrained minimization problem

$$(1.6) \quad J(u) = \inf_{v \in V} J(v)$$

admits a solution  $u \in V$ .

**Proof.** Let  $c := \inf_{v \in V} J(v)$  and assume that  $(v_n)_{n \in \mathbb{N}}$  is a minimizing sequence, i.e.,  $J(v_n) \rightarrow c$  ( $n \rightarrow \infty$ ).

Since  $c < +\infty$  and in view of the coercivity of  $J$ , the sequence  $(v_n)_{n \in \mathbb{N}}$  is bounded. Consequently, there exist a subsequence  $N' \subset \mathbb{N}$  and  $u \in V$  such that  $v_n \rightharpoonup u$  ( $n \in \mathbb{N}'$ ). The weak lower semicontinuity of  $J$  implies

$$J(u) \leq \inf_{n \in \mathbb{N}'} J(v_n) = c ,$$

whence  $J(u) = c$ . •

**Theorem 1.8** **Existence and uniqueness**

Suppose that  $J : V \rightarrow \overline{\mathbb{R}}$  is a proper convex, lower semicontinuous, coercive functional. Then, the unconstrained minimization problem (1.6) has a solution  $u \in V$ .

If  $J$  is strictly convex, then the solution is unique.

**Proof.** In view of Lemma 1.5 the existence follows from Theorem 1.7. For the proof of the uniqueness let  $u_1 \neq u_2$  be two different solutions.

Then there holds

$$J\left(\frac{1}{2}(u_1 + u_2)\right) < \frac{1}{2} J(u_1) + \frac{1}{2} J(u_2) = \inf_{v \in V} J(v) ,$$

which is a contradiction. •

We recall that in the finite dimensional case  $V = \mathbb{R}^n$ , a necessary optimality condition for (1.6) is that  $\nabla J(u) = 0$ , provided  $J$  is continuously differentiable. This can be easily generalized to the infinite dimensional case.

### Definition 1.9 Gateaux-Differentiability

A functional  $J : V \rightarrow \overline{\mathbb{R}}$  is called **Gateaux-differentiable** in  $u \in V$ , if

$$J'(u; v) = \lim_{\lambda \rightarrow 0^+} \frac{J(u + \lambda v) - J(u)}{\lambda}$$

exists for all  $v \in V$ .  $J'(u; v)$  is said to be the **Gateaux-variation** of  $J$  in  $u \in V$  with respect to  $v \in V$ .

Moreover, if there exists  $J'(u) \in V^*$  such that

$$J'(u; v) = J'(u)(v) = \langle J'(u), v \rangle_{V^*, V} , \quad v \in V ,$$

then  $J'(u)$  is called the **Gateaux-derivative** of  $J$  in  $u \in V$ .

### Theorem 1.10 Necessary optimality condition

Assume that  $J : V \rightarrow \overline{\mathbb{R}}$  is Gateaux-differentiable in  $u \in V$  with Gateaux-derivative  $J'(u) \in V^*$ . Then, the variational equation

$$(1.7) \quad \langle J'(u), v \rangle_{V^*, V} = 0 \quad , \quad v \in V$$

is a necessary condition for  $u \in V$  to be a minimizer of  $J$ .

If  $J$  is convex, then this condition is also sufficient.

**Proof.** Let  $u \in V$  be a minimizer of  $J$ . Then, there holds

$$J(u \pm \lambda v) \geq J(u) \quad , \quad \lambda > 0 , \quad v \in V ,$$

whence

$$\langle J'(u), \pm v \rangle_{V^*, V} \geq 0 \quad , \quad v \in V ,$$

and thus

$$\langle J'(u), v \rangle_{V^*, V} = 0 \quad , \quad v \in V .$$

If  $J$  is convex and (1.7) holds true, then

$$J(u + \lambda(v - u)) = J(\lambda v + (1 - \lambda)u) \leq \lambda J(v) + (1 - \lambda)J(u) ,$$

and hence,

$$\begin{aligned} 0 &= \langle J'(u), v - u \rangle_{V',V} = \lim_{\lambda \rightarrow 0^*} \frac{J(u + \lambda(v - u)) - J(u)}{\lambda} \leq \\ &\leq J(v) - J(u) . \end{aligned}$$

•

### Exercise 1.11 Application to the membrane problem

Show that the membrane problem (1.5) has a unique solution  $u \in H_0^1(\Omega)$  that is the solution of the variational equation

$$(1.8) \quad \mu \int_{\Omega} \nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad , \quad v \in H_0^1(\Omega) .$$

Let us assume that  $f \in C(\Omega)$  and that the solution  $u \in H_0^1(\Omega)$  of (1.8) has the **regularity property**  $u \in C^2(\Omega) \cap C_0(\bar{\Omega})$ . Then, **Green's formula**

$$\int_{\Omega} u \frac{\partial v}{\partial x_i} \, dx = - \int_{\Omega} \frac{\partial v}{\partial x_i} v \, dx + \int_{\Gamma} n_i u v \, d\sigma \quad , \quad 1 \leq i \leq 2 ,$$

where  $\mathbf{n} := (n_1, n_2)^T$  is the exterior unit normal vector, yields

$$\int_{\Omega} \nabla u \cdot \nabla v \, dx = - \int_{\Omega} \Delta u v \, dx + \underbrace{\int_{\Gamma} \frac{\partial u}{\partial \mathbf{n}} v \, d\sigma}_{= 0} .$$

Consequently, (1.8) can be written as

$$(1.9) \quad \int_{\Omega} (-\mu \Delta u - f) v \, dx = 0 \quad , \quad v \in H_0^1(\Omega) .$$

Since (1.9) holds true for all  $v \in H_0^1(\Omega)$ , we conclude that

$$-\mu \Delta u(x) = f(x) \quad \text{f.f.a. } x \in \Omega .$$

By continuity, we find that  $u$  satisfies the **boundary value problem**

$$(1.10) \quad \begin{aligned} -\mu \Delta u(x) &= f(x) \quad , \quad x \in \Omega , \\ u(x) &= 0 \quad , \quad x \in \Gamma . \end{aligned}$$

### Definition 1.10 Euler equation

The boundary value problem (1.10) is referred to as the **Euler equation** associated with the minimization problem (1.5).

## 1.2 Existence and uniqueness results for variational equations

In this section, we are concerned with a fundamental **existence and uniqueness result** for variational equations in a Hilbert space setting. We assume that  $V$  is a Hilbert space with inner product  $(\cdot, \cdot)_V$  and associated norm  $\|\cdot\|_V$  and we refer to  $V^*$  as its algebraic and topological dual. Given a bilinear form  $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$  and an element  $\ell \in V^*$ , i.e., a bounded linear functional on  $V$ , we consider the **variational equation**:

Find  $u \in V$  such that

$$(1.11) \quad a(u, v) = \ell(v) \quad , \quad v \in V .$$

### Definition 1.11 Bounded bilinear forms

A bilinear form  $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$  is said to be **bounded**, if there exists a constant  $C \geq 0$  such that

$$(1.12) \quad |a(u, v)| \leq C \|u\|_V \|v\|_V \quad , \quad u, v \in V .$$

### Definition 1.12 $V$ -elliptic bilinear forms

A bilinear form  $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$  is called  **$V$ -elliptic**, if there exists a constant  $\alpha > 0$  such that

$$(1.13) \quad |a(u, u)| \geq \alpha \|u\|_V^2 \quad , \quad u \in V .$$

The following celebrated **Lemma of Lax-Milgram** is one the fundamental principles of the finite element analysis and states an existence and uniqueness result for variational equations of the form (1.11).

### Theorem 1.11 Lemma of Lax-Milgram

Let  $V$  be a Hilbert space with dual  $V^*$  and assume that  $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$  is a bounded,  $V$ -elliptic bilinear form and  $\ell \in V^*$ . Then, the variational equation (1.11) admits a unique solution  $u \in V$ .

**Proof.** We denote by  $A : V \rightarrow V^*$  the operator associated with the bilinear form, i.e.,

$$(1.14) \quad Au(v) = a(u, v) \quad , \quad u, v \in V .$$

We note that  $A$  is a bounded linear operator, since in view of (1.12)

$$\|Au\|_{V^*} = \sup_{v \neq 0} \frac{|Au(v)|}{\|v\|_V} \leq C \|u\|_V .$$

and hence,

$$\|A\| = \sup_{u \neq 0} \frac{\|Au\|_{V^*}}{\|u\|_V} \leq C .$$

In order to state the variational equation (1.11) as an operator equation in  $V$ , we further consider a lifting from  $V^*$  to  $V$  provided by the **Riesz map**

$$\begin{aligned} \tau : V^* &\rightarrow V \\ \ell &\longmapsto \tau\ell , \end{aligned}$$

given by

$$(1.15) \quad \ell(v) = (\tau\ell, v)_V \quad , \quad v \in V .$$

It is easy to verify that the Riesz map is an **isometry**, i.e.,

$$(1.16) \quad \|\tau\ell\|_V = \|\ell\|_{V^*} \quad , \quad \ell \in V^* .$$

In terms of the Riesz map, the variational equation (1.11) can be equivalently written as the **operator equation**

$$(1.17) \quad \tau(Au - \ell) = 0 .$$

We prove the existence and uniqueness of a solution of (1.17) by an application of the **Banach fixed point theorem**. to this end we introduce the map

$$(1.18) \quad \begin{aligned} T : V &\rightarrow V \\ Tv &:= v - \rho (\tau(Av - \ell)) \quad , \quad \rho > 0 . \end{aligned}$$

Obviously,  $u \in V$  is a solution of (1.17) if and only if  $u \in V$  is a fixed point of the operator  $T$ . To prove existence and uniqueness of a fixed point in  $V$ , we have to show that the operator  $T$  is a **contraction** on  $V$ , i.e., there exists a constant  $0 \leq q < 1$  such that

$$(1.19) \quad \|Tv_1 - Tv_2\|_V \leq q \|v_1 - v_2\|_V \quad , \quad v_1, v_2 \in V .$$

Indeed, setting  $w := v_1 - v_2$  we have

$$\begin{aligned} \|Tw\|_V^2 &= \|w - \rho \tau Aw\|_V^2 = \\ &= \|w\|_2^2 - 2\rho \underbrace{(\tau Aw, w)_V}_{= a(w, w)} + \rho^2 \underbrace{\|\tau Aw\|_V^2}_{= \|Aw\|_{V^*}^2} \leq \\ &\leq \left(1 - 2\rho\alpha + \rho^2 C^2\right) \|w\|_V^2 . \end{aligned}$$

Clearly, we have

$$\left(1 - 2\rho\alpha + \rho^2 C^2\right) < 1 \quad \iff \quad \rho < \frac{2\alpha}{C^2} ,$$



which gives the assertion. •

### 1.3 The Ritz-Galerkin method

In general, we cannot access the solution of the variational equation (1.11) analytically. Therefore, we have to resort to numerical methods. A widely used approach is the so-called **Ritz-Galerkin method**, where we are looking for an **approximate solution** in a suitable **finite dimensional subspace** of  $V_h \subset V$ :

Find  $u_h \in V_h$  such that

$$(1.20) \quad a(u_h, v_h) = \ell(v_h) \quad , \quad v_h \in V_h .$$

If  $a(\cdot, \cdot)$  is a bounded,  $V$ -elliptic bilinear form, then the Lemma of Lax-Milgram tells us that (1.20) admits a unique solution  $u_h \in V_h$ .

The appropriate choice of  $V_h$  is dominated by **two important issues** that will be the focus of our analysis in the subsequent chapters:

- **Construction of  $V_h$**  such that the **linear algebraic system** represented by (1.20) can be efficiently solved.
- **Estimation of the global discretization error  $u - u_h$ .**

The first issue crucially hinges on the proper selection of a **basis** of the finite dimensional subspace  $V_h$ :

$$(1.21) \quad V_h = \text{span} \{ \varphi_1, \dots, \varphi_{n_h} \} \quad , \quad \dim V_h = n_h .$$

We may represent the solution  $u_h$  as a linear combination of the basis functions according to

$$(1.22) \quad u_h = \sum_{j=1}^{n_h} \alpha_j \varphi_j .$$

Obviously, (1.20) is satisfied for any  $v_h \in V_h$  if and only if it holds true for all basis functions  $\varphi_i, \leq i \leq n_h$ . Therefore, inserting (1.21) into (1.20) and choosing  $v_h = \varphi_i, i = 1, \dots, n_h$ , we see that indeed (1.20) corresponds to the **linear algebraic system**

$$(1.23) \quad \sum_{j=1}^{n_h} a(\varphi_j, \varphi_i) \alpha_j = \ell(\varphi_i) \quad , \quad 1 \leq i \leq n_h .$$

Motivated by applications in structural mechanics, we define:

#### Definition 1.13 Stiffness matrix and load vector

The matrix  $A_h = (a_{ij})_{i,j=1}^{n_h}$  with entries

$$(1.24) \quad a_{ij} := a(\varphi_j, \varphi_i) \quad , \quad 1 \leq i, j \leq n_h .$$

is called the **stiffness matrix** and the vector  $b_h = (b_1, \dots, b_{n_h})^T$  with components

$$(1.25) \quad b_i := \ell(\varphi_i) \quad , \quad 1 \leq i \leq n_h$$

is referred to as the **load vector**.

The unknowns in (1.23) are the coefficients  $\alpha_i, 1 \leq i \leq n_h$  which constitute the solution vector  $\alpha_h = (\alpha_1, \dots, \alpha_{n_h})^T$ .

In summary, the linear algebraic system (1.23) can be concisely written as

$$(1.26) \quad A_h \alpha_h = b_h \ .$$

The second issue about the **accuracy of the Ritz-Galerkin** approach can be answered by an important orthogonality property of the error which is called **Galerkin orthogonality**.

### Lemma 1.6 Galerkin orthogonality

Under the assumptions of the Lax-Milgram Lemma, let  $u \in V$  and  $u_h \in V_h$  be the unique solutions of (1.11) and (1.20), respectively. Then, there holds

$$(1.27) \quad a(u - u_h, v_h) = 0 \quad , \quad v_h \in V_h \ .$$

**Proof.** Since  $V_h \subset V$ , (1.11) also holds true for all  $v_h \in V_h$ . Subtraction of (1.20) from (1.11) gives the assertion.

### Remark 1.1 Elliptic projection

The Galerkin orthogonality has an immediate and helpful **geometric interpretation**. If we additionally assume that the bilinear form  $a(\cdot, \cdot)$  is symmetric, it defines an inner product on  $V$ . Then, (1.27) tells us that  $u_h \in V_h$  is the projection of  $u \in V$  onto  $V_h$ . Since  $a(\cdot, \cdot)$  is  $V$ -elliptic,  $u_h \in V_h$  is said to be the **elliptic projection** of  $u \in V$  onto  $V_h$ .

In lights of the preceding remark, we expect the error  $u - u_h$  in the norm  $\|\cdot\|_V$  to be related to the **best approximation** of the solution  $u \in V$  by an element in  $V_h$ .

This holds true in the general case and is the result of another fundamental tool in the finite element analysis, the **Lemma of Céa**:

### Theorem 1.12 Lemma of Céa

Under the assumptions of the Lemma of Lax-Milgram let  $u \in V$  and  $u_h \in V_h$  be the unique solutions of (1.11) and (1.20), respectively. Then, there holds

$$(1.28) \quad \|u - u_h\|_V \leq \frac{C}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V \ .$$

**Proof.** Taking the  $V$ -ellipticity and the boundedness of  $a(\cdot, \cdot)$  as well as the Galerkin orthogonality (1.27) into account, for any  $v_h \in V_h$  we have

$$\begin{aligned} \alpha \|u - u_h\|_V^2 &\leq a(u - u_h, u - u_h) = \\ &= a(u - u_h, u - v_h) + \underbrace{a(u - u_h, v_h - u_h)}_{= 0} \leq \\ &\leq C \|u - u_h\|_V \|u - v_h\|_V, \end{aligned}$$

which proves (1.28). •

#### REFERENCES

- [1] Hiriart-Urruty, J. and L  mar  chal, C. (1993). *Convex Analysis and Minimization Algorithms*. Springer, Berlin-Heidelberg-New York